Màster Universitari
Anàlisi de Dades Òmiques / Omics Data Analysis
FACULTAT DE CIÈNCIES I TECNOLOGIA
UVIC | UVIC·UCC

Master of Science in Omics Data Analysis

Master Thesis

Integrated bioinformatics analysis of transcriptomics and immunological data to understand malaria immunity in naive and endemic populations

by

Carla Sánchez Almirall

Supervisor: Carlota Dobaño, ISGlobal, Hospital Clínic, Universitat de Barcelona

Co-supervisor: Gemma Moncunill, ISGlobal, Hospital Clínic, Universitat de Barcelona

Academic tutor: Josep Maria Serrat, Facultat de Ciències i Tecnologia, Universitat de Vic

Biosciences Department

University of Vic – Central University of Catalonia

12-09-2022

Malaria

Integrated bioinformatics analysis of transcriptomics and immunological data to understand malaria immunity in naive and endemic populations

Carla Sánchez[1,2],

1 Universitat de Vic - Universitat Central de Catalunya (UVic)

2 ISGlobal, Hospital Clínic, Universitat de Barcelona, Barcelona, Catalonia, Spain

Abstract

Motivation: Malaria is one of the deadliest diseases worldwide. Although there have been significant advances on its understanding and vaccine development, host immune responses involved in protection are not totally elucidated, and current vaccines are not highly efficacious, particularly in Africans. Controlled human malaria infections (CHMI) are useful tools to understand immunity. We aimed to characterize transcriptomics and immunological data from Africans and malaria-naive Europeans to address which is the effect of being pre-exposed to *Plasmodium falciparum* on such immune responses, and to compare protected and not protected Africans upon CHMI to identify protective immune signatures.

Methods: Applying the blood transcriptional module (BTM) framework and using Gene Ontology database, we characterized both the baseline and post-challenge transcriptomic and immunological profiles in 20 African and 5 European volunteers participating in a CHMI clinical trial in Gabon. We used DESeq2 for the differential expression analysis and we performed a gene set enrichment analysis (GSEA). In addition, we applied Camera to detect more differences than classical differential expression analysis as it accounts for the inter-gene correlation. In addition, DIABLO provided us the opportunity to perform correlations between the different omics datasets.

Results: Africans and Europeans had a distinct baseline profile. Genes overexpressed in Europeans were involved in B cells, T cells, natural killer (NK) cells pathways, whereas genes overexpressed in Africans were involved in neutrophils and monocytes (innate) pathways. In addition, Africans had higher levels of *P. falciparum* antibodies than Europeans but lower levels of eotaxin. Between Africans, although they did not have many baseline differences, they had different responses according to immunity to malaria. Non protected individuals had BTMs related to CD4+ T cells at day 11 post-infection, whereas protected individuals had BTMs enriched in B cells and natural killer cells. This type of studies shed light into immune cells and biomarkers which may underlie differential responses of Africans to infection and vaccination, but further investigations are needed to unravel the determinants and mechanisms of immunity.

Availability: Code is available at https://github.com/carlasanchez99/TFM.

Contact: carla.sanchez@uvic.cat

1 Introduction

Despite all attempts to reduce malaria, it is still a parasitic disease that generates huge morbidity and mortality in humans, being a major global public health challenge (World Health Organization [WHO], 2021). There were 241 million cases and 627 000 deaths of malaria worldwide in 2020 estimated by the WHO. Among the 4 species that can cause malaria, *Plasmodium falciparum* is the most lethal and prevalent. This parasite has a complex life cycle consisting of three stages, two taking place in the human host: pre-erythrocytic and erythrocytic. The pre-erythrocytic stage

starts when the bite of a female mosquito *Anopheles* release sporozoites into the human host. Then, sporozoites migrate to the liver to infect hepatocytes and multiplicate becoming merozoites. This first stage is asymptomatic and associated with few gene expression changes (Rothen et al., 2018). Then, merozoites are released into the blood stream, invading the red blood cells (erythrocytic stage) starting the symptomatic and highly pathogenic stage of malaria (Salamanca et al., 2019). This stage is associated with multiple changes, including strong immune responses that are also observed at gene expression level (Rothen et al., 2018).

After being repeatedly exposed to the parasite, naturally acquired immunity (NAI) against malaria parasites is developed (Achan et al., 2020). Although it prevents clinical manifestation and complications, it does not provide sterile immunity, since it does not prevent infection. Continued exposure to malaria parasites is required to be protected, otherwise immunity declines (Mischlinger et al., 2020). This protection is slowly developed, and it has high specificity in terms of species, stage, strain and variant. Children who live in malaria-endemic areas are the most vulnerable group as the maternal antibodies transferred decline and they start to develop NAI due to the parasite exposure (Dobbs & Dent, 2016).
Although it is known that NAI involves multiple responses, mainly against blood-stage antigens (located in rhoptries [RH5], merozoite surface [MSP2], apical membrane [AMA1]), it is still unclear which are the exact targets and mechanisms of protection (Shah et al., 2021).

This complex parasite life cycle, the polymorphisms of some target antigens, and a poor understanding of the host immune response (Gonzales et al., 2020), explain in part why the development of a highly effective malaria vaccine has failed. The most advanced vaccine is RTS,S/AS01E, against the pre-erythrocytic stage, developed to stop the parasite cycle before the start of the symptoms. Although it is recommended by the WHO for African children, it has a short and partial efficacy, therefore a more effective vaccine is needed (Laurens, 2020).

To rationally design second generation vaccines, we need to characterize better the immune responses acquired. In addition, most vaccines show lower efficacy in individuals from malaria-endemic areas compared to malaria naïve areas (Long & Zavala, 2016), and immune differences at baseline between the two populations may be responsible for such differences in vaccine responses.

Controlled human malaria infection (CHMI) is an excellent tool to study immune responses (Stanisic et al., 2018). CHMI traditionally consisted on the administration of bites of *Plasmodium*-infected insectary-raised mosquitoes. A new method, the PfSPZ Challenge, has been developed by Sanaria Inc, which consists on the injection via needle and syringe of infectious, aseptic, purified and cryopreserved *P. falciparum* sporozoites (NF54). Inoculating sporozoites allows both liver and blood-stage infection to develop. As soon as parasitemia is detected in the blood participants by microscopy, they are treated with antimalarials.

Here, we characterized the immune and transcriptional profiles of 5 malaria-naïve Europeans and 20 pre-exposed Africans at baseline and during infection to understand the effect of pre-exposure to malaria at different molecular and immunological levels. We also performed an analysis comparing Africans who were protected and the non-protected to identify immunity signatures. We used a multi-dimensional approach, comprising whole blood transcriptomics, whole genome expression microarrays, antibodies, cytokines and hematological data.
During the trial, all European participants developed parasitemia whereas 4 Africans did not, and overall detection of parasitemia in Africans was delayed compared to Europeans, showing partial protection.

2 Materials and Methods

Clinical trial

Participants were part of the LaCHMI-1 trial (Lell et al., 2018) conducted in Lambaréné, Gabon, in August 2014. Individuals considered malaria naïve did not

have history of *P. falciparum* malaria and no long-term (>5 years) residence in a malaria-endemic area. Semi-immune individuals were Africans living more than 10 years in a highly malaria-endemic area. There were 5 Europeans and 20 Africans.
Parasitemia before and during the CHMI trial was detected by Thick Blood Smear (TBS) or by polymerase chain reaction (PCR). Among the 20 African participants, 4 had low and asymptomatic parasitemia, 1 had gametocytemia by PCR and two had gametocytemia by both PCR and TBS before starting the trial. All participants were given clindamycin to cure possible *Plasmodium* infections. After treatment, all participants were free of sexual parasites 1 day before starting the trial. Volunteers were inoculated with 3,200 PfSPZ of PfSPZ Challenge by direct venous inoculation DVI (Day 0).

