Common virulence gene expression in naïve and severe malaria cases

Jan Stephan Wichers[1,2], Gerry Tonkin-Hill[3], Thorsten Thye[4], Ralf Krumkamp[4,5], Jan Strauss[1,2,6], Heidrun von Thien[1], Benno Kreuels[4], Judith A.M. Scholz[1], Helle Holm Hansson[7], Rasmus Weisel Jensen[7], Louise Turner[7], Freia-Raphaella Lorenz[8], Anna Schöllhorn[8], Iris Bruchhaus[1,6], Egbert Tannich[4,5], Rolf Fendel[8,9], Thomas D. Otto[10], Thomas Lavstsen[7], Tim-Wolf Gilberger[1,2,6], Michael F. Duffy[11], Anna Bachmann[1,2,5, #]

[1] Molecular Biology and Immunology, Bernhard Nocht Institute for Tropical Medicine, 20359 Hamburg, Germany
[2] Centre for Structural Systems Biology, 22607 Hamburg, Germany
[3] Wellcome Sanger Institute, Hinxton/Cambridge CB10 1SA, UK
[4] Epidemiology and Diagnostics, Bernhard Nocht Institute for Tropical Medicine, 20359 Hamburg, Germany
[5] German Center for Infection Research (DZIF), partner site Hamburg-Borstel-Lübeck-Riems, Germany
[6] Biology Department, University of Hamburg, 22609 Hamburg, Germany
[7] CMP, University of Copenhagen, 2200 Copenhagen, Denmark
[8] Institute of Tropical Medicine, University of Tübingen, 72074 Tübingen, Germany
[9] German Center for Infection Research (DZIF), partner site Tübingen, Germany
[10] Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow G12 8TA, UK
[11] Department of Microbiology and Immunology, University of Melbourne, Melbourne/Parkville VIC 3052, Australia

corresponding author

Running title: In host expression of *var* genes

Abstract

Sequestration of *Plasmodium falciparum*-infected erythrocytes to host endothelium through the parasite-derived *Pf*EMP1 adhesion proteins, is central to the development of malaria pathogenesis. *Pf*EMP1 proteins have diversified and expanded to encompass many sequence variants conferring the same array of human endothelial receptor binding phenotypes. Here, we analyzed RNA-seq profiles of parasites isolated from 32 infected travelers returning to Germany. Patients were categorized into either malaria naïve (n=15) or pre-exposed (n=17), and into severe (n=8) or non-severe (n=24) cases. Expression analysis of *Pf*EMP1-encoding *var* genes showed that severe malaria was associated with expression of *Pf*EMP1 containing the endothelial protein C receptor (EPCR)-binding CIDRα1 domain, whereas CD36-binding *Pf*EMP1 was linked to non-severe malaria outcomes. In addition, gene expression-guided determination of parasite age, suggested that circulating parasites from non-severe malaria patients were older than parasites from severe malaria patients. First-time infected patients were also more likely to develop severe symptoms and tended to be infected for a longer period, which thus appeared to select for parasites with more sequestration-efficiency and therefore more pathogenic *Pf*EMP1 variants.

Keywords: *P. falciparum*, *Pf*EMP1, RNA-seq, transcriptomics, variant surface antigens

Introduction

Despite considerable efforts during recent years to combat malaria, this disease remains a major threat to public health in tropical countries. The most severe clinical courses of malaria are due to infections with the protozoan species *Plasmodium falciparum.* In 2018, there were 228 million cases of malaria worldwide, resulting in more than 400,000 deaths (WHO, 2019). Currently, about half of the world population is living at risk of infection and more than 90% of the malaria deaths occur in Africa. In particular, children under the age of five and pregnant women suffer from severe disease. The virulence of *P. falciparum* is linked to the infected erythrocytes binding to endothelial cell surface molecules expressed on blood vessel walls. This phenomenon, known as sequestration, prevents the passage of infected erythrocytes through the spleen, which would otherwise remove the infected erythrocytes from the circulation and kill the parasite (Saul, 1999). The membrane proteins mediating sequestration are exposed to the host's immune system and through evolution *P. falciparum* parasites have acquired several multi-copy gene families coding for variant surface antigens (VSAs) allowing immune escape through extensive sequence polymorphisms. Endothelial sequestration is mediated by the *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1) family, which members have different binding capacities for host vascular tissue receptors such as CD36, EPCR,

