

## RESEARCH ARTICLE

## Machine learning prediction of malaria vaccine efficacy based on antibody profiles

Jacqueline Wistuba-Hamprecht<sup>1,2</sup>, Bernhard Reuter<sup>1,2\*</sup>, Rolf Fendel<sup>3,4,5</sup>, Stephen L. Hoffman<sup>6</sup>, Joseph J. Campo<sup>7</sup>, Philip L. Felgner<sup>8</sup>, Peter G. Kremsner<sup>3,4,5</sup>, Benjamin Mordmüller<sup>3,5,9</sup>, Nico Pfeifer<sup>1,2,4\*</sup>

**1** Department of Computer Science, University of Tübingen, Tübingen, Germany, **2** Institute for Biomedical Informatics, University of Tübingen, Tübingen, Germany, **3** Institute of Tropical Medicine, University of Tübingen, Tübingen, Germany, **4** German Center for Infection Research, partner site Tübingen, Tübingen, Germany, **5** Centre de Recherches Médicales de Lambaréné, Lambaréné, Gabon, **6** Sanaria Inc., Rockville, Maryland, United States of America, **7** Antigen Discovery Inc., Irvine, California, United States of America, **8** Department of Medicine, University of California Irvine, Irvine, California, United States of America, **9** Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, The Netherlands

\* [bernhard.reuter@uni-tuebingen.de](mailto:bernhard.reuter@uni-tuebingen.de) (BR); [nico.pfeifer@uni-tuebingen.de](mailto:nico.pfeifer@uni-tuebingen.de) (NP)



## OPEN ACCESS

**Citation:** Wistuba-Hamprecht J, Reuter B, Fendel R, Hoffman SL, Campo JJ, Felgner PL, et al. (2024) Machine learning prediction of malaria vaccine efficacy based on antibody profiles. *PLoS Comput Biol* 20(6): e1012131. <https://doi.org/10.1371/journal.pcbi.1012131>

**Editor:** Andrew J. Yates, Columbia University Irving Medical Center, UNITED STATES

**Received:** October 5, 2021

**Accepted:** May 2, 2024

**Published:** June 7, 2024

**Copyright:** © 2024 Wistuba-Hamprecht et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data and code used for running experiments and model fitting is available on a GitHub repository at: <https://github.com/msmdev/MalariaVaccineEfficacyPrediction>.

**Funding:** N.P. was supported by the DFG Cluster of Excellence Machine Learning – New Perspectives for Science (EXC 2064/1, project no. 390727645) and by the BMBF (Tübingen AI Center, FKZ: 01IS18039A). B.R. was supported by the German Federal Ministry of Education and Research

## Abstract

Immunization through repeated direct venous inoculation of *Plasmodium falciparum* (Pf) sporozoites (PfSPZ) under chloroquine chemoprophylaxis, using the PfSPZ Chemoprophylaxis Vaccine (PfSPZ-CVac), induces high-level protection against controlled human malaria infection (CHMI). Humoral and cellular immunity contribute to vaccine efficacy but only limited information about the implicated Pf-specific antigens is available. Here, we examined Pf-specific antibody profiles, measured by protein arrays representing the full Pf proteome, of 40 placebo- and PfSPZ-immunized malaria-naïve volunteers from an earlier published PfSPZ-CVac dose-escalation trial. For this purpose, we both utilized and adapted supervised machine learning methods to identify predictive antibody profiles at two different time points: after immunization and before CHMI. We developed an adapted multitask support vector machine (SVM) approach and compared it to standard methods, i.e. single-task SVM, regularized logistic regression and random forests. Our results show, that the multitask SVM approach improved the classification performance to discriminate the protection status based on the underlying antibody-profiles while combining time- and dose-dependent data in the prediction model. Additionally, we developed the new feature distance explainability (ESP\_Y) method to quantify the impact of single antigens on the non-linear multitask SVM model and make it more interpretable. In conclusion, our multitask SVM model outperforms the studied standard approaches in regard of classification performance. Moreover, with our new explanation method ESP\_Y, we were able to interpret the impact of Pf-specific antigen antibody responses that predict sterile protective immunity against CHMI after immunization. The identified Pf-specific antigens may contribute to a better understanding of immunity against human malaria and may foster vaccine development.

(BMBF): Tübingen AI Center, FKZ: 01IS18039A. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: J.J.C. is a salaried employee of Antigen Discovery Inc., the manufacturer of the protein microarrays and performer of the experiments from which the data was derived. S.L.H. is an owner and salaried employee of Sanaria Inc. the manufacturer of PfSPZ-CVac and the Sponsor of the clinical trial from which the data was derived. P.L.F. is founder and stock holder of Nanomune Inc. and Antigen Discovery Inc.

## Author summary

Developing an effective malaria vaccine is challenging. Malaria is a life-threatening disease caused by the plasmodium parasite, which has a complex multi-stage life-cycle and expresses several thousand proteins in a highly coordinated manner. To date, our understanding of the immune mechanisms mediating protection against *Plasmodium falciparum* (Pf) is incomplete. Proteome microarrays have been used earlier by our clinical collaboration partners to identify Pf-specific antibody profiles of malaria-naïve volunteers during immunization with attenuated Pf sporozoites (PfSPZ). We reused this data to compare the ability of three supervised machine learning methods to identify predictive antibody profiles after immunization and before controlled human malaria infection (CHMI). We adapted a multitask support vector machine (SVM) approach to analyze time-dependent Pf-induced antibody profiles from several time points in a single prediction model. Our multitask SVM approach outperforms the studied standard approaches in classification performance. Additionally, we developed a new explanation method, named *feature distance explainability* (ESPY), to interpret the impact of Pf-specific antigens. We applied ESPY on the multitask SVM model and identified diverse Pf-specific antigen sets after immunization and before CHMI. Furthermore, we showed that the identified Pf-induced antibody profiles vary among protected and non-protected individuals who had been exposed to different doses of PfSPZ.

## Introduction

Malaria is a major health problem: alone in 2022 it caused more than 249 million cases and approximately 608,000 deaths [1]. *Plasmodium falciparum* (Pf) is the causal agent of almost all malaria-related deaths. Children, pregnant women and malaria-naïve subjects are at high risk of developing severe malaria, whereas adult residents of highly endemic areas develop immunity that protects from severe disease [2–4]. In addition, proof-of-concept studies have shown that experimental inoculation of high doses of attenuated Pf sporozoites (PfSPZ) (the mosquito-to-human transmission stage of the parasite) can lead to sterile protection [5]. Nevertheless, developing an effective vaccine for Pf remains a huge challenge. Pf is genetically highly diverse, employs several immune evasion strategies and has a complex, multi-stage life-cycle, during which more than 5,300 genes are expressed [6]. As a result, our understanding of immune responses to Pf-specific antigens that mediate naturally acquired or experimentally induced protection is incomplete.

It has been shown that up to 100% protection against controlled human malaria infection (CHMI) can be achieved by immunization of malaria-naïve adults by direct venous inoculation (DVI) of radiation-attenuated Pf sporozoites (Sanaria PfSPZ Vaccine) [7] and by chemo-attenuated PfSPZ (Sanaria PfSPZ-CVac) [8–10]. In those studies, protection is defined as an immune state that prevents parasites from reaching the blood stage, whereas in non-protected volunteers (either non-immunized or not successfully immunized participants) parasites will invade red blood cells following an approximately 6-day-long liver stage [8]. Only the asexual blood stage of the parasite is responsible for the symptoms and complications of malaria. Pf-specific protein microarrays can be used to characterize the pattern of antibody reactivity to Pf-specific epitopes. In [8] we used a Pf-specific protein microarray with 7,455 protein fragments, representing about 91% of the Pf proteome, to determine the antibody reactivity profile of 40 immunized and non-immunized malaria-naïve individuals after immunization and before CHMI. In this previous study [8], we showed, that among the subjects who received the

highest dose of attenuated PfSPZ, all of whom were protected, twenty-two proteins were recognized on the Pf proteome microarray by more than half of the protected subjects [8]. A limitation of protein microarrays is that antibody reactivity profiles are characterized by a huge number of features in a comparatively small number of samples, a problem that is better known as the curse of dimensionality. Since machine learning methods became a famous choice for analyzing such high-dimensional data [11–13], we sought to complete our previous, primarily descriptive analyses of the data [8] to better understand and even predict PfSPZ-CVac-induced protective immunity.

Therefore, we adapted here a multitask support vector machine (multitask SVM) approach to identify predictive Pf-specific antibody profiles of protected and non-protected vaccinees and controls by integrating time- and dose-dependent data in a single prediction model. Combining related tasks into a single prediction model is more promising than training independent models with the data of each task, if the number of features is much greater than the number of samples [14]. SVM kernels can be used to model relationships between single related tasks and combine them into a sole prediction model—the multitask SVM.

Analysing such a large array of antibody profiles using a proteome microarray, strong correlations can be assumed, e.g., between fragments representing one protein, similar epitopes, and due to cross-reactivity [15]. In general, it is advised to remove strongly linearly correlated features, to avoid biasing the variable importance measure of the features [13] and to improve classification performance [16, 17]. Therefore, we assessed the classification performance of our adapted multitask SVM approach under conditions in which highly linearly correlated features were removed, and compared it to state-of-the-art methods, such as regularized logistic regression (RLR) [11], a standard SVM model with radial basis function (RBF) kernel, and a random forest (RF) approach. All these methods are known to be able to deal with high-dimensional data for the classification of protected versus non-protected vaccinees and controls [11, 13]. To enable an optimal comparison, we trained the three state-of-the-art methods, RLR, RF and RBF-SVM, respectively, either time-point-wise (task-wise) or time-point-combined (multi-time). In the former one, samples that belonged to different time points (that is, before and after immunization) were separately used to train the models, whereas in the latter one, the samples from both time points were combined to train the models (for more details see the materials and methods section). Our results show that the adapted multitask SVM approach improves the prediction performance when classifying protected PfSPZ-CVac vaccinees versus non-protected PfSPZ-CVac vaccinees and controls. Moreover, we can show that highly correlated features degraded classification performance of the state-of-the-art methods compared to our multitask SVM approach.