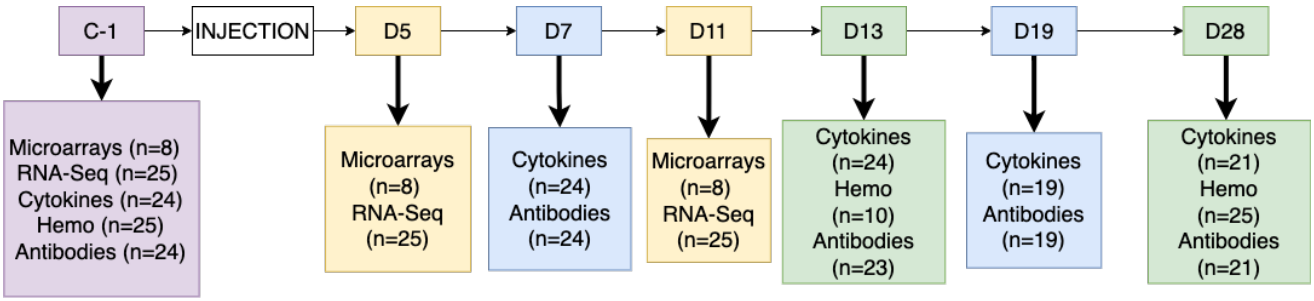


Figure 1: Flow chart of the study.

Statistical analysis of demographic and clinical variables

We assessed the association between the clinico-pathological variables and the two groups (Europeans and Africans) using a Fisher's exact test for categorical variables and Mann–Whitney U test or Student's t-test for the continuous ones.

Datasets

We had data available for all 25 individuals for RNA-Seq and hematological data and from 24 individuals we had antibodies and cytokines from the day before the injection (C-1) (Figure 1). Whole genome gene expression microarray data were available from 8 individuals, 3 samples for each one at study time points C-1, day (D)5 and D11. We also had RNA-Seq data for time points D5 and D11. For cytokines, hematological data and antibodies, there were time points D13 and D28. In addition, we had cytokines and antibodies data of D7 and D19.

There were time points where data was missing from some individuals.

We performed all data analyses using RStudio for macOS (RStudio; www.rstudio.com) and R x64 v.4.1.2 for macOS (R Foundation for Statistical Computing).

For all datasets, we first made an exploratory analysis consisting of a Principal Component analysis (PCA) using prcomp function to study sample aggregation, to detect group outliers and batch effects.

Whole genome gene expression microarrays

Microarrays were done at Progenika using 24 RNA samples (8 participants and 3 samples for each of them). Total RNA quantity and integrity were checked on Bioanalyzer and samples processed by Affymetrix standard protocols. The processed samples were hybridized to Affymetrix® GeneChip® microarrays (Human Gene 2.1 ST-24 array plate) using the GeneTitanTM platform. Softwares used in array processing were Command Console (AGCC GeneTitan® Control 4.3.0., Affymetrix®) for sample registration and GeneTitan control and ExpressionConsole (EC 1.3.1, Affymetrix®) for the Quality Control analysis.

The first steps of sample processing were the cDNA synthesis, fragmentation and labelling that were carried out with the kit Ovation Pico WTA System V2 and Encore Biotin Module (NuGen). The hybridization,

staining and scanning of the arrays were performed with equipment and protocols recommended by Affymetrix.

The microarrays dataset contained data from hybridization of each sample summarized with the RMA algorithm of Affymetrix and information about each of the probes deposited on the chip.

Data analysis of microarrays

In the microarray dataset there were 53617 probes and 24 samples. We performed a quality control to check if there were outliers or not and if the normalization was correct to ensure that the analysis could be done with all samples. Negative and positive control probes and probes with no annotated symbol were removed before the start of the differential expression analysis. Finally, 31135 probes remained.

The differential expression analysis was performed using the limma package v3.50.3 (Ritchie et al., 2015). First, unwanted variability was removed using sva function from sva package v.3.42.0 (Leek JT et al., 2021). Counts were already normalized with RMA method.

A linear model was then fitted for each gene with the lmFit function (adjusting for covariates donor and the surrogated variables calculated before) and contrasts were estimated with contrast.fit and eBayes functions. The p-values were corrected for multiple comparisons with the Benjamini and Hochberg (BH) method to control for the false discovery rate (FDR). To consider a differentially expressed gene in the comparisons, the filter applied was p-value <0.05 as using the p-value adjusted and logFC was too stringent and too few differentially expressed genes were found to perform an enrichment analysis.

Data analysis of RNA-Seq data

RNA-Seq data had been generated in a previous study (de Jong et al., 2021). We constructed an Expression Set object to perform the analysis. There were originally 58056 genes in the count matrix (de Jong et al., 2021).

For the differential gene expression analysis, we included 17437 genes, after filtering those with a minimum of 10 counts in a worthwhile number of samples. This step was performed using filterByExpr function as implemented in the edgeR R package v.3.36.0 (Robinson et al., 2010). Unwanted variability was removed using sva function as with microarrays.

Differential expression testing was conducted using the DESeq2 package, v.1.34.0 (Love et al., 2014). The DESeq2 workflow starts with a normalization by computing size factors, which addresses the differences not only in the library sizes, but also the library compositions. For each gene, a dispersion estimate is calculated. This dispersion value is equal to the squared coefficient of variation (variation divided by the mean). Then, a line is fit across the dispersion estimates of all genes versus the mean normalized counts of the genes. Dispersion values of each gene are shrunk towards the fitted line.

Next, a generalized linear model was fitted considering additional confounding variables that were age, sex, hemoglobin status, and the estimated surrogate variables obtained previously with sva. It uses negative binomial distribution for fitting count data. For a given contrast, a test for differential expression is carried out against the null hypothesis that the log fold change of the normalized counts of the gene in the given pair of groups is exactly zero. It adjusts p-values for multiple-testing. In almost all comparisons, to consider a differentially expressed gene in the comparisons, the filter applied was adjusted p-value < 0.05 and 2logFC > 1. When comparing semi-immune TBS+ and TBS -, the filter applied was p-value <0.05.

The analysis at D5 and D11 was conducted adjusting for baseline (C-1). We performed it with the Likelihood ratio test (LRT) which identifies any genes that show change in expression across the different levels. We provide the interaction term immune_status:t_point or tbs_pos:t_point to test any group-specific changes at any timepoints after time C-1.

Functional analysis

To determine whether differentially expressed genes were associated with a particular biological process or

molecular function, we performed a functional analysis using the clusterProfiler package v4.2.2 (Yu et al., 2012). First, we converted the gene symbols into ENTREZ ID using org.Hs.eg.db. package v3.14.0 (Marc Carlson, 2021). For the comparisons between naïve and semi-immune individuals, we used the GO biological process database with the enrichGO function, which is the world's largest source of information on genes' function. We also used BTMs (Blood Transcriptional Modules) from Li et al., 2014 in two approaches: with the differential expression analysis using gene set enrichment analysis (GSEA), and when assessing directly differences in gene sets without performing a differential expression analysis before. For this last approach, we used a combination of voom (Law et al., 2014)and camera (Wu & Smyth, 2012) both from limma package.

Camera is one of the few gene set tests that can properly account for intergene correlation in RNA-Seq data, it estimates the variance inflation factor for the gene expression that results from intergene correlation in the data and incorporates it into test procedures to control the apparent FDR. To perform voom, we used the same design matrix as the one used for the RNA-Seq analysis (adjusting for age, sex, hemoglobin status and the surrogate variables). Then we converted the genes into SYMBOLS and converted this list of gene identifiers into a list of indices for gene sets with ids2indices function. Next, we applied the camera function. To consider a significant genset, the filter was that the number of genes was bigger than 4 and FDR < 0.2 (Africans/Europeans comparison) or FDR < 0.25 (Africans TBS+/TBS- comparison). We filtered the TBA (bladder tumor antigen) BTMs gene sets as they are not related to what we are interested in.

Integrative omics analysis

We used DIABLO method (Data Integration Analysis for Biomarker discovery using Latent variable approaches for Omics studies) from mixOmics package v6.18.1 (Rohart et al., 2017) to integrate the significant features of RNA-Seq, microarray, antibody and cytokine data. DIABLO is a supervised approach that allows to integrate multiple datasets which variables also explain the categorical outcome of interest.

Before integration, individuals are filtered in order to have the same ones in all the datasets. Then, DIABLO is performed with bloc.splsda function. To see which are the correlations between the variables of the dataset, we first made a circosPlot. As we had a high number of variables, we only keept the ones that show a correlation > 0.8 and perform again the bloc.splsda function. We used the circosPlot function to show the correlations. We used the plotIndiv function with the result of the bloc.splsda to display how individuals grouped.