ICAM-1, PECAM1, receptor for complement component C1q (gC1qR) and CSA. *Pf*EMP1 proteins are known to mediate adhesion of infected erythrocytes to the linings of small blood vessels (Magallón-Tejada *et al*, 2016; Turner *et al*, 2013; Rowe *et al*, 2009). The long, variable, extracellular *Pf*EMP1 region responsible for receptor binding contains a single N-terminal segment (NTS main classes A, B and pam) and a variable number of different Duffy-binding like (DBL main classes are DBLα-ζ and pam) and cysteine-rich inter-domain region domains (CIDR main classes are CIDRα-δ and pam) domains (Rask *et al*, 2010). Based on recent findings, the sub-classification within main domain classes, e.g. the DBLβ subclasses 1 – 13, was questioned due to recombination occurring frequently between members of the different subclasses (Otto *et al*, 2019). *Pf*EMP1 molecules have been grouped into four categories (A, B, C and E) depending on the protein domain composition as well as the 5' upstream sequence, the chromosomal localization and the direction of transcription of their encoding *var* genes (Rask *et al*, 2010; Kyes *et al*, 2007; Kraemer & Smith, 2003; Lavstsen *et al*, 2003). Each parasite possesses about 60 *var* genes with approximately the same distribution over the different groups (Rask *et al*, 2010). About 10% of the genes belong to A-type *var* genes, typically encoding longer *Pf*EMP1 proteins with a head-structure containing a DBLα1 and either a CIDRα1 or a CIDRβ/γ/δ domain. In some A-type proteins, an ICAM-1-binding DBLβ1 or 3 domain follows this head-structure (Lennartz *et al*, 2017). Two conserved subfamilies also belong to the group A: the *var1* gene present in two different allele forms in the parasite population (3D7- and IT-type) (Otto *et al*, 2019) and the shortest *var3* gene family with only two extracellularly exposed domains (DBLα1.3 and DBLε8). Most *Pf*EMP1 proteins belong to the B- and C-families and have a DBLα0-CIDRα2-6 head structure attached to a DBLδ1-CIDRβ/γ domain combination. Some B-type proteins possess a DBLα2-CIDRα1 head-structure typically followed by additional domains including DBLβ12 domains suggested to bind gC1qR (Magallón-Tejada *et al*, 2016). Other B-type *Pf*EMP1 have a ICAM-1-binding DBLβ5 domain (Lennartz *et al*, 2019). The VAR2CSA *Pf*EMP1 binds placental CSA and cause pregnancy-associated malaria. The *var2csa* genes constitute the group E and most *P. falciparum* isolates possess a single or two gene copies of this inter-strain conserved *var* gene variant. Based on the head structure domains, *Pf*EMP1 divide into those with DBLα1 (A-type *Pf*EMP1), DBLα2 (B-type *Pf*EMP1) or DBLα0 (B- and C-type *Pf*EMP1) domains as well as those binding CD36 via CIDRα2-6 domains (B- and C-type *Pf*EMP1) or EPCR via CIDRα1 (found in A- and B-type *Pf*EMP1) (Otto *et al*, 2019). Accordingly, the head structure confers mutually exclusive binding properties, either to EPCR, CD36 or to an unknown receptor via the CIDRβ/γ/δ domain of some A-type proteins. Some domain sequence variants are found to often co-occur (Otto *et al*, 2019; Berger *et al*, 2013; Rask *et al*, 2010). For example, one domain cassette (DC), DC8, includes specific CIDRα1.1/8 subtypes capable of binding EPCR.