To identify and interpret informative features, i.e., single Pf-specific antigens, from the non-linear multitask SVM model, explainability models for non-linear machine learning models are needed, which motivated us to develop the *f*Eature *d*iStance *e*xPlainability (ESPY) method. ESPY is inspired by a feature importance measure for sequence-based non-linear predictions [18]. The ESPY values are directly derived from a multitask SVM (or general SVM) model. ESPY uses systematically and specifically triggered changes in the distance of a consensus sample to the classification boundary (the boundary that separates the datapoints into two sets, one of each class in a binary classification scenario) of the SVM to estimate the importance of features (for more details see the materials and methods section), but could be extended to any machine learning model that provides classification scores indicating to how certain a classification is. Consequently, we identified individual informative Pf-specific antigens by their respective ESPY value for protected PfSPZ-CVac vaccinees and non-protected PfSPZ-CVac vaccinees and controls. Additionally, we compared our ESPY method with the SHapley Additive exPlanation (SHAP) framework from Lundberg et al. [19] on simulated

data. The SHAP framework is an additive unified approach to derive feature importance values. The results of both explainability methods are similar on the simulated data, however ESPY significantly outperformed SHAP in run time: ESPY needed only a few seconds to compute feature importance values on the simulated data, while SHAP ran for more than 13 hours. Using our newly developed ESPY method, we show how to address the problem of explaining the predictions from a non-linear multitask SVM model based on single features. In summary, our adapted multitask SVM approach represents a classification method to integrate time- and dose-dependent data into a single prediction model, while ESPY provides explainability by means of identifying and evaluating informative single features from a non-linear model.

## Results

This section is structured into three main parts. In the first part, we show that our new multitask SVM approach can be used to classify with high accuracy protected PfSPZ-CVac vaccinees versus non-protected PfSPZ-CVac vaccinees and controls based on a subset of antibody (ab) intensity signals and excluding those that strongly linearly correlated above a Pearson correlation coefficient of  $pcc = 0.8$  for both time points (post-immunization and pre-CHMI). We highlight the classification performance of our multitask SVM approach at each single time point for both the whole Pf-specific proteome microarray and a selection of cell-surface Pf-antigens compared to standard machine learning approaches. To enable this comparison, we trained the standard machine learning models either on each single time point separately or on a combined set of all ab signal intensities for both time points and PfSPZ doses together.

In the second part, we illustrate how the ESPY values are used to quantify which Pf-specific antigens are informative in classifying protected PfSPZ-CVac vaccinees versus non-protected PfSPZ-CVac vaccinees and controls.

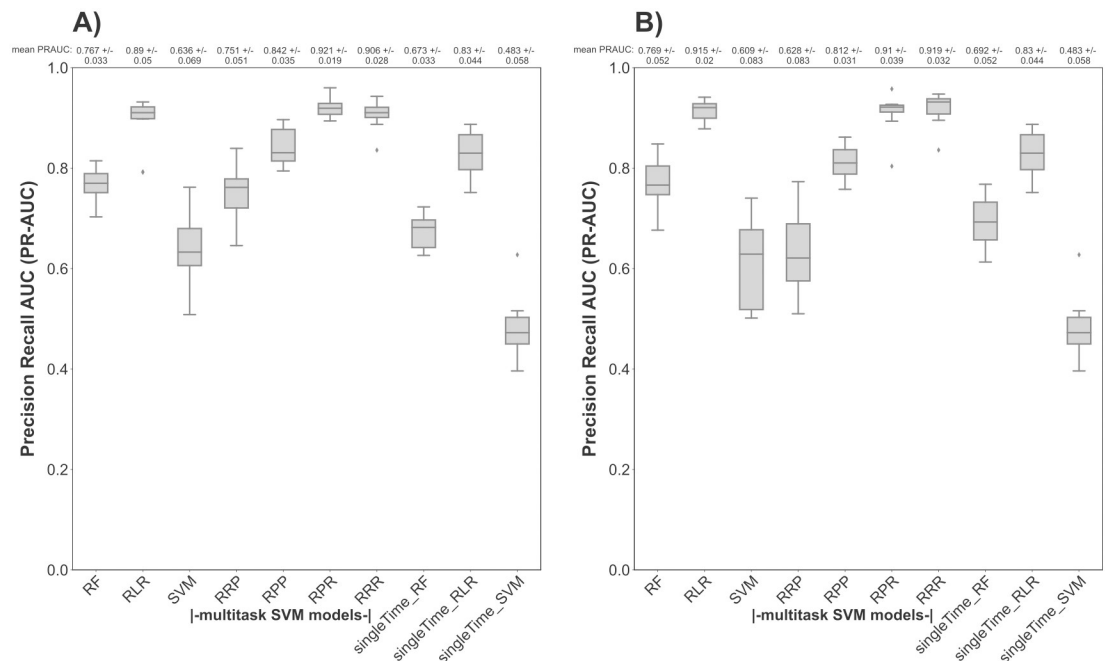
The third and last part of this section shows ESPY values for simulated data and compares those with the SHAP values from Lundberg et al. [19].

In the following lines, for a better understanding of our results, we briefly summarize which data we used and how we applied our methods (for more details please refer to the materials and methods section).

To compare our new multitask SVM approach with standard machine learning methods, i.e., standard single-task SVM, RLR, and RF, we used the Pf-specific antibody reactivity profile from the earlier published PfSPZ-CVac clinical trial by [8]. The Pf-specific ab reactivity profile contains Pf-specific ab-mediated responses of 40 individuals at two different time points (post-immunization and pre-CHMI). The 40 individuals were vaccinated with different doses of PfSPZ-CVac: placebo ( $n = 13$ ),  $3.2 \times 10^3$  PfSPZ ( $n = 9$ ),  $1.28 \times 10^4$  PfSPZ ( $n = 9$ ),  $5.12 \times 10^4$  PfSPZ ( $n = 9$ ). For each individual the protection status, i.e., protected or non-protected, was assessed by CHMI. The comparison of the classification performance and the identification of the informative Pf-specific antigens was done at each single time point (post-immunization and pre-CHMI) through the overall study. Antibody responses after CHMI (post-CHMI) were excluded from our malaria vaccine efficacy prediction analysis since controls underwent a CHMI at this time point as well, making it unfeasible to apply binary classification, due to the lack of non-protected controls.

### Classification of protected PfSPZ-CVac vaccinees versus non-protected PfSPZ-CVac vaccinees and controls from the PfSPZ-CVac mediated antibody response

We used a multitask SVM approach to build our prediction model. The multitask SVM builds upon a multitask kernel matrix that is constructed from the single-task kernels matrices via



**Fig 1. Performance of multitask SVM models in predicting the protection status based on the antibody reactivity profile per time point as compared to state-of-the-art approaches.** The PR-AUC score of the RF, the RLR, the single-task SVM (trained either on each single time point or on the combined time points), and the multitask SVM model (using different combinations of kernel functions) for predicting the protection status based on the whole proteome antibody profile per time point was assessed via 10-times repeated nested stratified 5-fold cross-validation. RF, RLR and single-task SVM models trained on each time point separately are labeled by the extension 'singleTime'. The mean PR-AUC score together with the standard deviation is displayed above each boxplot, with PR-AUC = 1 equating to perfect prediction and PR-AUC = 0.5 equating to random guessing. The PR-AUC performance of the different applied models is shown (A): at post-immunization and (B): at pre-CHMI.

<https://doi.org/10.1371/journal.pcbi.1012131.g001>

element-wise multiplication (see materials and methods section). A critical step in building a multitask SVM model is to find the right combination of single-task kernels to encode the relationships found in the input data. We explored the following combinations of a time point kernel, an antibody signal kernel, and a dose kernel: the time point kernel matrix was calculated using a RBF kernel function, while the antibody signal and the dose kernels were either calculated using a polynomial or RBF kernel function. The prediction performance, the PR-AUC score (*Area Under the Precision Recall Curve*) of each studied kernel combination at each single time point was assessed via a 10-times repeated nested stratified 5-fold cross-validation over a grid of kernel parameters Table A in [S1 Appendix](#). Due to the fact, that in our datasets the number of features outweighs the number of samples, we removed strongly linear correlated features above a Pearson correlation coefficient of  $pcc = 0.8$  for the whole and a selection of cell-surface antibody reactivity profile, as earlier described by [13]. For details please refer to the materials and methods section.

Based on prediction performance comparisons between the studied kernel combinations (Table A in [S1 Appendix](#)), the combination of three RBF kernels (RRR) and the combination of RBF and polynomial kernels (RPR) into a multitask kernel matrix resulted in the highest nested cross-validation PR-AUC score compared to all other kernel combinations (Fig 1). First of all, we compared our multitask SVM approach with three state-of-the-art approaches, namely RLR (with elastic net regularization), RF and single-task RBF-SVM at each single time point based on the complete antibody reactivity profile (Fig 1). Moreover, to illustrate the

differences in classification performance between our multitask SVM approach and state-of-the-art approaches we trained the state-of-the-art approaches either on each time point separately (stated by the extension “singleTime” for simplicity) ( $n = 40$ ) or on the combined time points ( $n = 80$ ). For details please refer to the materials and methods section. As shown in (Fig 1), the highest mean PR-AUC scores of the multitask SVM approaches were achieved by the kernel combinations RRR and RPR after immunization (Fig 1A) and before CHMI (Fig 1B) compared to the single-task RBF-SVM, the RLR and RF model trained on each time point separately or on the combined time points. Only before CHMI (Fig 1B) the RLR model trained on both time points achieved a similar mean PR-AUC score as the multitask SVM model for the RPR kernel combinations. The prediction performance of the state-of-the-art models, RF, RLR and single-task RBF-SVM was also assessed via a 10-times repeated nested stratified 5-fold cross-validation per time point (as detailed in the materials and methods section).

Finally, for a comprehensive view we compared the classification performances of our multitask SVM model and the state-of-the-art approaches for different Pearson correlation coefficients. As shown in Fig A in S1 Appendix, our multitask SVM model (for the kernel combinations RPR and RRR) is robust in its PR-AUC scores with less variance over different Pearson correlation coefficients in comparison to the state-of-the-art approaches after immunization (Fig A A in S1 Appendix) and before CHMI (Fig A B in S1 Appendix).