Antibodies, cytokines and hematological analysis

Antibody data was obtained through in-house quantitative suspension array technology (qSAT) assays applying the xMAP™ technology (Luminex Corp., Texas) (Sánchez et al., 2020). Antibody dataset contains data of IgM, IgG, IgG1, IgG2, IgG3 and IgG4 against 21 antigens either from the pre-erythrocytic (CelTOS, CSP, LSA-1, and SSP-2), erythrocytic (AARP, CyRPA-1, CyRPA-2, DBLa, EBA-175, MSP-3, PTRAMP, RH1, RH2, RH4, RH5) or both stages (AMA-1[3D7] AMA-1[FVO], EXP-1, MSP-1 19, MSP-1 42[3D7] and MSP-1 42[FVO]) of *P. falciparum*.

We used cytokines, chemokines and growth factors measured with the Cytokine Human Magnetic 30-Plex Panel from Life Technologies™ (Aguilar et al., 2019) including the following: epidermal growth factor (EGF), fibroblast growth factor (FGF), granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), hepatocyte growth factor (HGF), vascular endothelial growth factor (VEGF), tumor necrosis factor (TNF), interferon (IFN)-α, IFN-γ, interleukin (IL)-1RA, IL-1β, IL-2, IL-2R, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12(p40/p70), IL-13, IL-15, IL-17, IFN-γ induced protein (IP-10), monocyte chemoattractant protein (MCP-1), monokine induced by IFN-γ (MIG), macrophage inflammatory protein (MIP)-1α, MIP-1β and regulated on activation normal T cell expressed and secreted (RANTES) and eotaxin.

All cytokines and antibodies were in pg/mL and median fluorescence intensity (MFI) units, respectively, and were $\log_{10}$ transformed for further analysis.

The hematological dataset was obtained from the clinical trial (Lell et al., 2018) containing hemoglobin, leukocytes, lymphocytes, neutrophils, eosinophils and platelets information.
Normality of the data was tested using the Shapiro-Wilk test. If the data was not normal distributed, the Mann–Whitney U test was applied to see whether there were significant differences between groups. For the normal distributed data, the F-test was used to see whether variances were equal or not and then the Student's t-test was applied to determine if there were significant differences. We corrected the p-values (FDR) through the BH method.

3 Results

Description of participants

RNA-Seq, antibody, cytokine, whole genome gene expression microarrays and hematological data from a total of 25 participants were analyzed (Figure 1). Among the 25 individuals, there were 5 Europeans (malaria naïve) and 25 Africans (malaria exposed). Europeans were significantly older than Africans.

All European participants developed parasitemia tested by TBS and by PCR during the CHMI trial, as expected. However, among the 20 Africans, 12 were positive by TBS and 16 were positive by PCR, reflecting protective immunity for some Africans. All the 4 individuals who tested negative for parasitemia by PCR were females (*p*-value: 0.03).
Africans took longer to develop parasitemia by TBS than Europeans, suggesting partial immunity, but no significant differences were seen in days to parasitemia by PCR. Sex and hemoglobin status distribution were not significantly different between groups.

Differences between Africans and Europeans at baseline

A total of 309 genes were found to be differentially expressed between groups at baseline (before sporozoite inoculation) using RNA-Seq data. 247 were upregulated in Africans and 63 were upregulated in Europeans. Some of these 63 genes (*TYROBP, GAB2, LAT2, LILRA2, CXCR2* and *TREM1*) take part in immune related pathways defined in the GO database as “myeloid cell activation”, “granulocyte activation” and “superoxide anion generation”. The rest of the genes and the ones upregulated in Africans are not involved in any immune related pathway using GO database.

Variables	Europeans	Africans	*p*-value
Number (n, %)	5 (20)	20 (80)	
Sex, male (n, %)	1 (20)	12 (60)	NS
Age, years (mean, SD)	26.2 ± 2.04	21.6 ± 2.34	<0.001
HbAA hemoglobin genotype (%)	5 (100)	11 (55)	NS
Parasitemic subjects by TBS (n, %)	5 (100)	12 (60)	NS
Days to parasitemia by TBS (mean, min, max)	12.4 (12-14)	22.15 (13-28)	<0.001
Parasitemic subjects by PCR	5 (100)	16 (80)	NS
Days to parasitemia by PCR (mean, min, max)	8 (7-14)	7.85 (0-28)	NS

Table 1: Description of study participants. The statistical significance was considered when p-value < 0.05; SD: standard deviation; NS: not significant; TBS: Thick Blood Smear; PCR: Polymerase Chain Reaction.

Performing GSEA, we did not find any BTMs enriched with these DEG. However, we found a total of 106 BTMs with the analysis using camera (Supplementary Table 1). Gene sets upregulated in Africans were related mainly with B cells, T cells, natural killer (NK) cells and platelet activation. Furthermore, the gene sets upregulated in Europeans were more focused in neutrophils, monocytes, dendritic cells, interferon responses and antigen presentation.

Malaria antibody responses were overall higher in African participants than Europeans at baseline, as expected, because the later had not been exposed to malaria. The biggest differences were found for IgG and IgG1, IgG3 responses. There were no significant differences in IgM and few in IgG4 and IgG2

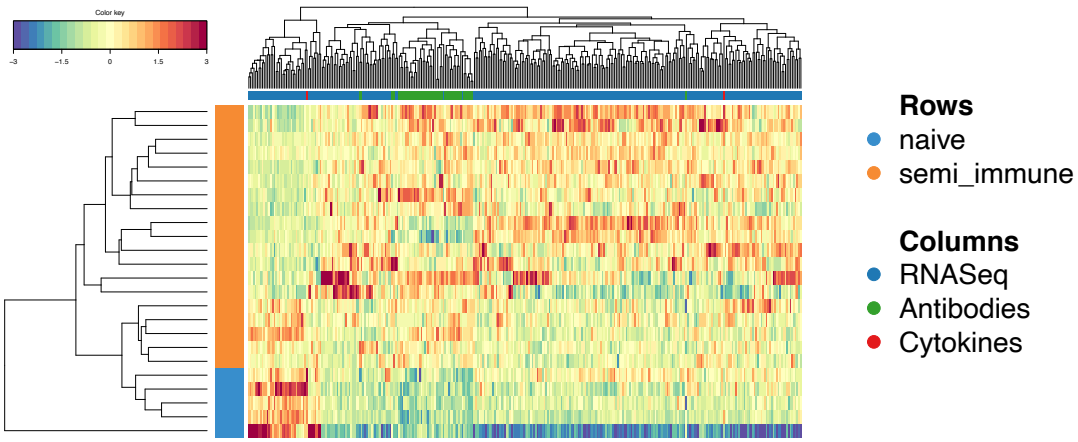


Figure 2: Heatmap of the 309 DEG, 47 antibodies and 2 cytokines significantly different between Africans and Europeans at baseline.

responses. Antibody differences were found to pre-erythrocytic antigens: CelTOS, CSP, SSP-2; erythrocytic antigens: DBLα, EBA-175, LSA-1, MSP-3, RH2, PTRAMP, CyRPA-1, AARP; and antigens expressed in both stages: EXP-1, MSP-1_{19}, MSP-1_{42}, MSP3.

From the 30 cytokines analyzed, Europeans had higher levels of eotaxin (FDR < 0.05) than Africans whereas Africans had higher levels of EGF (FDR < 0.1). In addition, from the hematological data, there were only significantly differences in eosinophils, which were higher in Africans.

When using DIABLO to combine the differentially expressed genes, cytokines and antibodies that also show differences, Africans and Europeans had quite a different baseline profile (Figure 2). Individuals clustered within each group although there was one individual (L1-001, European) who differed; this individual had lower counts in these selected genes than the other individuals. When plotting the individuals as the result of the splsda, the scatter plots of each of the 3 datasets showed also a good separation between the groups (Figure 3). These plots use 2 variates as a combination of features trying to maximize the covariance between variables and groups.

Association between DEG, antibodies and cytokines showed that the genes that had a positive correlation with antibodies were the ones overexpressed in Africans, and the genes that were upregulated in Europeans had a negative correlation with antibodies (Figure 4). Eotaxin positively correlated with genes such as *LILRA2, HLX, NINJ1, TREM1…* that are involved in immune response pathways such as leukocyte activation, pattern recognition receptor signaling pathway and positive regulation of defense response (GO database).