Group A *var* gene expression has been associated with severe forms of malaria, whereas mild malaria may be associated with group C expression (Avril *et al*, 2012; Claessens *et al*, 2012; Lavstsen *et al*, 2012; Falk *et al*, 2009; Warimwe *et al*, 2009; Kyriacou *et al*, 2006; Rottmann *et al*, 2006; Jensen *et al*, 2004; Kirchgatter & Del Portillo, 2002). Conflicting results were reported on group B expression during severe disease; however, this is likely explained by the fact that a subset of B-type *Pf*EMP1 share EPCR and ICAM-1 receptor-binding phenotypes with A-type *Pf*EMP1. Indeed, consensus from a range of gene expression studies is that severe malaria is associated with expression of *Pf*EMP1 with EPCR-binding CIDR α1 domains (Jespersen *et al*, 2016; Kessler *et al*, 2017; Storm *et al*, 2019; Shabani *et al*, 2017; Mkumbaye *et al*, 2017; Bernabeu *et al*, 2016; Magallón-Tejada *et al*, 2016). No other domain has been consistently associated with severity of disease, but elevated expression of some specific DCs has also been associated with severe disease, in particular the EPCR-binding DC8 (DBLα2-CIDRα1.1-DBLβ12-DBLγ4/6), DC13 (DBLα1.7-CIDRα1.4) and DC4 (DBLα1.4-CIDRα1.6-DBLβ3), but also the DC5 found in A-type *Pf*EMP1 (DBLγ12-DBLδ5-CIDRβ3/5) and DC6 (DBLγ14,-DBLζ5-DBLε4) found in both A- and B-type *Pf*EMP1 (Bernabeu *et al*, 2016; Magallón-Tejada *et al*, 2016; Berger *et al*, 2013; Avril *et al*, 2013, 2012; Claessens *et al*, 2012; Lavstsen *et al*, 2012). In this study, we used an RNA-seq based analysis to study gene expression with special emphasis on *var* genes in parasites from hospitalized travelers returning from malaria endemic countries with certified *P. falciparum* infections. Individuals were clustered into i) first-time infected and ii) pre-exposed individuals on the basis of serological data or into iii) severe and iv) non-severe cases according to medical reports. Our multi-dimensional analysis reveals a clear association of domain cassettes with EPCR-binding properties with a naïve immune status and severe malaria, whereas CD36-binding *Pf*EMP1 proteins and the conserved *var1*-3D7 allele were expressed at higher levels in pre-exposed patients and non-severe cases. Interestingly, circulating parasites from severe cases tended to be younger than parasites from non-severe cases, indicating that the EPCR-binding phenotype confers more efficient sequestration of infected erythrocytes.

Results

Cohort characterization

This study is based on a cohort of 32 malaria patients hospitalized in Hamburg, Germany. Parasitemia, recorded clinical symptoms and patient sub-grouping are summarized in Table 1. For ten patients, the present malaria episode was their first recorded *P. falciparum* infection. Nine individuals had previously experienced malaria episodes according to the medical reports, whereas malaria exposure was unknown for 13 patients. In order to determine if patients already had an immune response to *P. falciparum* antigens, indicative of previous exposure to malaria, plasma samples were analyzed by a Luminex-based assay displaying the antigens

AMA1, MSP1 and CSP (Table S1). Immune responses to AMA1 and MSP1 are known to be long-lasting and seroconversion to AMA1 is assumed to occur after only a single or very few malaria infections (Drakeley *et al*, 2005). Principal component analysis of the Luminex data resulted in separation of the patients into two discrete groups corresponding to first-time infected adults ('naïve cluster') and malaria pre-exposed individuals ('pre-exposed cluster') (Figure 1A). The 13 patients with unknown malaria exposure status could be grouped into either the naïve or pre-exposed groups defined by the PCA of the antigen reactivity. The only outlier in the clustering was a 19-year-old patient (#21) from Sudan, who reported several malaria episodes during childhood, but clustered with the malaria-naïve patients.