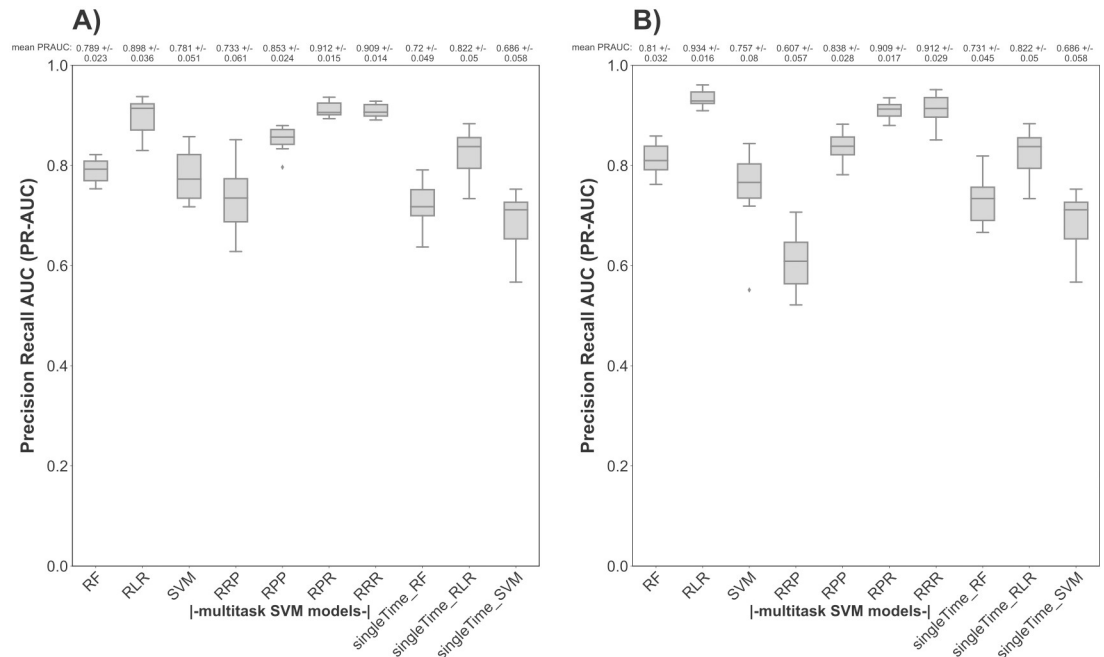
In a second step, we used our multitask SVM approach to analyze the antibody profile against pre-selected Pf-specific cell-surface antigens. Fig 2 illustrates the prediction performance (PR-AUC) of the compared models based on the measured selective cell-surface antibody reactivity profile per time point. Again, the combination of three RBF kernels (RRR) and the combination of RBF and polynomial kernels (RPR) into a multitask kernel matrix resulted in the highest nested cross-validation PR-AUC score compared to all other kernel combinations at post-immunization (Fig 2A) and pre-CHMI (Fig 2B). The single-task RBF-SVM, trained either on each single time point separately or on the combined time points, performed relatively poorly in comparison to the multitask SVM approach. The RLR model trained on the combined time points achieved a higher mean PR-AUC score only at pre-CHMI (Fig 2B) compared to the multitask SVM approach. However, at post-immunization (Fig 2A) the RLR approach achieved lower mean PR-AUC values.

For the selective dataset, we could also show, that our multitask SVM approach (for the kernel combinations RPR and RRR) is robust in its PR-AUC scores with less variance over different Pearson correlation coefficients in comparison to the state-of-the-art approaches after immunization (Fig B A in S1 Appendix) and before CHMI (Fig B B in S1 Appendix).

Overall these results demonstrate that our multitask SVM approach provides state-of-the-art performance in classifying protected PfSPZ-CVac vaccinees versus non-protected PfSPZ-CVac vaccinees and controls per time point by combining different tasks into a single model. Also, strongly linear correlated features have only a small effect on the prediction performance of our multitask SVM approach in comparison to the state-of-the-art approaches (Figs A and B in S1 Appendix).

### **Informative Pf-specific antigens for successful classification of protected PfSPZ-CVac vaccinees versus non-protected PfSPZ-CVac vaccinees and controls**

Evaluation of informative Pf-specific antigens to exhibit an antibody profile at post-immunization and pre-CHMI is essential for predicting and improving vaccine-induced protective immunity. To identify antibody profiles from the underlying PfSPZ-CVac dataset, where the number of Pf-specific antigens ( $p = 7,455$ ) is much higher than the number of individuals

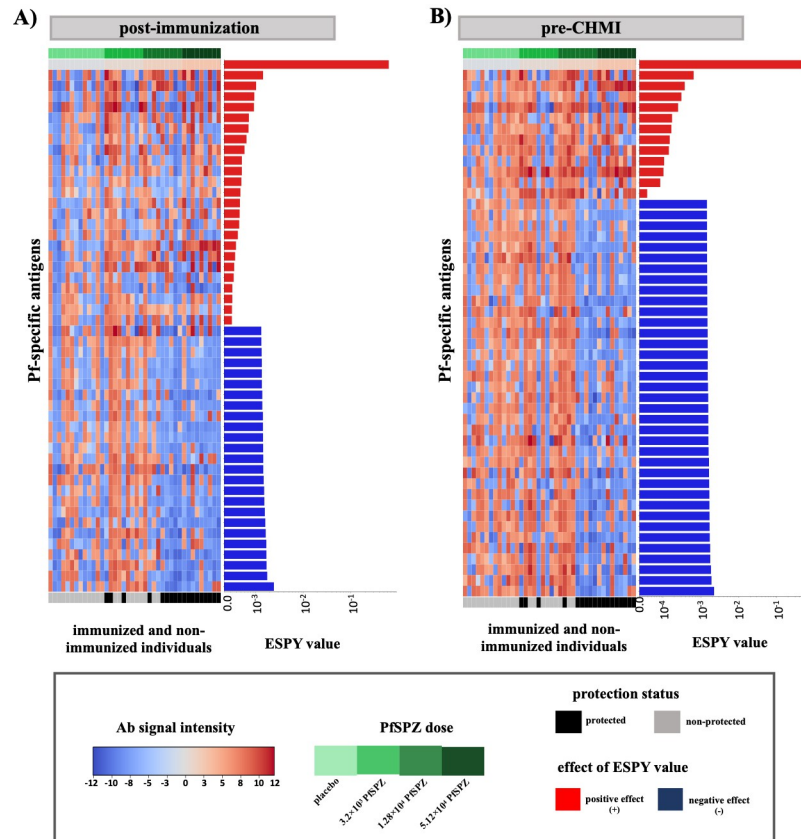


**Fig 2. Performance of multitask SVM models in predicting the protection status based on cell-surface antibody reactivity profile per time point as compared to state-of-the-art approaches.** The PR-AUC score of the RF, the RLR, the single-task SVM (trained either on each single time point or on the combined time points), and the multitask SVM model (using different combinations of kernel functions) for predicting the protection status based on the selective cell-surface proteome antibody profile per time point was assessed via 10-times repeated nested stratified 5-fold cross-validation. RF, RLR and single-task SVM models trained on each time point separately are labeled by the extension ‘singleTime’. The mean PR-AUC score together with the standard deviation is displayed above each boxplot, with PR-AUC = 1 equating to perfect prediction and PR-AUC = 0.5 equating to random guessing. The PR-AUC performance of the different applied models is shown (A): at post-immunization, (B): at pre-CHMI.

<https://doi.org/10.1371/journal.pcbi.1012131.g002>

( $n = 40$ ), we combined data from several time points and tasks into a non-linear multitask model and identified and evaluated informative Pf-specific antigens per time point from this model (for details please refer to the materials and methods section). First, we evaluated the best parameter setting for the kernel combinations in the multitask SVM models by a 10-times repeated stratified 5-fold grid-search cross-validation for both the whole proteome microarray and the pre-selected set of cell-surface antigens. The best kernel combinations and the associated parameters were selected based on the highest mean PR-AUC score per time point. Again the kernel combinations of the “RPR” and “RRR” performed equally well (achieving the highest mean PR-AUC scores), whereas the kernel combination “RRR” resulted in a more stable prediction performance with lower standard deviation than the kernel combination “RPR” for the multitask SVM approach. Table B in [S1 Appendix](#) shows the combinations of kernels and the associated parameters of the multitask SVM that achieved the highest mean AUC score, both based on the whole Pf-specific proteome microarray and the pre-selected cell-surface antigens. The kernel combinations of choice for both datasets were the RBF kernel for time point similarity, the RBF kernel for the antibody signal intensity similarity, and the RBF kernel for the PfSPZ-specific dose similarity at post-immunization and pre-CHMI.

Second, we evaluated the Pf-induced antibody profiles of all individuals based on immunoreactivity to 574 Pf-specific antigens after removing strongly linearly correlated features above a Pearson correlation coefficient of  $pcc = 0.8$ . [Fig 3](#) shows the antibody profile (at post-



**Fig 3. Antibody profile of protected and non-protected vaccinees and the placebo group against informative Pf-specific antigens.** Informative Pf-specific antigens of the complete proteome microarray were evaluated at post-immunization and pre-CHMI. Pf-specific antigens identified to be important by ESPY evaluation showed either a high antibody signal intensity in protected vaccinees or unprotected vaccinees and controls. The top 50 Pf-specific antigens with the highest ESPY values are shown (A) at post-immunization and (B) at pre-CHMI. The heatmap plot shows the antibody signal intensity, while the bars on the right side of each figure show the importance and effect of each feature based on the ESPY value. ESPY values of Pf-specific antigens, that were evaluated to have a positive effect on the protection status classification are colored in red, while blue-colored bars represents antigens, that have a negative effect.

<https://doi.org/10.1371/journal.pcbi.1012131.g003>

immunization and pre-CHMI) for the top 50 informative Pf-specific antigens selected using the ESPY values from the multitask SVM model. To better understand the informativeness of evaluated Pf-specific antigens in the classification of protected versus non-protected vaccinees and controls, the ESPY value has two properties: the absolute ESPY value of a Pf-specific antigen reflects how much a single Pf-specific antigen affected the prediction. Whereas, the *effect* label reflects the direction on the classification. A positive *effect* “+” denotes that the evaluated informative feature is more similar to the positive sample, and vice versa for a negative *effect* “-”.

Immunized protected individuals, who received the highest dose of PfSPZ-CVac, showed a higher antibody reactivity against CSP, PfEMP1, MSP2, MSP4, LSA1, conserved (membrane) proteins with unknown function, and intra-cellular/trans-membrane proteins, at post-immunization. The identified Pf-specific antigens were assessed to have a positive effect (red bars) on the classification of protected versus non-protected individuals (Fig 3A) by ESPY evaluation and are therefore more similar to protected vaccinees. The informative Pf-specific



antigens with the highest ESPY value were membrane proteins, conserved proteins of unknown function and intracellular proteins. This evaluation was done considering, the PfSPZ-dose as an informative feature with the highest ESPY value and a positive effect on the classification of protected vaccinees versus non-protected vaccinees and controls at both time points and datasets.

Zinc finger protein, PHISTb, PfEMP1, intracellular proteins, and proteins of unknown function were evaluated (by ESPY) to have a negative effect (blue bars) in the protection status classification after immunization (Fig 3A) in immunized individuals who received a lower dose or were located in the placebo group.

At pre-CHMI (Fig 3B), a smaller set of twelve Pf-specific antigens, namely CSP, LSA1, GLURP and proteins of unknown function were evaluated to have a positive effect on the classification of protected vaccinees versus non-protected vaccinees and controls, and showed a higher antibody signal intensity in immunized protected individuals. By contrast, all other informative Pf-specific antigens, like zinc finger protein, ETRAMP5, PfEMP1, and mostly conserved proteins of unknown function and intracellular proteins were evaluated to have a negative effect on the protection status classification, and showed a higher antibody signal intensity in non-protected vaccinees and controls. The top 50 informative Pf-specific antigens for the whole proteome microarray at post-immunization and pre-CHMI based on ESPY evaluation, are listed in Tables C and D in [S1 Appendix](#).