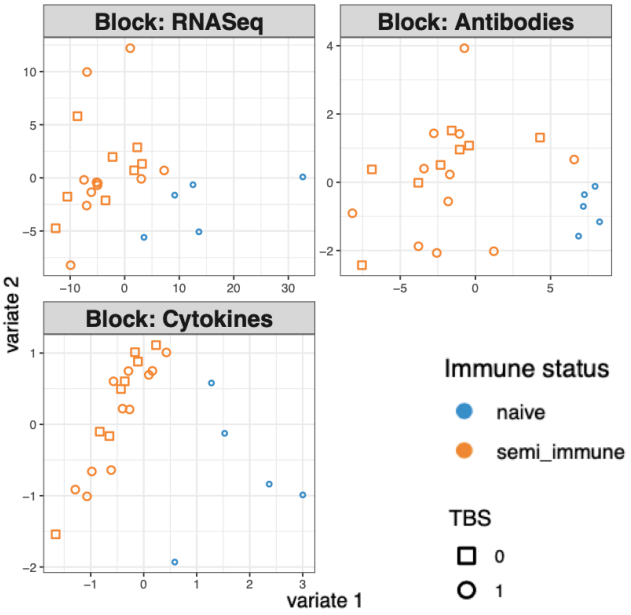


Figure 3: Scatter plots results of the splslda function using the 309 DEGs, 47 antibodies and 2 cytokines.

Differences between Africans and Europeans at different time points

When comparing both groups at each of the 3 time points, only 9 DEGs were common in all of them (Supplementary Figure 1). At day 5, Europeans still have upregulated pathways as “immune response-regulating signaling pathway”, “neutrophil activation involved

in immune response", "response to type I interferon" etc. The most differences in the number of genes were observed at D11, with 1049 DEGs. From those 1049, the ones upregulated in Africans are now the ones related to immune pathways as "response to interferon-gamma", "antigen processing and presentation of peptide antigen", "T cell mediated immunity" etc using GO database, contrary to what happened at C-1.

There were also different antibody responses between the groups, but only at time point D28 Europeans had statistical higher levels of IgM to MSP-1_{42}[FVO], MSP-1_{42}[3D7], MSP-1_{19}, EXP-1 and DBL-α.

At D7, apart from eotaxin, FGF and IL1-β were also higher in Europeans, and IL-10 in Africans. The time point which showed more significant differences was D13, being IL-8, IL-2, IL-1RA, IFN-γ, VEGF, HGF, IL-15, MCP-1, MIP-1β, MIG, GM-CSF and IL-6 significantly higher in Europeans compared to Africans.

Protective signatures at baseline; comparison between TBS+ and TBS- Africans

Looking at what genes were different at baseline between protected (TBS-, n=8) and non-protected African individuals (TBS+, n=12), we found 654 DEGs. Some of the genes upregulated in the TBS- group are involved in the immunological pathways "complement activation", "B cell receptor signaling pathway" etc us-

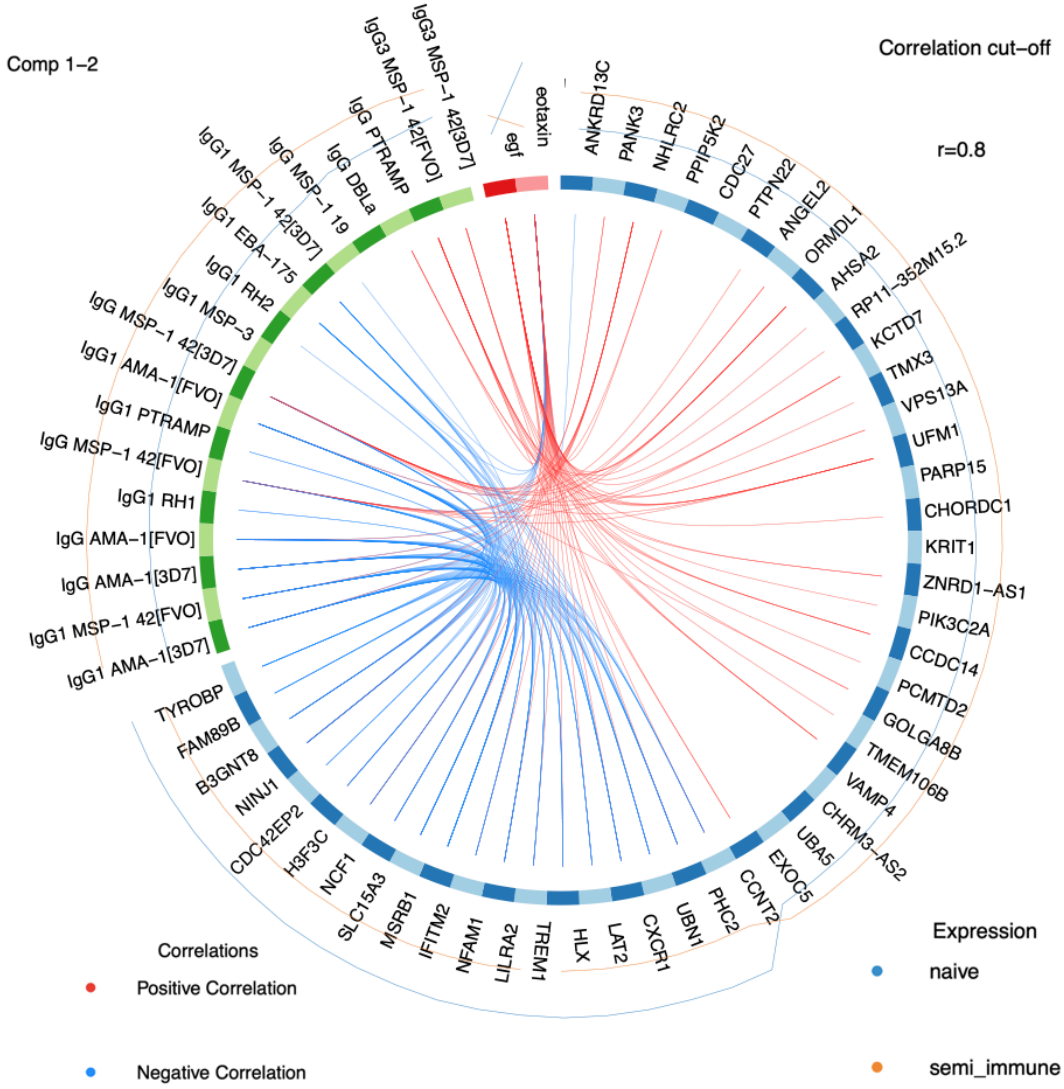


Figure 4: CircosPlot showing correlation>0.8 between DEG, antibodies and cytokines significantly different between Europeans and Africans at baseline (C-1).

We did not detect any immunological pathways enriched with the genes upregulated in TBS+ individuals. Furthermore, we did not find any BTMs enriched with the DEGs performing GSEA. Using camera, we observed few BTMs enriched in each group, BTMs upregulated in TBS- are related to T cells and BTMs upregulated in TBS+ are related to platelets, neutrophils ets

We also performed an analysis of antibody, cytokine and hematological data to add them to DIABLO. We only find significant differences in antibodies (FDR < 0.05) between Africans TBS+ and TBS– at baseline. The significant ones were IgG3 and IgG4 against both strains of AMA-1 antigens and IgG3 against CelTOS.

Taking advantage that we had RNA microarray data pre-challenge, we also carried out a differential expression analysis between Africans that were TBS- (n=2) and the TBS+ (n=6) at baseline. We detected 1819 DEGs (p-value <0.05) between the groups. From these 1819 there are 77 DEGs that overlapped with DEGs obtained through the RNA-Seq analysis.