Plasma samples were further subjected to i) a merozoite-directed antibody-dependent respiratory burst (mADRB) assay (Kapelski *et al*, 2014), ii) a PEMS-specific ELISA and iii) a protein microarray with 228 *P. falciparum* antigens (Borrmann, 2020). Analysis of these serological assays in relation to the patient clustering confirmed the expected higher and broader antigen recognition by ELISA, protein microarray, and stronger ability to induce burst of neutrophils by serum from the group of malaria pre-exposed patients (Figure 1B–D, Table S1). Data from all the serological assays were next used for an unsupervised random forest machine learning approach to build models predictive of individual's protective status. This algorithm confirmed the classification of patient #21 as being non-immune (Figure 1E). The calculated variable importance highlighted the relevance of the mADRB assay, the ELISA and the Luminex to allocate patients into cluster (Figure 1F). A multidimensional scaling plot was used to visualize cluster allocation and patient #26 positioned at the borderline to pre-exposed patients (Figure 1E). The patient was grouped into the naïve cluster in accordance with the Luminex data and the medical report showing that this patient from Jamaica was infected during his first trip to Africa.

Using protein microarrays, the antibody response against described antigens was analyzed in detail. As expected, pre-exposed individuals showed significantly elevated IgG antibody responses against a wide range of parasite antigens, especially typical blood stage markers, including MSP1, MSP2, MSP4, MSP10, EBA175, ring exported protein 1 (REX1), and AMA1 (Figure 1D, upper panel). Markers for a recent infection, MSP1, MSP4, GLURP and Early transcribed membrane protein 5 (ETRAMP5) (Van Den Hoogen *et al*, 2019), were significantly elevated in the pre-exposed individuals in comparison to the defined first-time infected group. In addition, further members of the ETRAMP family, including ETRAMP10, ETRAMP14, ETRAMP10.2 and ETRAMP4, and also antibodies against pre-erythrocytic antigens, such as CSP, STARP and LSA3, were highly elevated. Similar effects were detectable for IgM antibodies; previous exposure to the malaria parasite led to higher antibody levels (Figure 1D, lower panel).

Based on medical reports eight patients from the malaria-naïve group were considered as more severe cases, having a substantially higher parasitemia and an impaired function of diverse organs especially the brain (Table 1). Patient #1 was included into the severe group due to increasing parasitemia during hospitalization and circulating schizonts indicative of a very high sequestering parasite biomass associated with severity (Bernabeu *et al*, 2016). The remaining 24 cases were summarized into the non-severe malaria group. Comparing the IgG antibody response of severe and non-severe cases within the previously malaria-naïve group, elevated antibody levels were detected in the severe subgroup. The highest fold change was observed in antibodies directed against intracellular proteins, such as DnaJ protein, GTPase-activating protein or heat shock protein 70 (Figure S1). Interestingly, IgM antibodies against ETRAMP5 were detectable in the severely infected individuals, suggesting they were infected for a prolonged period of time compared to the mild malaria population (Helb *et al*, 2015; Van Den Hoogen *et al*, 2019).

RNA-seq transcriptomics

Parasites were isolated from the venous blood of all patients for subsequent transcriptional profiling. Transcriptome libraries were sequenced for all 32 patient samples. The number of trimmed reads ranged between 29,142,684 and 82,000,248 (median: 41,383,289) within the individual libraries derived from patients. The proportion of total reads specific for *P. falciparum* were 87.7% (median; IQR: 76.7–91.3). Lower percentages (12.4% and 15.68%) were obtained only for patient isolates #1 and #2 (Table S2). These samples were not subjected to globin-mRNA depletion due to their low RNA content after multiple rounds of DNase treatment. Consequently, less than one million *P. falciparum* reads were obtained for each of these samples. Therefore, samples from patient #1 and #2 were omitted from assembly due to low coverage, but included in the differential gene expression analysis.