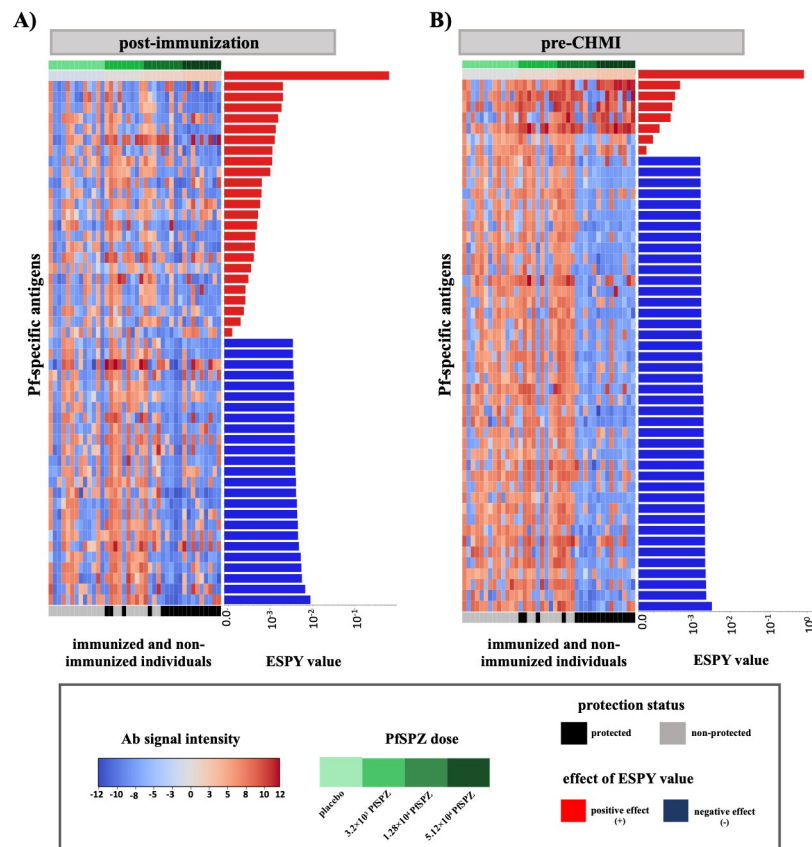
Third, we evaluated Pf-induced antibody profiles of all individuals using immunoreactivity to the pre-selected 188 Pf-specific cell surface antigens after removing strongly linearly correlated features above a Pearson correlation coefficient of  $pcc = 0.8$ . Fig 4 shows the antibody profiles (at post-immunization and pre-CHMI) for the top 50 informative Pf-specific cell-surface antigens based on ESPY evaluation of the multitask SVM model. At post-immunization (Fig 4A) MSP2, PfEMP1, rifin and membrane proteins of unknown function were evaluated to have a positive effect on the classification of protected vaccinees versus non-protected vaccinees and controls. By contrast, mainly zinc finger protein, PHISTb, PHISTc, PfEMP1 and conserved proteins of unknown function were evaluated to have a negative effect on the protection status classification.

At pre-CHMI, again a small set of seven informative Pf-specific antigens, namely CSP, MSP2, PfEMP1, LSA1 and (membrane) proteins of unknown function were evaluated to have a positive effect on the classification of protected vaccinees versus non-protected vaccinees and controls. All other Pf-specific antigens were evaluated to have a negative effect on the protection status classification. Those were mainly rifin, ETRAMP5, PHISTb, PHISTc, zinc finger proteins, (membrane) proteins of unknown function, PfEMP1 and MSP7, and showed a higher antibody signal intensity in unprotected vaccinees and the control group (Fig 4B).

Overall, the identified informative Pf-specific cell-surface antigens are well-known pre-erythrocytic and erythrocytic Pf-specific antigens, like CSP, ETRAMP, MSP, LSA, PfEMP1, PHISTb/c, rifin, zinc finger protein, and other Pf-specific antigens of unknown function. The top 50 informative Pf-specific antigens for the pre-selected cell-surface proteome microarray at post-immunization and pre-CHMI based on the ESPY value are listed in Tables E and F in [S1 Appendix](#).

### ESPY versus SHAP: Evaluation of informative features on simulated data

Fig 5A shows the informative features as identified by ESPY evaluation of a RBF-SVM trained on simulated data. As described in detail in the materials and methods section (Simulated data), the simulated data set consists of 500 samples and 1000 features, where 15 features are defined as informative features and the remaining ones as uninformative features. We used a SVM model with a RBF kernel to evaluate the ESPY values on simulated data. The RBF-SVM

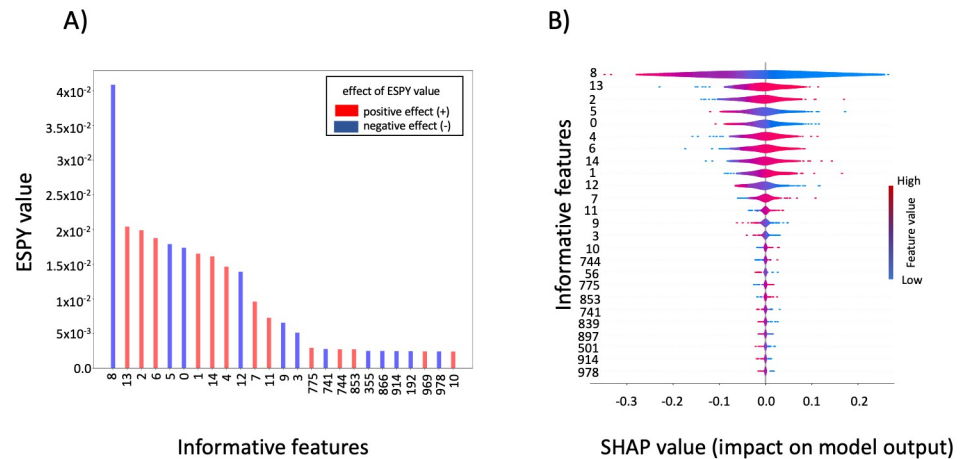


**Fig 4. Antibody profile of protected and non-protected vaccinees and the control group against informative cell-surface Pf-specific antigens.** Informative Pf-specific antigens against pre-selected Pf-specific cell-surface antigens were evaluated at post-immunisation and pre-CHMI. Pf-specific antigens identified to be important by ESPY evaluation showed either a high antibody signal intensity in protected vaccinees or unprotected vaccinees and controls. The top 50 Pf-specific antigens with the highest ESPY values are shown (A) at post-immunization and (B) at pre-CHMI. The heatmap plot shows the antibody signal intensity, while the bars on the right side of each figure show the importance and effect of each feature based on the ESPY value. ESPY values of Pf-specific antigens that were evaluated to have a positive effect on the protection status classification are colored in red, while blue-colored bars represent antigens that have a negative effect.

<https://doi.org/10.1371/journal.pcbi.1012131.g004>

model achieved the highest AUC score for values of the regularization parameter  $C = 10$  and the RBF kernel parameter  $\gamma = 0.001$  (AUC of 0.92 during a stratified 5-fold grid-search cross-validation on the training dataset, and an AUC of 0.81 on the hold-out test dataset). The aforementioned parameter combination was then used to train a RBF-SVM model on the training dataset, which was used for ESPY evaluation afterwards. Since the ESPY evaluation is based on the computation of the distances from the classification boundary (refer to the materials and methods section), the procedure for the evaluation of informative features is the same for a single-task or a multitask SVM.

Fig 5A shows the top 25 informative features and their respective ESPY values. The first 14 features have higher ESPY values and are thus sorted from the left to the right of the shown bar plot. They are by definition the informative features in the simulated data. Only the fifteenth feature (of the top 15 informative features in the simulated data), labeled by *feature10*, had lower ESPY values and thus were shifted to the end of the list of the top 25 informative



**Fig 5. Informative features selected using ESPY and SHAP values on the simulated data.** We show the top 25 features of the simulated data that were evaluated to be informative by ESPY and SHAP. A) The first 14 out of 15 informative features have higher ESPY values in comparison to rest of the features. For each feature the effect with the positive class is indicated by the color of the bars: a negative effect with the positive class (negative ESPY effect) is indicated by blue color and positive effect with the positive class (positive ESPY effect) is indicated by red color. B) Also for the SHAP evaluation, the first 14 out of 15 informative features have the highest mean SHAP values and are thus sorted to the top of the shown violin summary plot. The SHAP value and the color coded original feature value are used to indicate the change of each feature in the model prediction towards the positive or negative class. A high SHAP value indicates a change towards the positive class and vice versa for a low SHAP value.

<https://doi.org/10.1371/journal.pcbi.1012131.g005>

features. The remaining ten of 25 features had lower ESPY values and are by definition uninformative features of the simulated data.

We compared our ESPY method with the SHAP (SHapley Additive exPlanations) framework from Lundberg et al. [19]. Fig 5B shows the evaluated SHAP values. Again, the first 14 features have a higher mean SHAP value and are thus sorted to the top of the shown violin summary plot. Once again, the fifteenth feature (*feature10*) shows a lower mean SHAP value in comparison to the top 14 informative features. Comparing the class association (whether a feature is more similar to the positive class or vice versa), it can be seen in Fig 5 that, while the ranking by the absolute ESPY and SHAP values of several of the top 15 features is permuted, the evaluated associations of the individual top 15 features are similar. SHAP took more than 13 hours to evaluate the features on a Dell XPS 13 with Intel i7-10510U CPU, 4 Cores and 16GB of memory.

In comparison, ESPY evaluated the features in only 2.08 seconds on the same laptop. As mentioned by [20], a faster run-time of the SHAP framework could be achieved by clustering the background dataset (i.e., the training dataset) used to evaluate SHAP using k-means clustering with not too many cluster centers (e.g.,  $k = 50$ ). However, in scenarios as studied here, where the data has many more dimensions than samples (curse of dimensionality) ( $n \gg m$ ), k-means is unlikely to find a good clustering of the data. In summary, we could show that our ESPY method and the SHAP framework perform similarly on the task of informative feature, which reflects a real world dataset. However, the SHAP framework needed much more computation time compared to our approach.

## Discussion

In this study, we used a Pf-specific proteome microarray covering about 91% of the entire Pf proteome to identify PfSPZ-CVac-induced antibody profiles. As earlier suggested by Felgner

et al. and Trieu et al. [21, 22] and later shown by Obiero et al. and Wichers et al. [6, 23, 24], the immune response against human malaria is induced by a wide range of Pf-specific antigens. Here, we adapted supervised machine learning methods to identify predictive antibody profiles from immunized and non-immunized PfSPZ-CVac individuals.