Using DIABLO, we combined the 1819 DEGs with the mentioned antibodies and found a high correlation between IgG4 AMA-1[3D7] and genes (*BST2, CCR8, CD80, EIF2AK2, EXOSC4, IL12B, MX2, OAS1, OAS2, OAS3, TRIM25 and UBE2J1*) involved in immunological pathways, they are upregulated in TBS+ group. Focusing on these genes, Pearson correlation

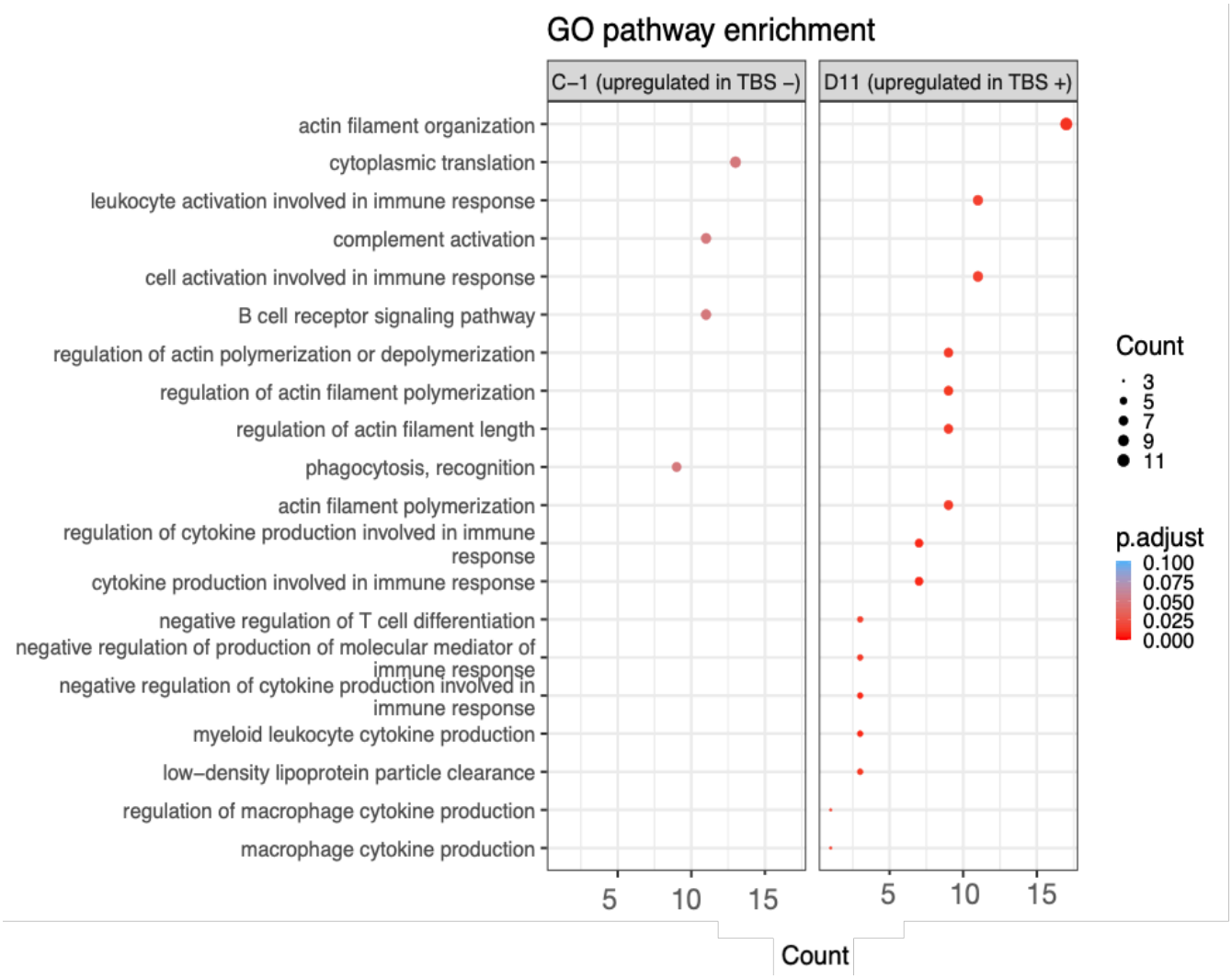


Figure 5: GO pathways enriched at C-1 in TBS- individuals and at D11 in TBS+ individuals.

was performed and CD80 was the gene that most correlated with IgG4 AMA-1[3D7] (Supplementary Figure 2).

Comparison among TBS+ and TBS- Africans at D11 and D13

While there were 4 enriched pathways at C-1 in TBS- and 0 in TBS+, there were 40 enriched pathways at D11 (adjusting for baseline) in TBS+ and 0 in TBS- using GO database (Figure 5). These pathways are mainly related to leukocytes and cytokines. At D11, although there were no BTMs enriched performing GSEA, 40 BTMs were enriched (FDR < 0.25) using camera method (Supplementary Table 2). Among these 40 BTMs, 7 were upregulated in TBS- individuals and 33 in TBS+ individuals. While these 7 BTMs are related to B cells, NK and innate activation (similar BTMs were upregulated in Africans compared to naïve at baseline), the other 33 are more related to CD4 T cells and cell cycle.

4 Discussion

Here, we studied the baseline profiles of Africans and Europeans to determine the effect of pre-exposure to malaria on the immune profile using gene expression, antibody, cytokine and hematological data. In addition, inside the Africans group, we compared the baseline profiles of individuals protected (TBS-) and individuals who developed parasitemia (TBS+) to identify signatures of protection.

Whereas previous studies of malaria CHMI are focused on the longitudinal response (de Jong et al., 2021), i.e. the response to infection, we carried out an analysis comparing the baseline profiles. The immune status at baseline can predict responses to vaccination and infection outcomes (Moncunill et al., 2020) (Tsang et al., 2020). Here, we found that Africans and Europeans have a clear distinguishable baseline profile.

Africans had enriched pathways related to B and T cells compared to Europeans, which may be a sign of acquired immunity developed in part to exposition to malaria. However, this could also be due to genetic differences between populations and differential exposure to other environmental factors (at gene expression levels we cannot distinguish between them). In contrast, the levels of antibodies against specific *Plasmodium* antigens are higher in Africans, confirming malaria pre-exposure and acquired immunity in Africans at baseline. Nevertheless, malaria-specific antibodies and memory B cells by themselves have not been found to be protective from *P. falciparum* infection or clinical disease (Nogaro et al., 2011), which might explain why some of the Africans developed parasitemia and others did not. On the contrary, Europeans had enriched pathways related to neutrophils and interferon type I response, the later having been involved in protective responses against blood-stage infection or in immune evasive mechanism depending on host–parasite genetics (Sebina & Haque, 2018). It could be that malaria exposure blunts such neutrophil and immune responses as a mechanism of tolerance to the malaria disease.

While in most time points during the CHMI Africans had higher levels of antibodies than Europeans, some cytokines were higher in Europeans than Africans almost in all time points post-infection, indicating a higher cytokine response in Europeans than Africans. This could reflect a higher inflammatory response in naives in comparison to exposed who may have more tolerance to infection.

Of note, among the Africans, although at baseline we did not find many immune enriched pathways between protected and non-protected individuals, more differences were found at D11, suggesting that despite the similar baseline transcriptomic profiles, the responses induced by *Plasmodium* are different.
In addition, in TBS- individuals, we identified high correlations between IgG4 AMA-1[3D7] and genes involved on immunological pathways.

It is important to note that what happens at the transcriptomic level does not always translate at the downstream level. This could explain why non-protected individuals at D11 had BTMs related to CD4 T cells, whereas different studies show that $CD4^+$ T cells have a protective role against malaria (Süss et al., 1988). Here, we could be detecting an innate

activation and differentiation of $CD4^+$ T cells in the unprotected ones while the protected ones, who may have more parasitemia control, may be less activated by innate responses. Protected individuals may have activation of malaria-specific memory T cells but being at low frequency it may have gone undetected by transcriptomics.
On the contrary, protected individuals had BTMs enriched in B cells and natural killer cells, both previously associated with malaria protection.

In this study, combining camera and voom using BTMs has been key to identify the gene sets enriched in each group. While using GSEA and BTMs with the genes differentially expressed we did not find any BTM enriched, camera was more sensitive. Camera allows to investigate sets of genes when the individual gene associations are not strong. Using camera has also been very important and provided advantages as it accounts for the inter-gene correlation, which is ignored when doing a classical gene expression analysis that is important especially in this kind of studies.

BTMs (Li et al., 2014)are widely used in systems immunology studies as they are mainly focused to immune pathways and immune cells. The BTMs used in this analysis were done characterizing the innate and adaptive immune responses to different vaccines using systems biological approaches (Li et al., 2014). Using these BTMs has prevented us from losing information as happened when using general databases as GO. GO is not only focused on the immunology field, therefore it has more pathways to test but less immune. In addition, the use of DIABLO allowed us to correlate the different datasets, which helped us understand the relationship between the different kind of omics.

One of the limitations of this study has been the small sample size that has restricted to obtain conclusions with greater statistical weight in the analysis. In addition to the small sample size, the number of samples of the datasets were sometimes different, and there were some missing data, which made the analysis less comparable between datasets (e.g. microarrays vs. RNAseq) and limited data integration. Another limitation mentioned above was that Africans and Europeans have a different genetic background and are also exposed to other diseases or environmental factors, which did not allow us to determine that all the differences are due to exposure to malaria.