Variation in parasite ages in the different patient samples was analyzed with a mixture model in accordance to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) using published data from López-Barragán *et al.* as a reference (López-Barragán *et al*, 2011). Parasites from first-time infected and pre-exposed patients revealed no obvious difference in the proportion of the different parasite stages (Figure 2A) or in median age (Figure 2B, Table 2). However, a small bias towards younger parasites in the severe cases was observed with a median age of 8.2 hpi (IQR: 8.0–9.8) in comparison to 9.8 hpi in the non-severe cases (IQR: 8.2–11.4) (Figure 2C,D, Table 2). None of the samples revealed high proportions of late trophozoites (all <3%), schizonts (0%) or gametocytes (all <6%) (Figure 2A, C).

Differential* var *gene expression on gene, domain and homology block level

To correlate individual *var* genes with a naïve immune status or disease severity differential *var* gene expression was analyzed according to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018). First, *var* gene assembly was performed by assembling each sample separately, which reduces the risk for generating false chimeric genes and results in longer contigs compared to a combined all sample assembly approach. In total, 6,441 contigs with over 500 bp-length were generated with an N50 of 2,302 bp and a maximum length of 10,412 bp (Data S1). For 5,488 contigs *Pf*EMP1 domains could be annotated, the remaining contain only homology blocks defined by Rask *et al.* (Rask *et al*, 2010) (Table S3, S4).

Var allele level:

A median of 200 contigs (IQR: 137–279) with >500 bp was assembled per sample. Almost half of the transcripts, for which a *Pf*EMP1 domain annotation could be made (47.2%: 2,592), showed >98% nucleotide identity for at least 80% of the length of a *var* gene in the varDB (Otto, 2019). Moreover, 203 transcripts matched with over 1 kb overlap and 98% identity to *var* genes from the 15 reference genomes (Otto *et al*, 2018a) (Table S4). The Salmon RNA-seq quantification pipeline, which identifies equivalence classes allowing reads to contribute to the expression estimates of multiple transcripts, was used to estimate expression levels for each transcript. As a result, it accounts for the redundancy present in our whole set of *var* gene contigs from all separate sample-specific assemblies. We compared this approach with Corset which previously has been used to investigate differential expression of *var* genes in severe malaria and found it gave similar results (see methods) (Tonkin-Hill *et al*, 2018). Due to the high diversity in *var* genes both of these approaches are only able to identify significant associations between transcripts and phenotypes when there is sufficient similarity within the associated sequences.

By comparing transcripts expressed in first-time infected patients with those from parasites isolated from pre-exposed patients using the Salmon approach, transcript levels were higher for twelve and lower for three genes in malaria-naïve hosts (Table 3, Figure 3A, B, Table S6). Assembled alleles of the conserved subfamilies, *var1*, *var2csa* and *var3*, were also found at higher frequencies in first-time infected patients. Notably, the *var1-IT* allele was expressed at higher levels in parasites from first-time infected patients, whereas the *var1-3D7* allele was expressed at higher levels in parasites from pre-exposed and non-severe patients (Figure 3, Table 3, 4). This was confirmed by mapping normalized reads from all patients to the *var1-3D7* and *var1-IT* allele forms (Figure S2). Several *var* fragments from B- or C-type *var* genes were associated with a naïve immune status and three transcripts from A, DC8 and B-type *var* genes as well as *var2csa* were linked to severe malaria patients (Figure 3, Table 3, 4, Table S6, S7).