Due to the large number of antibody signal intensities in comparison to the small number of individuals per time point, we set up a customized multitask SVM approach. Multitask SVM models are known to perform very well on prediction problems that profit from combining related data into a single model to increase the number of samples for the prediction task [14]. We compared our final multitask SVM approach with state-of-the-art machine learning approaches, namely regularized logistic regression (RLR), random forest (RF), and a single-task SVM model with a RBF kernel, after immunization and before CHMI. Overall, the RLR model, the single-task SVM model, and the RF approach, which were either trained on each single time point separately or on samples from both time points, showed a lower performance in comparison to our multitask SVM approach, where both time points were combined into a single prediction model.

Using our multitask SVM approach, we show how to profit from combining time- and dose-dependent data from multiple time points into a sole prediction model, whenever the number of samples at each single time point is small. Moreover, we show that our multitask SVM approach is less affected by highly correlated features over a range of Pearson correlation coefficients and achieves robust accuracy scores in comparison to the state-of-the-art approaches (Figs A and B in [S1 Appendix](#)). This might be of great interest in the biomedical context, where the number of available samples is often limited and strong correlations between features can be assumed. Here, in the analysis of a large proteome microarray ( $p = 7,455$ ) was analyzed, we assumed strong correlations due to e.g., similar epitopes, between fragments representing one protein, and cross-reactivity [15]. Furthermore, we excluded the antibody responses after CHMI (post-CHMI) from our malaria vaccine efficacy prediction analysis and our implementation of our multitask SVM approach due to the reason that the control group underwent a CHMI at this time point as well.

For the evaluation of informative Pf-specific antigens from the multitask SVM approach, we decided to only consider only the first top 50 Pf-specific antigens with the highest ESPY values per time point. Using ESPY evaluation, we estimated the contribution of each single Pf-specific antigen towards the classification of protected vaccinees and non-protected vaccinees and controls. In this analysis strongly linearly correlated features do not have the same influence on the evaluation procedure as they might have it in the RLR and RF approach. With this analysis we could show that PfSPZ-CVac immunized protected individuals react against a broad spectrum of known and unknown Pf-specific antigens, such as PfEMP1, CSP, MSP2/4, LSA1, GLURP conserved (membrane) proteins of unknown function, and intra-cellular/trans-membrane proteins after immunization and before CHMI.

The antibody breadth of PfSPZ-CVac immunized individuals varied based on the received PfSPZ-CVac dose. The number of identified Pf-specific antigens showed an overall, medium to high antibody reactivity among individuals who received a low and medium dose of PfSPZ-CVac and in the control group. Assuming that this result is not induced due to a high background noise, we hypothesize that these antigens are recognized by humoral immunity of individuals that were never in contact with human malaria before. This might be explained through cross-reactivity. Murugan et al. [25] supports the idea that the repertoire before immunization is important for the generation of high affinity antibodies (in this case anti-CSP). Further analysis of the antibody profile before vaccination of those individuals who underwent a successful immunization is needed to determine whether a high abundance of antibodies correlates with a higher chance of vaccination. Protected individuals who received

the highest dose of PfSPZ-CVac showed higher antibody reactivity against a comparatively small set of Pf-specific cell-surface antigens of the pre-erythrocytic and erythrocytic stage, like CSP, PfEMP1, MSP2/4, LSA1, GLURP and conserved Pf (membrane) proteins with unknown function. In agreement with these findings, Mordmüller et al. [8] showed that LSA1, MSP4, GLURP and conserved Pf proteins with unknown function were recognized by more than half of the protected individuals in the high-dose group, whereas, PfEMP1 was not recognized by at least 5/9 of the protected individuals in the high-dose group. Further, Mordmüller et al. [8] showed by ELISA that all of those protected individuals strongly reacted against CSP. Our analysis of Pf-specific antigen identification evaluated most of the informative proteins found by Mordmüller et al. [8], as well as new ones that seem to improve the prediction performance. Antibodies against CSP have significant functional activity in the protection against human malaria [8, 26], and CSP is a dominant antigen in the early (pre-erythrocytic) phase of the infection. However, the most advanced malaria vaccine candidates, RTS,S (Mosquirix) and R21/Matrix-M, confer only limited and short-lived protection against clinical malaria in the former one [27–29] and in the latter one the R21/Matrix-M candidate achieved the WHO-stated vaccine efficacy goal over 75% [30, 31] against Pf clinical malaria but only in one specific malaria endemic area. We hypothesize that sterile protection against human malaria induced through PfSPZ-CVac is not alone conferred by the humoral immunity, but rather is the result of both humoral and cellular immune responses to a number of different antigens. Recent findings from many clinical studies [7, 32–35], administering chemoattenuated PfSPZ or irradiated sporozoites, reproducibly confirm that PfSPZ-based vaccines induce an increase in cellular immune responses. CD4 T cells, CD8 T cells and  $\gamma\delta$  T cells are supposed to be primary effectors in the elimination of parasite-infected hepatocytes. Especially memory CD8 T cells are associated with the direct killing of infected hepatocytes; at least in animal models of malaria [36]. However, these processes occur in the liver and are therefore difficult to study in humans. Malaria vaccine candidates, such as RTS,S (Mosquirix) or R21/Matrix-M, target one major surface protein and thereby are not able to induce highly potent cellular immune responses against infected hepatocytes. Thus, the additive effect of antibodies in the ensemble of a cellular immune response against sporozoites is likely to lead to robust sterile protection and is yet missing in single protein based vaccine attempts. Further analysis of T- and B-cell mediated immune response against human malaria is hence needed for the prediction of a successful immunization.

A clear limitation of this study is that proteins on the microarray are not in their natural context and may have different conformation and post-translational modifications [37]. Furthermore, the microarray chip contains different concentrations of Pf-specific antigens per spot resulting in the problem of optimal quantification of informative Pf-specific antigens. This might also account for the relatively small number of identified Pf-specific antigens that were related to the protection status. In addition, the biological life cycle of the parasite is complex and the metabolically active parasite resides mostly intra-cellularly in hepatocytes and erythrocytes, with only transient extracellular phases. Therefore, the response is likely to mainly target cell-surface expressed antigens of merozoites, the extracellular form of Pf, and parasite-antigens that are presented by the infected cell. Importantly, the small number of identified Pf-specific antigens were sufficient to discriminate the protection status of protected vaccinees and non-protected vaccinees and controls. In general, all methods we have used are limited, if the antibody profiles are not similar between subjects. Here, the antibody breadth might be more important than a specific pattern favoring sterile protection against human malaria.

## Conclusion

The large number of more than 5,300 genes that are expressed during the life cycle of Pf, may explain why we still do not know which antigens are central to the induction of a sterile protection against human malaria. Proteome microarrays enabled us to measure antibody reactivity against Pf-specific antigens representing about 91% of the Pf proteome. Felgner et al., Obiero et al., and Illingworth et al. [6, 21, 38] earlier suggested that the immune response against human malaria is induced by a wide range of these proteins, yet only a few of these proteins have been clinically tested as a malaria vaccine. In medical trial studies, where new anti-malaria vaccine candidates are tested, the number of samples is often restricted to a small size. To overcome this problem, here we combined time- and dose-dependent data of PfSPZ antibody profiles of immunized and non-immunized individuals from multiple time-points into one sole prediction model. This approach is beneficial, since clinical vaccine studies are usually limited in their number of samples. Additionally, we proposed the new ESPY method to explain predictions from a non-linear SVM model. We could show, on simulated data, that ESPY evaluation can identify all informative features and provides explanations comparable to the SHAP framework for kernel SVMs. We successfully applied the ESPY method to find informative Pf-specific antigens for the prediction of protected and non-protected PfSPZ-CVac vaccinees and controls based on their antibody reactivity profiles. Our findings might help to extend the knowledge about Pf-specific antigens that induce B-cell activation. However, to fully understand the immune response against human malaria, a further step will be to include T-cell activation and RNA-seq data.

## Materials and methods

### Ethics statement

The study was approved by the ethics committee of the medical faculty and the university clinics of the University of Tübingen (project number 537/2013AMG1). All trial participants were thoroughly informed and gave written informed consent before any study procedure, recording of data or analysis was carried out.

### Data

**Proteome microarray.** We analyzed the Pf-specific antibody reactivity profile from the PfSPZ-CVac clinical trial TÜCHMI-002 (ClinicalTrials.gov Identifier: NCT02115516), which has been previously described [8]. In brief, Pf-specific antibody-mediated response profiles of 40 malaria-naïve individuals, vaccinated three times over 8 weeks with placebo (normal saline) or different doses of PfSPZ ( $3.2 \times 10^3$  PfSPZ,  $1.28 \times 10^4$  PfSPZ,  $5.12 \times 10^4$  PfSPZ) by direct venous inoculation under chloroquine chemoprophylaxis (Sanaria PfSPZ-CVac), were measured at four different time points: before vaccination (I-1), following the third vaccination (III+14), one day before CHMI (C-1), and four weeks after CHMI (C+28). The proteome microarray contains 7,455 Pf-specific protein fragments representing 4,805 unique Pf genes of the NF54 Pf strain. This resulted in a dataset containing 40 individuals per time point with Pf-specific antibody intensity signals from 7,455 Pf-specific fragments. At each dosage, 9 individuals were vaccinated (for a total of 27 across 3 PfSPZ-CVac doses) and 13 individuals were allocated to the placebo group. For each individual, the protection status is defined by the primary efficacy endpoint as described earlier by Mordmüller et al. [8]. An individual was considered not protected against malaria, if any parasitemia was detected by thick blood smear, and protected, if no parasite was detected within 21 days following CHMI. All parasitemic volunteers were treated promptly with a highly active antimalarial. The data underlying this study

were obtained from proteome microarray data as described by Mordmüller et al. [8]: “Raw spot and local background fluorescence intensities, spot annotations and sample phenotypes were imported and merged in R, where all subsequent procedures were performed. Fore-ground spot intensities were adjusted by local background by subtraction, [ . . . ].” Baseline (I-1) antibody responses of immunized and control individuals were subtracted from the data generated post-immunization (III+14 and C-1) for each individual to focus on PfSPZ-CVac induced antibody responses [6]. Antibody responses after CHMI (C+28) were excluded from our malaria vaccine efficacy prediction analysis, because also controls underwent CHMI at this time point. Thus, after CHMI, there are no samples for the unprotected class anymore and, therefore, applying binary classification models will not be feasible. The resulting dataset contains 80 samples associated to 40 patients at two different time points (III+14 and C-1). Subsequently, the Pf-specific antibody signal intensities were arcsine transformed.