5 Conclusion

In conclusion, the effect of being pre-exposed to malaria was correlated to enriched pathways of B cells, T cells and NK cells compared to naïve individuals. Although we observed few transcriptomic differences among protected Africans and the non-protected ones at baseline, protected individuals had overexpression of genes involved in B cells and NK pathways whereas the non-protected ones have pathways enriched in T cells and this is consistent with their role in immunity. The use of BTMs, camera and DIABLO methods had been keys in this study, helping us to integrate and understand the transcriptomic and immunological profiles. For future studies, it will be important the use of these methods as they gave more information and help us understand what happens than just analyzing each dataset separately.

6 Data availability

All code used in this study is available at https://github.com/carlasanchez99/TFM.
Further information is available upon reasonable request.

7 Acknowledgements

I would like to thank the supervisors of this project, Carlota Dobaño and Gemma Moncunill, for their great interest, patience, motivation and for encouraging me during all these months and for giving me the opportunity to be part of ISGlobal. Also, I would like to thank Josep Maria Serrat for being the University supervisor. Finally, I would like to thank to my family and friends, for all the unconditional support this year.

8 Bibliography

Achan, J., Reuling, I. J., Yap, X. Z., Dabira, E., Ahmad, A., Cox, M., Nwakanma, D., Tetteh, K., Wu, L., Bastiaens, G. J. H., Abebe, Y., Manoj, A., Kaur, H., Miura, K., Long, C., Billingsley, P. F., Sim, B. K. L., Hoffman, S. L., Drakeley, C., … D'Alessandro, U. (2020). Serologic Markers of Previous Malaria Exposure and Functional Antibodies Inhibiting Parasite Growth Are Associated With Parasite Kinetics Following a Plasmodium falciparum Controlled Human Infection. *Clinical Infectious Diseases*, *70*(12), 2544–2552. https://doi.org/10.1093/cid/ciz740

Aguilar, R., Campo, J. J., Chicuecue, S., Cisteró, P., Català, A., Luis, L., Ubillos, I., Galatas, B., Aide, P., Guinovart, C., Moncunill, G., & Dobaño, C. (2019). Changing plasma cytokine, chemokine and growth factor profiles upon differing malaria transmission intensities. *Malaria Journal*, *18*(1), 406. https://doi.org/10.1186/s12936-019-3038-x

de Jong, S. E., van Unen, V., Manurung, M. D., Stam, K. A., Goeman, J. J., Jochems, S. P., Höllt, T., Pezzotti, N., Mouwenda, Y. D., Betouke Ongwe, M. E., Lorenz, F.-R., Kruize, Y. C. M., Azimi, S., König, M. H., Vilanova, A., Eisemann, E., Lelieveldt, B. P. F., Roestenberg, M., Sim, B. K. L., … Yazdanbakhsh, M. (2021). Systems analysis and controlled malaria infection in Europeans and Africans elucidate naturally acquired immunity. *Nature Immunology*, *22*(5), 654–665. https://doi.org/10.1038/s41590-021-00911-7

Dobbs, K. R., & Dent, A. E. (2016). Plasmodium malaria and antimalarial antibodies in the first year of life. *Parasitology*, *143*(2), 129–138. https://doi.org/10.1017/S0031182015001626

Gonzales, S. J., Reyes, R. A., Braddom, A. E., Batugedara, G., Bol, S., & Bunnik, E. M. (2020). Naturally Acquired Humoral Immunity Against Plasmodium falciparum Malaria. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.594653

Laurens, M. B. (2020). RTS,S/AS01 vaccine (Mosquirix™): an overview. *Human Vaccines & Immunotherapeutics*, *16*(3), 480–489. https://doi.org/10.1080/21645515.2019.1669415

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Zhang Y, Storey JD, & Torres LC. (2021). *sva: Surrogate Variable Analysis. R package version 3.42.0.*

Lell, B., Mordmüller, B., Dejon Agobe, J.-C., Honkpehedji, J., Zinsou, J., Mengue, J. B., Loembe, M. M., Adegnika, A. A., Held, J., Lalremruata, A., Nguyen, T. T., Esen, M., KC, N., Ruben, A. J., Chakravarty, S., Lee Sim, B. K., Billingsley, P. F., James, E. R., Richie, T. L., … Kremsner, P. G. (2018). Impact of Sickle Cell Trait and Naturally Acquired Immunity on Uncomplicated Malaria after Controlled Human Malaria Infection in Adults in Gabon. *The American Journal of Tropical Medicine and Hygiene*, *98*(2), 508–515. https://doi.org/10.4269/ajtmh.17-0343

Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D. S., Johnson, S. E., Milton, A., Rajam, G., Kasturi, S., Carlone, G. M., Quinn, C., Chaussabel, D., Palucka, A. K., Mulligan, M. J., Ahmed, R., Stephens, D. S., Nakaya, H. I., & Pulendran, B. (2014). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nature Immunology*, *15*(2), 195–204. https://doi.org/10.1038/ni.2789

Long, C. A., & Zavala, F. (2016). Malaria vaccines and human immune responses. *Current Opinion in Microbiology*, *32*, 96–102. https://doi.org/10.1016/j.mib.2016.04.006

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Marc Carlson. (2021). *org.Hs.eg.db: Genome wide annotation for Human. R package version 3.14.0.*

Mischlinger, J., Rönnberg, C., Álvarez-Martínez, M. J., Bühler, S., Paul, M., Schlagenhauf, P., Petersen, E., & Ramharter, M. (2020). Imported Malaria in Countries where Malaria Is Not Endemic: a Comparison of Semi-immune and Nonimmune Travelers. *Clinical Microbiology Reviews*, *33*(2). https://doi.org/10.1128/CMR.00104-19

Moncunill, G., Scholzen, A., Mpina, M., Nhabomba, A., Hounkpatin, A. B., Osaba, L., Valls, R., Campo, J. J., Sanz, H., Jairoce, C., Williams, N. A., Pasini, E. M., Arteta, D., Maynou, J., Palacios, L., Duran-Frigola, M., Aponte, J. J., Kocken, C. H. M., Agnandji, S. T., … Dobaño, C. (2020). Antigen-stimulated PBMC transcriptional protective signatures for malaria immunization. *Science Translational Medicine*, *12*(543). https://doi.org/10.1126/scitranslmed.aay8924

Nogaro, S. I., Hafalla, J. C., Walther, B., Remarque, E. J., Tetteh, K. K. A., Conway, D. J., Riley, E. M., & Walther, M. (2011). The Breadth, but Not the Magnitude, of Circulating Memory B Cell Responses to P. falciparum Increases with Age/Exposure in an Area of Low Transmission. *PLoS ONE*, *6*(10), e25582. https://doi.org/10.1371/journal.pone.0025582

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Rothen, J., Murie, C., Carnes, J., Anupama, A., Abdulla, S., Chemba, M., Mpina, M., Tanner, M., Lee Sim, B. K., Hoffman, S. L., Gottardo, R., Daubenberger, C., & Stuart, K. (2018). Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period. *PLOS ONE*, *13*(6), e0199392. https://doi.org/10.1371/journal.pone.0199392

Salamanca, D. R., Gómez, M., Camargo, A., Cuy-Chaparro, L., Molina-Franky, J., Reyes, C., Patarroyo, M. A., & Patarroyo, M. E. (2019). Plasmodium falciparum Blood Stage Antimalarial Vaccines: An Analysis of Ongoing Clinical Trials and New Perspectives Related to Synthetic Vaccines. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02712

Sánchez, L., Vidal, M., Jairoce, C., Aguilar, R., Ubillos, I., Cuamba, I., Nhabomba, A. J., Williams, N. A., Díez-Padrisa, N., Cavanagh, D., Angov, E., Coppel, R. L., Gaur, D., Beeson, J. G., Dutta, S., Aide, P., Campo, J. J., Moncunill, G., & Dobaño, C. (2020). Antibody responses to the RTS,S/AS01E vaccine and Plasmodium falciparum antigens after a booster dose within the phase 3 trial in Mozambique. *Npj Vaccines*, *5*(1), 46. https://doi.org/10.1038/s41541-020-0192-7

Sebina, I., & Haque, A. (2018). Effects of type I interferons in malaria. *Immunology*, *155*(2), 176–185. https://doi.org/10.1111/imm.12971