Domain level:

Similar to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018), we quantified domain level expression by aligning reads to their single best 'hit' in the combined assembly. Domains were identified within each transcript and the sum of the read counts corresponding to domains with the same classification was calculated to provide domain level read counts. This showed that different EPCR-binding CIDRα1 domain variants and other domains found in DCs with CIDRα1 domains were expressed at significantly higher levels in first-time infected patients (Table 5, Figure 4A, B, Table S6). Thus, besides domains from DC8 (DBLα2, CIDRα1.1, DBLβ12) and DC13 (DBLα1.7, CIDRα1.4), a single domain from DC15 (DBLα1.2) were increased upon infection of malaria-naïve individuals. The DBLα1.2 domain was in all of the 32 gene assemblies with adjacent CIDR domains annotated linked to an EPCR binding CIDRα1 domain, 56% of these were a CIDRα1.5 domain (Table S3, S4). The CIDRα1.6 domain from DC4 failed to reach statistical significance (p_{adj}=0.07), but was 2.25x times higher expressed in parasites infecting naïve patients (Table S6). In addition to domains associated with EPCR-binding *Pf*EMP1, parasites from first-time infected patients showed significantly more transcripts encoding the CIDRδ1 domain of DC16 (DBLα1.5/6-CIDRδ1/2) (Table 5, Figure 4). In general, all domains associated with the same domain cassettes showed the same trend even if some domains failed to reach statistical significance (Table S6). Moreover, the DBLβ6 domain was among the top hits of significantly higher expressed domains in the naïve patient cluster. The DBLβ6 is associated with A-type *var* genes and often found adjacent to DC15 and DC16 (Otto *et al*, 2019) thus supporting the association of both domain cassettes with a naïve immune status. Domain types found expressed at lower levels in malaria-naïve included several domains from the *var1-3D7* allele (DBLα1.4, DBLγ15, DBLε5) as well as NTSB and N-terminal head domains from B- and C-type *Pf*EMP1 (DBLα0.13/22/23) with CD36-binding CIDR domains (CIDRα2.8/9,6) and the C-terminal CIDRγ11 domain (Table 5, Figure 4A, B).

When comparing the severe sample set to the non-severe, domains of DC8 (DBLα2, CIDRα1.1, DBLβ12) and DC15 (DBLα1.2) were found associated with severe disease (Table 6, Figure 4C, D, Table S7). The DC16 consists either of a DBLα1.5 or 1.6 attached to a CIDRδ1 domain. The DBLα1.6 domain was found expressed at higher levels in severe malaria patients whereas the DBLα1.5 domain was found to be highly expressed in non-severe cases. The CIDRδ1 domain showed no association with disease group, and the DBLα1.5 domain type was generally expressed at a very high level in multiple patient isolates (Table 6, Figure 4). As observed for pre-exposed individuals, domain types expressed at significantly higher levels in non-severe cases included the CIDRα1.3 domain from the *var1-3D7* allele as well as N-terminal head domains from B- and C-type *Pf*EMP1 with CD36-binding capacity (DBLα0.23, CIDRα2.4/9)(Table 6, Figure 4C, D).

Since the subclassification of domains is debatable (Otto *et al*, 2019) and different domain subclasses confer the same binding phenotype (Lau *et al*, 2015), the domains of the N-terminal head structure were grouped according to their binding phenotype and the normalized read counts (TPM) were summarized for each patient (Figure 4E, F). This showed clear differences were observed for DBL and CIDR domains associated with EPCR- or CD36-binding *Pf*EMP1. As expected, the EPCR-binding phenotype as well as the CIDRγ3 domain were associated with the naïve and more severe cases, the CD36-binding phenotype with the pre-exposed and non-severe patients.

Homology block level:

Within *Pf*EMP1 sequences 628 homology blocks were defined (Rask *et al*, 2010) and 613 were available for download and subsequent analysis from the VARDOM server (Tonkin-Hill *et al*, 2018). Homology block expression levels were obtained by aggregating read counts for each block after first identifying all occurrences of the block within the combined transcript assembly. Transcripts encoding blocks number 255, 584 and 614, all typically located within DBLβ domains of DC8 and CIDRα1-containing type A *Pf*EMP1 (Table 7, Figure 5A, B, Table S6), number 557, located in the inter-domain region between DBLβ and a DBLγ domains (no *Pf*EMP1 type association) and block number 155 found in NTSA, were found associated with a naïve immune status. All these blocks are found in A-, B/A- or B-type genes. Conversely, transcripts encoding block 88 from DBLα0 domains and 269 from ATSB were found at lower levels in malaria-naïve patients indicating that B- and C-type genes are more frequently expressed in pre-exposed individuals (Table 7, Figure 5A, B). Two homology blocks, 591 and 559, associated with B-type *Pf*EMP1 were found to be lower expressed in severe malaria cases (Table 8, Figure 5C, D, Table S7).