In a second step, we defined a set of cell surface Pf-antigen fragments from the whole set of Pf-specific fragments of the proteome microarray. Cell surface Pf-antigen fragments and Pf-antigen fragments of uncharacterized proteins ( $m = 1,194$ ) were selected from the proteome microarray supplement information based on their protein name/description, representing extracellular/membrane and uncharacterized proteins. In this subset, we assumed that responses to cell surface antigens are over-represented compared to intra-cellular antigens.

**Simulated data.** To evaluate how our feature importance measurement performs, we used python and `sklearn.datasets.make_classification(n_samples = 500, n_features = 1000, n_informative = 15, n_redundant = 0, n_repeated = 0, n_classes = 2, random_state = 42)` from the *Scikit-learn* [39] package to generate a random two-class classification problem. Said method creates clusters of points that are standard normally distributed about vertices of an 15-dimensional hyper-cube and assigns an equal number of clusters to each class. The resulting data matrix  $X \in \mathbb{R}^{500 \times 1000}$  consists of 15 informative feature columns, while the remaining feature columns are filled with random noise.

## Prediction models for time-series data

Our proposed approach for identifying Pf-specific immune signatures from protected and non-protected individuals tackles two problems. First, identifying appropriate machine learning models that are able both to deal with high-dimensional data and can learn from time and dose dependent data. Second, identifying informative Pf-specific antigens between protected and non-protected individuals.

Since individuals were exposed to different PfSPZ doses in the PfSPZ-CVac clinical trial at two (actually three, but we excluded the third time point, due to the reasons described in section Proteome microarray) consecutive time points, we built different prediction models to analyze the dependent time-series data. All models are binary classifiers that predict the protection state of the Pf-specific proteome microarray-based on antibody reactivity profile.

In a first scenario, the RLR [11], RF, and single-task RBF-SVM models were trained separately for both time-points. For this purpose, the dataset was split into two smaller “single-time” datasets corresponding to the time points III+14 and C-1, respectively. Each dataset contains antibody intensity signals for all PfSPZ doses ( $3.2 \times 10^3$  PfSPZ,  $1.28 \times 10^4$  PfSPZ,  $5.12 \times 10^4$  PfSPZ), associated to a single time point, augmented by the PfSPZ dose as auxiliary information.

Tasks, that are related to each other, can be used in a prediction model simultaneously (so-called *multitask-learning*) [40, 41]. Considering the small number of samples per time point and the very large amount of features (sample size  $N_{\text{patients}} = 40$ , number of features  $P = 7,455$ ), establishing a prediction model that uses related samples from additional time points is

more promising than training independent models with data for the specific task only. Therefore, in a second scenario, we cast the prediction problem into a multitask learning problem and treat prediction based on time point, PfSPZ dose, and antibody intensity signal data as separate tasks. To implement the multitask approach, we represented the relationships between individuals for each task (time point, PfSPZ dose, antibody intensity signals) by a separate feature matrix. Within the multitask approach we were able to classify individuals into protected and non-protected using the measured immune profile of all 40 individuals for each time point (III+14 and C-1) and dose ( $3.2 \times 10^3$  PfSPZ,  $1.28 \times 10^4$  PfSPZ,  $5.12 \times 10^4$  PfSPZ) at once in one model resulting in 80 samples.

In a third scenario, to accomplish a fair comparison between our multitask SVM approach and the other methods, namely RLR, RF and RBF-SVM, we additionally trained those on the original “multi-time” dataset, containing 80 samples associated to 40 patients at two different time points (III+14 and C-1) augmented by the PfSPZ dose. For details please refer to the subsequent section Prediction performance assessment.

### A multitask SVM approach for time- and dose-dependent proteome data

Kernels of SVMs can be used to model relationships between single related tasks and combine them into one prediction model [42, 43]. Combining such small single datasets, that are related to each other, can lead to an improvement in classification. Kernel-based multitask learning can be achieved by the element-wise product of two kernel matrices. According to the *Schur product theorem* [44], the Hadamard (element-wise) product of two positive (semi-)definite matrices is also a positive (semi-)definite matrix. For the measured immune profile, an antibody signal intensity matrix  $X_{\text{antibody}} \in \mathbb{R}^{N \times P}$  is given, where  $N$  is the number of samples ( $N = 2 \times N_{\text{patients}}$ ) and  $P$  the number of features. The value at entry  $x_{np}$  is the detected antibody signal intensity of feature  $p$  for sample  $n$ . Additionally, two vectors are given: the time-series vector  $\gamma_t \in \mathbb{R}^N$ , where  $t$  represents the date of antibody profile collection, and the dose vector  $\gamma_d \in \mathbb{R}^N$ , where  $d$  represents the dose of PfSPZ. To simulate the clinical trial of PfSPZ-CVac over the two time points, we used a radial basis function (RBF) kernel to represent the relationship between the individuals based on time points, resulting in the kernel matrix  $K_1(n_b, n_t)$ . As a representation of the relationship between the individuals based on the administered PfSPZ dose, we used either an RBF or a polynomial kernel function, resulting in a kernel matrix  $K_2(n_d, n_{d'})$ . Finally, the relationship between the individuals based on antibody signal intensities was represented by either an RBF or a polynomial kernel function, resulting in a kernel matrix  $K_3(n_p, n_{p'})$ . The resulting kernel matrices were combined by element-wise multiplication:

$$K_{\text{multitask}}((n_t, n_d, n_p), (n_{t'}, n_{d'}, n_{p'})) := K_1(n_t, n_{t'}) \times K_2(n_d, n_{d'}) \times K_3(n_p, n_{p'}) \quad (1)$$

where  $K_{\text{multitask}}$  is a positive semi-definite (psd) kernel matrix, containing all feature intensities signals for two time points (III+14 and C-1) and all PfSPZ doses.

### Adapted spectral translation approach

The *Schur product theorem* [44] states that the Hadamard product of two psd matrices is psd as well. In practice, due to numerical issues arising from limited machine precision, the multitask kernel matrix  $K_{\text{multitask}}$ , resulting from the element-wise multiplication of the single-task kernel matrices, might be slightly disturbed and not psd within numerical precision. To deal with this issue, we implemented the following procedure.



After element-wise multiplication of the single-task kernel matrices, we applied a variant of the so-called *spectral translation approach* [45]. Originally, as described by Vert et al. [45], the smallest negative eigenvalue is subtracted from (i.e., its absolute value is added to) the diagonal of a non-psd symmetric matrix, resulting in a psd kernel matrix. This approach exploits the fact that a psd matrix is symmetric and all its eigenvalues are non-negative per definition. By adding the absolute value of the smallest negative eigenvalue to its diagonal, the spectrum of the non-psd matrix is effectively shifted upwards and, thus, the matrix becomes psd. Since, in practice, again due to numerical instabilities, this approach sometimes does not result in a psd matrix, we adapted it into an iterative procedure as follows: We computed the real eigenvalues  $e_i$  of  $K_{\text{multitask}}$  and tested if any of them were negative. If negative eigenvalues were found,  $\max(\{\tilde{\epsilon}, \min_i(e_i)\})$  was added to the diagonal of  $K_{\text{multitask}}$ , with  $\tilde{\epsilon} = 10^3 \cdot \epsilon$  and  $\epsilon$  being the machine precision. This approach was repeated for a maximum of 1000 iterations (if the matrix was not psd after 1000 iterations, a value of 1.0 was added to the diagonal and the resulting matrix was used) or until no further negative eigenvalues were found, thus resulting in a psd multitask kernel matrix  $K_{\text{multitask}}$ . Based on the multitask kernel matrix, we trained and tested different settings of parameters to create a predictive model.

### Dealing with strongly linearly correlated features

To better understand and predict a successful vaccination, identification and interpretation of single Pf-specific antigens is important. In the underlying dataset we are dealing with a large number of features compared to a small number of samples. Further, strong correlations between the features can be assumed due to, e.g., several fragments representing a single protein, similar epitopes, and cross-reactivity [15]. Reducing the number of irrelevant and uninformative features can increase performance [13, 16, 17]. As earlier shown by Valletta et al. [13], a Pearson correlation coefficient of  $pcc = 0.8$  seems to be a reasonable threshold for removing strongly linearly correlated features of such proteome microarray data. Nonetheless, we assessed the influence of correlated features on the prediction performance of the studied machine learning models, by removing features that were linearly correlated above a certain Pearson correlation threshold  $pcc$ , varying the threshold between 0.1 and 1.0 in steps of 0.1. The results of this analysis are visualized in supplementary figures Figs A and B in [S1 Appendix](#). For both datasets, the whole proteome microarray (Fig A in [S1 Appendix](#)) and the pre-selected cell surface microarray (Fig B in [S1 Appendix](#)), we can show, that the prediction performance of the state-of-the-art methods is more influenced by strongly linearly correlated features than our multitask SVM approach with a kernel combination of either 'RRR' or 'RPR'. Furthermore, the Pearson correlation coefficient of  $pcc = 0.8$  seems to be also a reasonable threshold for our datasets. Therefore, we removed strongly linearly correlated features above a Pearson correlation coefficient of  $pcc = 0.8$ , which corresponds to around 8% of the original whole proteome microarray and to around 16% of the original pre-selected cell surface microarray.

### Prediction performance assessment

To assess the prediction performance of the studied machine learning models, we proceeded as follows. We compared multitask SVM models with single- and multi-time single-task RBF-SVM, RF, and RLR models.