Shah, Z., Naung, M. T., Moser, K. A., Adams, M., Buchwald, A. G., Dwivedi, A., Ouattara, A., Seydel, K. B., Mathanga, D. P., Barry, A. E., Serre, D., Laufer, M. K., Silva, J. C., & Takala-Harrison, S. (2021). Whole-genome analysis of Malawian Plasmodium falciparum isolates identifies possible targets of allele-specific immunity to clinical malaria. *PLOS Genetics*, *17*(5), e1009576. https://doi.org/10.1371/journal.pgen.1009576

Stanisic, D. I., McCarthy, J. S., & Good, M. F. (2018). Controlled Human Malaria Infection: Applications, Advances, and Challenges. *Infection and Immunity*, *86*(1). https://doi.org/10.1128/IAI.00479-17

Süss, G., Eichmann, K., Kury, E., Linke, A., & Langhorne, J. (1988). Roles of CD4- and CD8-bearing T lymphocytes in the immune response to the erythrocytic stages of Plasmodium chabaudi. *Infection and Immunity*, *56*(12), 3081–3088. https://doi.org/10.1128/iai.56.12.3081-3088.1988

Tsang, J. S., Dobaño, C., VanDamme, P., Moncunill, G., Marchant, A., Othman, R. ben, Sadarangani, M., Koff, W. C., & Kollmann, T. R. (2020). Improving Vaccine-Induced Immunity: Can Baseline Predict Outcome? *Trends in Immunology*, *41*(6), 457–465. https://doi.org/10.1016/j.it.2020.04.001

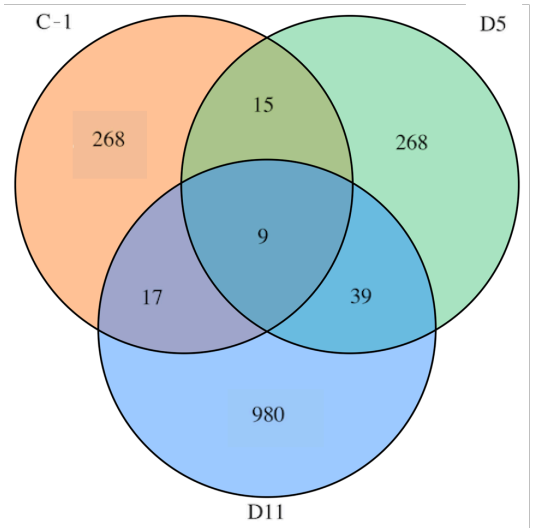
Wu, D., & Smyth, G. K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Research*, *40*(17), e133–e133. https://doi.org/10.1093/nar/gks461

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

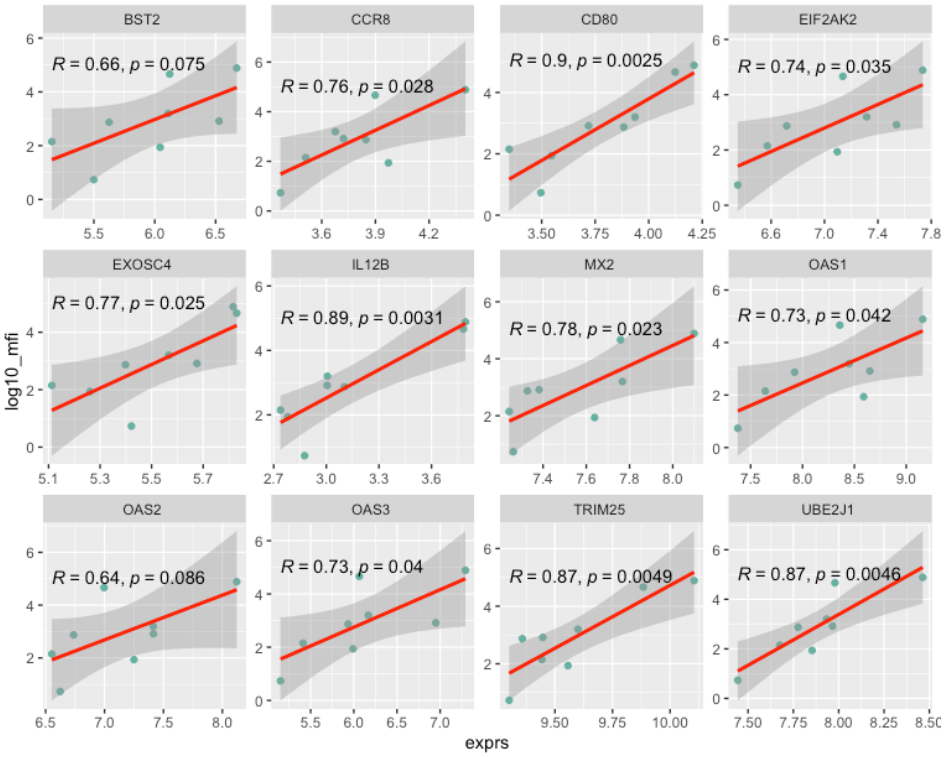
9 Supplementary material

Supplementary Figures

Supplementary Figure 1: *Venn diagram intersecting the DEGs of RNA-Seq data preforming DESeq2 analysis between Africans and Europeans. C-1: day before the injection; D5: day 5 post-infection; D11: day 11 post-infection. D5 and D11 are adjusted for baseline.*



Supplementary Figure 2: *Pearson's correlation between IgG4 AMA-1[3D7] and genes involved in immunological pathways (differentially expressed between TBS+ and TBS- with 8 individuals). R: value of the statistic test; p: p-value*



Supplementary Tables

Supplementary Table 1: *Gene sets significantly (FDR < 20%) associated with pre-exposition to malaria from analysis between Africans and Europeans at baseline. Down: upregulated in Africans. Up: upregulated in Europeans. FDR: False discovery rate.*

Direction	Geneset	P-value	FDR
Down	enriched in B cells (II) (M47.1)	0	0,00001
Down	enriched in B cells (I) (M47.0)	0	0,00003
Down	spliceosome (M250)	0,00003	0.00038
Down	mitosis (TF motif CCAATNNSNNNGCG) (M169)	0,0001	0.00107
Down	Plasma cell surface signature (S3)	0.00027	0.00257
Down	enriched in B cells (VI) (M69)	0.00052	0.00423
Down	enriched in NK cells (I) (M7.2)	0.00054	0.00434
Down	cell cycle (I) (M4.1)	0.00071	0.00556
Down	mitotic cell cycle in stimulated CD4 T cells (M4.9)	0.00101	0.00724
Down	transmembrane transport (II) (M191)	0.00107	0.00751
Down	heme biosynthesis (I) (M171)	0.00152	0.00945
Down	cell movement, Adhesion & Platelet activation (M30)	0.00168	0.00977
Down	cell cycle (III) (M103)	0.00216	0.01215
Down	platelet activation & blood coagulation (M199)	0.00351	0.01823
Down	E2F1 targets (Q4) (M10.1)	0.00419	0.02143
Down	cell division (M37.3)	0.00443	0.02202
Down	enriched in B cells (V) (M47.4)	0.00502	0.02427
Down	enriched in B cells (IV) (M47.3)	0.00517	0.02431
Down	enriched for TF motif TNCATNTCCYR (M232)	0.00537	0.02456
Down	CD4 T cell surface signature Th1-stimulated (S6)	0.0057	0.02573
Down	G protein coupled receptors cluster (M155)	0.01035	0.04228
Down	PLK1 signaling events (M4.2)	0.0114	0.04601
Down	enriched in T cells (I) (M7.0)	0.01469	0.05599
Down	enriched in G-protein coupled receptors (M130)	0.0188	0.06717
Down	cytoskeleton/actin (SRF transcription targets) (M145.1)	0.0201	0.06936
Down	enriched for ubiquitination (M138)	0.02022	0.06936
Down	lymphocyte generic cluster (M60)	0.02136	0.07227
Down	enriched for TF motif PAX3 (M179)	0.03392	0.10559
Down	cell adhesion (M51)	0.03417	0.10559
Down	enriched in B cells (III) (M47.2)	0.03787	0.11495
Down	mitotic cell cycle (M4.7)	0.04695	0.13768
Down	integrin cell surface interactions (I) (M1.0)	0.04696	0.13768
Down	enriched in NK cells (II) (M61.0)	0.05813	0.16343
Down	DC surface signature (S5)	0.06551	0.18121
Down	extracellular region cluster (GO) (M189)	0.06645	0.18216
Down	T cell differentiation via ITK and PKC (M18)	0.06758	0.18216
Down	chemokine cluster (II) (M27.1)	0.06798	0.18216
Down	collagen, TGFB family et al (M77)	0.06861	0.18244
Down	extracellular matrix (I) (M2.0)	0.07294	0.19245
Up	enriched in neutrophils (I) (M37.1)	0	0
Up	enriched in monocytes (II) (M11.0)	0	0