A summary of the differential *var* gene expression data on the multi-, single- and subdomain level can be found in Figure 6.

Var expression profiling by DBLα-tag sequencing

To supplement our RNA-seq analysis with an orthogonal analysis, we performed DBLα-tag RT-PCR combined with deep amplicon sequencing on 30 of our patient samples (Lavstsen *et al*, 2012). 851 to 3,368 reads with a median of 1,666 over all samples were analyzed. Identical DBLα-tag sequences were clustered to generate relative expression levels of each unique *var* gene tag. Overall, the relative expression levels were similar for sequences found in both the RNA-seq and the DBLα-tag approach with a mean log2(DBLα-PCR/RNA-seq) of 0.4 (CI of 95%: -2.5–3.3) determined by Bland-Altman plotting (Figure S3). In median, 82.6% of all detected individual DBLα-tag sequence clusters with >10 reads (92.9% of all DBLα-tag reads)

were found in the RNA-seq approach; and 81.8% of the upper 75th percentile of RNA-seq contigs (with DBLα tag sequences) were found by the DBLα-tag approach.

Unique DBLα-tag sequences were searched for near identical sequences among all known *var* genes on varDB Version V1.1 (Otto, 2019) using Varia tool (Mackenzie *et al*, 2020). Nearly identical database sequences were found for ~85% of the DBLα-tag sequences allowing prediction of these query gene's domain annotation (Table S10). In line with the RNA-seq data we found DBLα1 and DBLα2 sequences enriched in first-time infected patients and the severe malaria patients. Conversely, a significant higher proportion of DBLα0 sequences was found in pre-exposed individuals and mild cases (Figure 7A, B). No difference was observed in the number of reads or unique DBLα-tags detected between patient groups, although a trend towards more DBLα-tag clusters could be observed in first-time infected patients and severe cases (Figure 7A, B). A prediction of the NTS and CIDR domains adjacent to the DBLα domain showed a significant higher proportion of NTSA in severe cases as well as EPCR-binding CIDRα1 domains in first-time infected and severe cases. Expression of *var* genes encoding NTSB and CIDRα2-6 domains were significantly associated with in pre-exposed and non-severe cases (Figure 7A, B). Analysis of *var* expression in relation to other domains, showed *var* transcripts with DBLβ, γ and ζ or CIDRγ domains were more frequently expressed in first-time infected and severe malaria patients whereas those encoding DBLδ and CIDRβ were less frequent (Figure 7C, D).

Assessing expression in relation to domain subtype, CIDRα1.1/5, DBLβ12, DBLγ2/12 and DBLα2 were associated with severe malaria whereas CIDRα3.1/4, DBLα0.12/16, and DBLδ1 associated with non-severe cases (Figure 7D). Overall, these data corroborated the main observations from the RNA-seq analysis, confirming the association of EPCR-binding *Pf*EMP1 variants with development of severe malaria symptoms and CD36-binding *Pf*EMP1 variants with establishment of less severe infections in semi-immune individuals.

Correlation of* var *gene expression with antibody levels against head structure CIDR domains

A detailed analysis of the antibody repertoire of the patients against head structure CIDR domains of *Pf*EMP1 was carried out using a panel of 19 different EPCR-binding CIDRα1 domains, 12 CD36-binding CIDRα2–6 domains, three CIDRδ1 domains as well as a single CIDRγ3 domain (Obeng-Adjei *et al*, 2020; Bachmann *et al*, 2019). Additionally, the minimal binding region of VAR2CSA was included (Figure 8). Generally, plasma samples from malaria-naïve as well as severe cases showed lower MFI values for all antigens tested in comparison to samples from pre-exposed or non-severe cases (Figure 8A, B). Mann-Whitney U testing revealed significant differences for CIDRα2–6, CIDRδ1 and CIDRγ3, but not for EPCR-binding CIDRα1 domains.