For the RLR models we used the so-called *elastic net* regularization as implemented in the `sklearn.linear_model.LogisticRegression` class of the *Scikit-learn* [46] package. Elastic net RLR uses a combination of  $\ell_1$  and  $\ell_2$ - regularization, where  $\rho$  defines the compromise between

$\ell_1$  ( $\rho = 1$ ) and  $\ell_2$  ( $\rho = 0$ ) penalty, and minimizes the following cost function:

$$\min_{w,c} \left( \frac{1-\rho}{2} w^T w + \rho \|w\|_1 + C \sum_{i=1}^N \log(\exp(-y_i(X_i^T w + c)) + 1) \right) \quad (2)$$

We tuned and assessed the single- and multi-time RLR models during a 10-times repeated nested stratified 5-fold cross-validation, as described by Krstajic (Algorithm 2 in [47]) and implemented in the *RepeatedStratifiedNestedCV* class of the *nestedcv* [48] package, over a grid of  $\rho$  and the regularization parameter  $C$  (Table A in S1 Appendix). The single- and multi-time single-task RBF-SVM models (using the C-Support Vector Classification method in *sklearn.svm.SVC*) were tuned and assessed likewise by performing a 10-times repeated nested stratified 5-fold cross-validation over a grid of the RBF kernel coefficient  $\gamma$  and regularization parameter  $C$  (Table A in S1 Appendix). In case of single- and multi-time RF models, we used the *RandomForestClassifier* implemented in *sklearn.ensemble* and assessed its performance likewise by executing a 10-times repeated nested stratified 5-fold cross-validation over a number of parameters, namely the number of trees  $n_{\text{trees}}$  and the maximal number of features  $\text{max\_features}$  per split (Table A in S1 Appendix). The multitask SVM models were deployed by pre-computing the multitask kernel matrices as described above in section A multitask SVM approach for time- and dose-dependent proteome data for different parameter combinations (listed in Table A in S1 Appendix) and feeding the pre-computed kernel matrices into the C-Support Vector Classification method implemented in *sklearn.svm.SVC*. The models were assessed and optimized over a parameter grid (listed in Table A in S1 Appendix) using 10-times repeated nested stratified 5-fold cross-validation utilizing *nestedcv.RepeatedStratifiedNestedCV*. All models were optimized and evaluated on exactly the same cross-validation folds. In case of multitask SVM models and multi-time RF, RLR, and single-task RBF-SVM models, test sets (aka test “folds” in the cross-validation context) were always constituted of samples from a certain single time point (III+14 or C-1), while the associated train folds were constituted from samples from both time points (III+14 and C-1). It was ensured that all samples taken from patients that were represented by a sample included in a test fold were excluded from the associated disjoint train fold. Model performances were measured using the *Area Under the Precision Recall Curve* (PR-AUC) where a higher PR-AUC score equates to perfect prediction and a lower PR-AUC score to random guessing. Precision recall curves are commonly used to evaluate the prediction performance of a model with imbalanced data, that is, when the number of samples of one class (e.g. class 0) is much higher than the number of samples of the other class (e.g. class 1) [49]. To assess the prediction performance of the underlying models, namely RLR, single-task RBF-SVM, RF and multitask SVM models, we performed 10 repetitions of nested stratified 5-fold cross-validation for each model, as described in detail above. To evaluate the PR-AUC, the predictions on the test sets of the models trained on the train folds were collected and used to calculate a nested cross-validation PR-AUC score for every repetition as described by Krstajic (Algorithm 2 in [47]). The final nested cross-validation score is then calculated as the average of the resulting scores of each repetition.

### Tuning models for feature evaluation

Before using the multitask SVM approach to evaluate informative features, we had to optimize the parameters. The multitask SVM models were deployed as described in the previous section. They were tuned over a parameter grid (listed in Table A in S1 Appendix) using 10-times repeated stratified 5-fold grid-search cross-validation, as described by Krstajic (Algorithm 1 in [47]), utilizing *nestedcv.RepeatedGridSearchCV*. Again, all models were optimized and evaluated on exactly the same cross-validation folds. In the case of multitask SVM models, test folds

always comprised of samples from a certain single time point (III+14 or C-1), while the associated train folds comprised samples of both time points (III+14 and C-1). It was ensured that all samples taken from patients that were represented by a sample included in a test fold were excluded from the associated disjoint train fold.

The obtained parameters were then used to predict the mean PR-AUC performance on the hold-out test dataset for the whole proteome microarray and the pre-selected cell-surface proteome microarray and are listed in Table B in [S1 Appendix](#). Afterwards, the trained model (either trained on the whole proteome microarray or on the pre-selected cell-surface proteome microarray) was used to evaluate the ESPY value (as explained in detail in the following section ESPY (fEature diStance exPlainability)) for each single feature on the test data.

### ESPY (fEature diStance exPlainability)

We propose ESPY values as a measure of importance of each single feature based on a (multi-task) SVM model. The method is inspired by a feature importance measure for sequence-based non-linear predictions proposed by Pfeifer and Lengauer [18]. ESPY uses systematically and specifically triggered changes in the distance of a consensus sample to the classification boundary of the SVM to estimate the importance of features as described below. The distance to the classification boundary can be computed for a sample  $\vec{x}$ , the optimized parameter  $\alpha \in \mathbb{R}^n$ , offset  $b$ , and class labels  $y_i \in [0, 1]$ :

$$d(\vec{x}) = \sum_{i=1}^n y_i (\alpha_i k(\vec{x}, \vec{x}_i) + b) \tag{3}$$

The idea of the approach is to determine if a change in the antibody intensity signal of each feature at a time results in a distance change towards the negative or positive side of the classification boundary. If the signal leads to a change to the positive side, the antibody response is more similar to the positive samples and vice versa. To compare the change in the antibody intensity signal of each feature at a time, a consensus sample is generated. The consensus sample  $\vec{x}^{Q_2}$  is element-wise defined as the median (alias second quartile  $Q_2$ ) antibody intensity signal  $x_j^{Q_2} = \text{median}(X_j)$  over all samples per feature  $X_j$ .  $X_j$  is the  $j$ -th column of the data matrix  $X \in \mathbb{R}^{N \times M}$ , which was constructed by appending the time point vector  $\gamma_t \in \mathbb{R}^N$  and the dose vector  $\gamma_d \in \mathbb{R}^N$  as columns to the antibody signal intensity matrix  $X_{\text{antibody}} \in \mathbb{R}^{N \times P}$ . To evaluate the ESPY value and therefore the importance of each feature, we measured the change in distance when varying the consensus at a certain feature  $j$  to the first quartile  $Q_1$  (25% percentile) and third quartile  $Q_3$  (75% percentile), respectively:

$$d_j^{Q_1} = d(\vec{x}_j^{Q_1}) - d(\vec{x}^{Q_2}) \tag{4}$$

$$d_j^{Q_3} = d(\vec{x}_j^{Q_3}) - d(\vec{x}^{Q_2}) \tag{5}$$

where  $\vec{x}_j^{Q_p} := (x_1^{Q_p}, x_2^{Q_p}, \dots, x_{j-1}^{Q_p}, x_j^{Q_p}, x_{j+1}^{Q_p}, \dots, x_m^{Q_p})^T$  and  $x_j^{Q_p}$  is the  $p$ th quartile of  $X_j$ . Based on this,

we define the ESPY value per feature  $j$  as:

$$I_{\text{ESPY},j} := |d_j^{Q_1}| + |d_j^{Q_3}|, \quad (6)$$

$$M_{\text{ESPY},j} := \begin{cases} + & , & d_j^{Q_3} \geq 0 \wedge d_j^{Q_1} \leq 0 \wedge d_j^{Q_3} \neq d_j^{Q_1} \\ - & , & d_j^{Q_3} \leq 0 \wedge d_j^{Q_1} \geq 0 \wedge d_j^{Q_3} \neq d_j^{Q_1} \\ 0 & , & d_j^{Q_3} = d_j^{Q_1} = 0 \\ \text{NAN} & , & \text{else} \end{cases}, \quad (7)$$

with  $+$ ,  $-$ ,  $0$ , and  $\text{NAN}$  denoting a *positive effect*, *negative effect*, *no effect*, and an *undefined effect* respectively. Finally, the ESPY value per feature  $j$  is normalized by dividing each ESPY value by the sum of the ESPY values of all features.

**ESPY evaluation on simulated data.** To evaluate the ESPY values of each feature in the simulated dataset based on 7, we trained an RBF-SVM model. At first, the simulated dataset was split into training and test sets (70% for training, 30% for testing). This was followed by hyper-parameter tuning using stratified 5-fold grid-search cross-validation (as implemented in *sklearn.model\_selection.GridSearchCV*) over a grid of the RBF  $\gamma$  parameter and the regularization parameter  $C$  (see Table A in S1 Appendix) on the training data, to evaluate the best parameter combination for the RBF-SVM model. The obtained parameters were then used to predict the AUC performance on the hold-out test dataset. Afterwards, the trained model was used to evaluate the ESPY value for each single feature on the test data.

We compared our explanation and feature extraction method, namely ESPY, with the SHapley Additive exPlanations (SHAP) method [19], a game-theoretic approach for making any machine learning model interpretable. Particularly, we applied the *shap.KernelExplainer* class of the *shap* [50] package on the trained RBF-SVM model and evaluated the features with the highest importance scores.

## Supporting information

**S1 Appendix. This PDF file compiles supplementary information of this paper.**  
(PDF)

## Acknowledgments

The authors thank Nicolas Kersten and Freia-Raphaella Lorenz for helpful remarks and discussions during the course of this work. The clinical trial was funded by the German Federal Ministry of Education and Research (BMBF) through the German Center for Infection Research (DZIF) (DZIF TTU 03.802). Proteome microarray data was retrieved from Mordmüller, B., Surat, G., Lagler, H. et al. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature* 542, 445–449 (2017). <https://doi.org/10.1038/nature21060> for which the following statements hold: The manufacture of the investigational product by Sanaria was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under SBIR award numbers 5R44AI058375 and 5R44AI055229. Proteome microarray manufacture and experiments from which the data was derived were supported by NIAID SBIR grant 5R44AI066791 and funding from the Bill & Melinda Gates Foundation.

## Author Contributions

**Conceptualization:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter, Rolf Fendel, Benjamin Mordmüller, Nico Pfeifer.

**Data curation:** Jacqueline Wistuba-Hamprecht, Rolf Fendel, Stephen L. Hoffman, Joseph J. Campo, Philip L. Felgner, Peter G. Kremsner, Benjamin Mordmüller.

**Formal analysis:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter, Nico Pfeifer.

**Funding acquisition:** Nico Pfeifer.

**Investigation:** Peter G. Kremsner, Benjamin Mordmüller.

**Methodology:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter, Nico Pfeifer.

**Project administration:** Rolf Fendel, Benjamin Mordmüller, Nico Pfeifer.

**Resources:** Stephen L. Hoffman, Joseph J. Campo, Peter G. Kremsner, Benjamin Mordmüller.

**Software:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter.

**Supervision:** Bernhard Reuter, Benjamin Mordmüller, Nico Pfeifer.

**Validation:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter.

**Visualization:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter.

**Writing – original draft:** Jacqueline Wistuba-Hamprecht, Rolf Fendel, Benjamin Mordmüller, Nico Pfeifer.

**Writing – review & editing:** Jacqueline Wistuba-Hamprecht, Bernhard Reuter, Rolf Fendel, Stephen L. Hoffman, Joseph J. Campo, Philip L. Felgner, Peter G. Kremsner, Benjamin Mordmüller, Nico Pfeifer.