Up	TLR and inflammatory signaling (M16)	0	0
Up	NK cell surface signature (S1)	0	0
Up	platelet activation - actin binding (M196)	0	0
Up	enriched in monocytes (IV) (M118.0)	0	0
Up	suppression of MAPK signaling (M56)	0	0
Up	plasma membrane, cell junction (M162.0)	0	0
Up	enriched in activated dendritic cells/monocytes (M64)	0	0
Up	formyl peptide receptor mediated neutrophil response (M11.2)	0	0
Up	recruitment of neutrophils (M132)	0	0
Up	myeloid cell enriched receptors and transporters (M4.3)	0	0,00001
Up	enriched in monocytes (surface) (M118.1)	0	0,00001
Up	regulation of transcription, transcription factors (M213)	0	0,00002
Up	blood coagulation (M11.1)	0	0,00003
Up	immune activation - generic cluster (M37.0)	0	0,00003
Up	enriched in neutrophils (II) (M163)	0	0,00008
Up	enriched in monocytes (III) (M73)	0,00001	0.00011
Up	lysosomal/endosomal proteins (M139)	0,00001	0.00011
Up	platelet activation (III) (M42)	0,00001	0.00011
Up	RA, WNT, CSF receptors network (monocyte) (M23)	0,00001	0.00014
Up	platelet activation (II) (M32.1)	0,00001	0.00019
Up	cell junction (GO) (M4.13)	0,00003	0.00038
Up	cell cycle and growth arrest (M31)	0,00003	0.00041
Up	antiviral IFN signature (M75)	0.00012	0.0012
Up	inflammatory response (M33)	0.00016	0.00163
Up	enriched in antigen presentation (III) (M95.1)	0.00017	0.00164
Up	type I interferon response (M127)	0.00043	0.00388
Up	regulation of localization (GO) (M63)	0.00045	0.00399
Up	regulation of antigen presentation and immune response (M5.0)	0.00052	0.00423
Up	extracellular matrix, complement (M140)	0.00077	0.00586
Up	platelet activation (I) (M32.0)	0,0009	0.00667
Up	myeloid, dendritic cell activation via NFkB (II) (M43.1)	0.00091	0.00667
Up	BCR signaling (M54)	0.00126	0.00845
Up	MAPK, RAS signaling (M100)	0.00137	0.00887
Up	KLF12 targets network (M32.3)	0.00151	0.00945
Up	cell cycle and transcription (M4.0)	0.00159	0.00956
Up	TLR8-BAFF network (M25)	0.00162	0.00958
Up	AP-1 transcription factor network (M20)	0.0021	0.01203
Up	RIG-1 like receptor signaling (M68)	0.00236	0.01304
Up	transmembrane transport (I) (M87)	0.00289	0.0155
Up	viral sensing & immunity; IRF2 targets network (II) (M111.1)	0.00302	0.01593
Up	chemokines and inflammatory molecules in myeloid cells (M86.0)	0.00442	0.02202
Up	interferon alpha response (II) (M158.1)	0.00513	0.02431
Up	viral sensing & immunity; IRF2 targets network (I) (M111.0)	0.00756	0.03323
Up	inflammasome receptors and signaling (M53)	0.00821	0.03518
Up	enriched in antigen presentation (II) (M95.0)	0.00867	0.03625
Up	Naive B cell surface signature (S8)	0.00954	0.03943
Up	regulation of signal transduction (M3)	0.01193	0.04758
Up	endoplasmic reticulum (M37.2)	0.01292	0.05096
Up	innate activation by cytosolic DNA sensing (M13)	0.01415	0.05516

Up	CORO1A-DEF6 network (I) (M32.2)	0.01461	0.05599
Up	cell adhesion (GO) (M117)	0.01666	0.06211
Up	cytoskeletal remodeling (M32.8)	0.01722	0.0635
Up	extracellular matrix (II) (M2.1)	0.01833	0.06616
Up	CORO1A-DEF6 network (II) (M32.4)	0.01906	0.06739
Up	B cell development (M9)	0.02149	0.07227
Up	transmembrane transport (SLC cluster) (M154.1)	0.02641	0.08712
Up	CD1 and other DC receptors (M50)	0.03201	0.10165
Up	activated dendritic cells (M67)	0.03326	0.10466
Up	lipid metabolism, endoplasmic reticulum (M92)	0.03763	0.11495
Up	complement and other receptors in DCs (M40)	0.04475	0.13391
Up	CCR1, 7 and cell signaling (M59)	0.0449	0.13391
Up	myeloid cell cytokines, metallopeptidases and laminins (M78)	0.05248	0.15256
Up	golgi membrane (I) (M113)	0.0541	0.15592
Up	translation initiation factor 3 complex (M245)	0.06367	0.17754
Up	respiratory electron transport chain (mitochondrion) (M231)	0.0675	0.18216

Supplementary Table 2: *Gene sets significantly (FDR < 25%) associated with protection to malaria from analysis between Africans TBS+ and TBS- at D11 (adjusting for baseline). Down: upregulated in TBS+. Up: upregulated in TBS-.*

Direction	Geneset	P-value	FDR
Down	cell cycle (I) (M4.1)	0	0
Down	plasma cells & B cells, immunoglobulins (M156.0)	0	0,00002
Down	mitotic cell division (M6)	0	0.00011
Down	PLK1 signaling events (M4.2)	0	0.00031
Down	cell division - E2F transcription network (M4.8)	0,00001	0.00052
Down	cell cycle and transcription (M4.0)	0,00003	0.00186
Down	E2F transcription factor network (M8)	0,0001	0.00469
Down	cell cycle (III) (M103)	0.00011	0.00469
Down	mitotic cell cycle (M4.7)	0.00013	0.00483
Down	cell division (stimulated CD4+ T cells) (M46)	0.00019	0.00645
Down	G protein coupled receptors cluster (M155)	0.00029	0.00887
Down	E2F1 targets (Q4) (M10.1)	0.00031	0.00887
Down	mitotic cell cycle in stimulated CD4 T cells (M4.9)	0,0004	0.01062
Down	T cell surface signature (S0)	0.00058	0.01421
Down	olfactory receptors (M228)	0.00081	0.01845
Down	cell cycle (II) (M4.10)	0.00142	0.02898
Down	cell division in stimulated CD4 T cells (M4.6)	0.00163	0.03102
Down	mitotic cell cycle in stimulated CD4 T cells (M4.5)	0.00257	0.04401
Down	mitotic cell cycle - DNA replication (M4.4)	0.00285	0.04437
Down	mismatch repair (I) (M22.0)	0.00351	0.05234
Down	Wnt signaling pathway (M206)	0.00385	0.05502
Down	mitotic cell cycle in stimulated CD4 T cells (M4.11)	0.00439	0.06028
Down	C-MYC transcriptional network (M4.12)	0.00535	0.06794
Down	Ran mediated mitosis (M15)	0.00567	0.06886
Down	platelet activation and degranulation (M85)	0.00707	0.0782
Down	CD4 T cell surface signature Th1-stimulated (S6)	0.00879	0.09417
Down	E2F1 targets (Q3) (M10.0)	0.01034	0.1075
Down	Rho GTPase cycle (M4.14)	0.01451	0.13829
Down	DNA repair (M76)	0.01857	0.16782
Down	metabolism of steroids (M225)	0.01859	0.16782
Down	cell adhesion (M51)	0.01926	0.1694
Down	enriched in nuclear pore complex interacting proteins (M247)	0.02224	0.19069
Down	enriched in B cells (II) (M47.1)	0.02369	0.19821
Up	enriched in NK cells (I) (M7.2)	0.00144	0.02898
Up	DC surface signature (S5)	0.01121	0.10982
Up	putative targets of PAX3 (M89.1)	0.02865	0.22796
Up	RA, WNT, CSF receptors network (monocyte) (M23)	0.02924	0.22796
Up	B cell development (M9)	0.03019	0.2301
Up	innate activation by cytosolic DNA sensing (M13)	0.03295	0.2421
Up	plasma cells, immunoglobulins (M156.1)	0.03317	0.2421