## References

1. WHO. World malaria report 2023. WHO. 2023;.
2. Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature medicine*. 1998; 4(3):358–360. <https://doi.org/10.1038/nm0398-358> PMID: 9500614
3. Doolan DL, Dobaño C, Baird JK. Acquired immunity to malaria. *Clinical microbiology reviews*. 2009; 22(1):13–36. <https://doi.org/10.1128/CMR.00025-08> PMID: 19136431
4. Wu HM. Evaluation of the sick returned traveler. In: *Seminars in diagnostic pathology*. vol. 36. Elsevier; 2019. p. 197–202.
5. CLYDE DF, MOST H, McCarthy V, Vanderberg J, et al. Immunization of man against sporozoite-induced falciparum malaria. *American journal of medical sciences*. 1973; 266(3):169–77.
6. Obiero JM, Campo JJ, Scholzen A, Randall A, Bijker EM, Roestenberg M, et al. Antibody Biomarkers Associated with Sterile Protection Induced by Controlled Human Malaria Infection under Chloroquine Prophylaxis. *MSphere*. 2019; 4(1):e00027–19. <https://doi.org/10.1128/mSphereDirect.00027-19> PMID: 30787114
7. Seder RA, Chang LJ, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, et al. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science*. 2013; 341(6152):1359–1365. <https://doi.org/10.1126/science.1241800> PMID: 23929949
8. Mordmüller B, Surat G, Lagler H, Chakravarty S, Ishizuka AS, Lalremruata A, et al. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature*. 2017; 542(7642):445–449. <https://doi.org/10.1038/nature21060> PMID: 28199305
9. Mwakingwe-Omari A, Healy SA, Lane J, Cook DM, Kalhori S, Wyatt C, et al. Two chemoattenuated PfSPZ malaria vaccines induce sterile hepatic immunity. *Nature*. 2021; 595(7866):289–294. <https://doi.org/10.1038/s41586-021-03684-z> PMID: 34194041

10. Sulyok Z, Fendel R, Eder B, Lorenz FR, Kc N, Karnahl M, et al. Heterologous protection against malaria by a simple chemoattenuated PfSPZ vaccine regimen in a randomized trial. *Nature communications*. 2021; 12(1):1–10. <https://doi.org/10.1038/s41467-021-22740-w> PMID: 33947856
11. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005; 67(2):301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
12. Finney OC, Danziger SA, Molina DM, Vignali M, Takagi A, Ji M, et al. Predicting antidiarrhoeal immunity using proteome arrays and sera from children naturally exposed to malaria. *Molecular & Cellular Proteomics*. 2014; 13(10):2646–2660. <https://doi.org/10.1074/mcp.M113.036632> PMID: 25023128
13. Valletta JJ, Recker M. Identification of immune signatures predictive of clinical protection from malaria. *PLoS computational biology*. 2017; 13(10):e1005812. <https://doi.org/10.1371/journal.pcbi.1005812> PMID: 29065113
14. Zhang Y, Yang Q. A survey on multi-task learning. arXiv preprint arXiv:170708114. 2017;.
15. Muellenbeck MF, Ueberheide B, Amulic B, Epp A, Fenyo D, Busse CE, et al. Atypical and classical memory B cells produce *Plasmodium falciparum* neutralizing antibodies. *Journal of Experimental Medicine*. 2013; 210(2):389–399. <https://doi.org/10.1084/jem.20121970> PMID: 23319701
16. Bolón-Canedo V, Sánchez-Maróñ A, Alonso-Betanzos A. On the effectiveness of discretization on gene selection of microarray data. In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*. IEEE; 2010. p. 1–8.
17. Inza I, Larranaga P, Blanco R, Cerrolaza AJ. Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial intelligence in medicine*. 2004; 31(2):91–103. <https://doi.org/10.1016/j.artmed.2004.01.007> PMID: 15219288
18. Pfeifer N, Lengauer T. Improving HIV coreceptor usage prediction in the clinic using hints from next-generation sequencing data. *Bioinformatics*. 2012; 28(18):i589–i595. <https://doi.org/10.1093/bioinformatics/bts373> PMID: 22962486
19. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st international conference on neural information processing systems*; 2017. p. 4768–4777.
20. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc.; 2017. p. 4765–4774. Available from: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
21. Felgner PL, Roestenberg M, Liang L, Hung C, Jain A, Pablo J, et al. Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Scientific reports*. 2013; 3:3549. <https://doi.org/10.1038/srep03549> PMID: 24351974
22. Trieu A, Kayala MA, Burk C, Molina DM, Freilich DA, Richie TL, et al. Sterile protective immunity to malaria is associated with a panel of novel *P. falciparum* antigens. *Molecular & Cellular Proteomics*. 2011; 10(9):M111–007948.
23. Wichers JS, Tonkin-Hill G, Thye T, Krumkamp R, Kreuels B, Strauss J, et al. Common virulence gene expression in adult first-time infected malaria patients and severe cases. *Elife*. 2021; 10:e69040. <https://doi.org/10.7554/eLife.69040> PMID: 33908865
24. de Jong SE, van Unen V, Manurung MD, Stam KA, Goeman JJ, Jochems SP, et al. Systems analysis and controlled malaria infection in Europeans and Africans elucidate naturally acquired immunity. *Nature Immunology*. 2021; 22(5):654–665. <https://doi.org/10.1038/s41590-021-00911-7> PMID: 33888898
25. Murugan R, Buchauer L, Triller G, Kreschel C, Costa G, Martí GP, et al. Clonal selection drives protective memory B cell responses in controlled human malaria infection. *Science immunology*. 2018; 3(20):eaap8029. <https://doi.org/10.1126/sciimmunol.aap8029> PMID: 29453292
26. Murphy SC, Deye GA, Sim BKL, Galbiati S, Kennedy JK, Cohen KW, et al. PfSPZ-CVac efficacy against malaria increases from 0% to 75% when administered in the absence of erythrocyte stage parasitemia: A randomized, placebo-controlled trial with controlled human malaria infection. *PLoS pathogens*. 2021; 17(5):e1009594. <https://doi.org/10.1371/journal.ppat.1009594> PMID: 34048504
27. RTS SCTP. A phase 3 trial of RTS, S/AS01 malaria vaccine in African infants. *New England Journal of Medicine*. 2012; 367(24):2284–2295. <https://doi.org/10.1056/NEJMoa1208394>
28. RTS SCTP. Efficacy and safety of the RTS, S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites. *PLoS medicine*. 2014; 11(7):e1001685. <https://doi.org/10.1371/journal.pmed.1001685>
29. Rts S. 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. *The Lancet*. 2015; 386(9988):31–45. [https://doi.org/10.1016/S0140-6736\(15\)60721-8](https://doi.org/10.1016/S0140-6736(15)60721-8)

30. 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. *The Lancet*. 2021; 397(10287):1809–1818. [https://doi.org/10.1016/S0140-6736\(21\)00943-0](https://doi.org/10.1016/S0140-6736(21)00943-0) PMID: 33964223
31. Dattoo MS, Natama HM, Somé A, Bellamy D, Traoré O, Rouamba T, et al. Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial. *The Lancet Infectious Diseases*. 2022; 22(12):1728–1736. [https://doi.org/10.1016/S1473-3099\(22\)00442-X](https://doi.org/10.1016/S1473-3099(22)00442-X) PMID: 36087586
32. Bijker EM, Bastiaens GJ, Teirlinck AC, van Gemert GJ, Graumans W, van de Vegte-Bolmer M, et al. Protection against malaria after immunization by chloroquine prophylaxis and sporozoites is mediated by preerythrocytic immunity. *Proceedings of the National Academy of Sciences*. 2013; 110(19):7862–7867. <https://doi.org/10.1073/pnas.1220360110> PMID: 23599283
33. Bijker EM, Teirlinck AC, Schats R, van Gemert GJ, van de Vegte-Bolmer M, van Lieshout L, et al. Cytotoxic markers associate with protection against malaria in human volunteers immunized with *Plasmodium falciparum* sporozoites. *The Journal of infectious diseases*. 2014; 210(10):1605–1615. <https://doi.org/10.1093/infdis/jiu293> PMID: 24872326
34. Lyke KE, Ishizuka AS, Berry AA, Chakravarty S, DeZure A, Enama ME, et al. Attenuated PfSPZ Vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection. *Proceedings of the National Academy of Sciences*. 2017; 114(10):2711–2716. <https://doi.org/10.1073/pnas.1615324114> PMID: 28223498
35. Roestenberg M, McCall M, Hopman J, Wiersma J, Luty AJ, van Gemert GJ, et al. Protection against a malaria challenge by sporozoite inoculation. *New England Journal of Medicine*. 2009; 361(5):468–477. <https://doi.org/10.1056/NEJMoa0805832> PMID: 19641203
36. Hoffman SL, Doolan DL. Malaria vaccines—targeting infected hepatocytes. *Nature medicine*. 2000; 6(11):1218–1219. <https://doi.org/10.1038/81315> PMID: 11062530
37. Vigil A, Davies DH, Felgner PL. Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future microbiology*. 2010; 5(2):241–251. <https://doi.org/10.2217/fmb.09.127> PMID: 20143947
38. Illingworth JJ, Alanine DG, Brown RE, Marshall J, Bartlett HE, Silk S, et al. Functional comparison of blood-stage *Plasmodium falciparum* malaria vaccine candidate antigens. *Frontiers in Immunology*. 2019; 10:1254. <https://doi.org/10.3389/fimmu.2019.01254> PMID: 31214195
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 12:2825–2830.
40. Caruana R. Multitask learning. In: *Learning to learn*. Springer; 1998. p. 95–133.
41. Ben-David S, Schuller R. Exploiting task relatedness for multiple task learning. In: *Learning Theory and Kernel Machines*. Springer; 2003. p. 567–580.
42. Jebara T. Multi-task feature and kernel selection for SVMs. In: *Proceedings of the twenty-first international conference on Machine learning*. ACM; 2004. p. 55.
43. Micchelli CA, Pontil M. Kernels for Multi-task Learning. In: *Advances in neural information processing systems*; 2005. p. 921–928.
44. Schur J. *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen*. De Gruyter. 1911;.
45. Vert JP, Saigo H, Akutsu T. Convolution and local alignment kernels. *Kernel methods in computational biology*. 2004; p. 131–154. <https://doi.org/10.7551/mitpress/4057.003.0009>
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. scikit-learn is a Python module for machine learning (Github), version 1.0.2; Available from: <https://github.com/scikit-learn/scikit-learn>.
47. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics*. 2014; 6(1):1–15. <https://doi.org/10.1186/1758-2946-6-10> PMID: 24678909
48. Reuter B. NestedCV: repeated stratified nested cross-validation. (Github);. <https://github.com/msmdev/NestedCV>.
49. Qi Q, Luo Y, Xu Z, Ji S, Yang T. Stochastic optimization of areas under precision-recall curves with provable convergence. *Advances in neural information processing systems*. 2021; 34:1752–1765.
50. Slundberg SM. A Unified Approach to Interpreting Model Predictions;. Available from: <https://github.com/slundberg/shap>.