Another way of analyzing the samples is to take the average MFI from an unrelated control cohort plus two standard deviations as a cut off for seropositivity for calculation of the coverage of antigen recognition (Cham *et al*, 2010). By doing so, almost half of the tested antigens were recognized by pre-exposed (median: 46.3%) and non-severe patients (median: 45.1%), but only 1/4 of the antigens were recognized by first-time infected patients (median: 24.4%) and 1/20 by severely ill patients (median: 4.9%). Apart from controls, antigens recognized by over 60% of the pre-exposed and/or non-severe patient sera were i) four CIDRα1 domains capable of EPCR-binding (CIDRα1.5, CIDRα1.6, CIDRα1.7 and the DC8 domain CIDRα1.8), ii) two CD36-binding CIDRα domains (CIDRα2.10, CIDRα3.1) and iii) two CIDR domains with unknown binding phenotype (CIDRδ1 and CIDRγ3) (Figure 8C, Table S1).

Differential gene expression analysis of core genes

Global gene expression analysis (Tonkin-Hill *et al*, 2018) identified 420 genes to be higher and 236 to be lower expressed in first-time infected patients, together corresponding to 11.3% of all *P. falciparum* genes (Table S9). Analysis of gene set enrichment analysis (GSEA) of GO terms and KEGG pathways showed a significantly lower expression level for genes with several GO terms involved in antigenic variation and host cell remodeling in first-time infected patients (Figure 9A, Table S9). These analysis results may be distorted, since variant surface antigens like *var*, *rif*, *stevor*, *surf* and *pfmc-2tm* are largely clone-specific and reads from the clinical isolates would map only to homologous regions in 3D7 genes, which may actually not be present in the clinical isolate. Therefore, we manually screened differentially expressed genes known to be involved in *var* gene regulation or correct display of *Pf*EMP1 at the host cell surface (Table S9). On single gene level, genes encoding the Maurer's cleft proteins MAHRP1 and REX2 (Spycher *et al*, 2003; Spielmann *et al*, 2006), REX4 and MSRP7 located within the host cell cytosol (Spielmann *et al*, 2006; Heiber *et al*, 2013), the erythrocyte membrane-located glycophorin binding protein GBP130 (Perkins, 1988) as well as Sir2a and SWIB known to be involved in *var* gene regulation (Tonkin *et al*, 2009; Wang & Zhang, 2020) were expressed at significantly lower levels in malaria-naïve patients. MAHRP1 is essential for translocation of *Pf*EMP1 to the surface of infected erythrocytes (Spycher *et al*, 2006, 2008) and was suggested to be part of the *Pf*EMP1 loading hub (McHugh *et al*, 2020); on the contrary, deletion of the REX2 and REX4-encoding genes via chromosome breakage was associated with the loss of cytoadherence, but not with aberrant trafficking of *Pf*EMP1 (Nacer *et al*, 2011; Chaiyaroj *et al*, 1994; Day *et al*, 1993). Genetic ablation of GBP130 increased the membrane rigidity of infected erythrocytes without negative impact on cytoadherence under flow conditions (Maier *et al*, 2008).

In contrast, several PHIST encoding genes were found expressed at higher levels in first-time infected patients including the lysine-rich membrane-associated PHISTb protein (LyMP),

previously reported to interact with the ATS domain (Oberli *et al*, 2014, 2016). Furthermore, the exported proteins FIKK9.6 (Nunes *et al*, 2007), MSRP5 and MSRP6 (Heiber *et al*, 2013) as well as PF3D7_0721100 showed a significant increase compared to pre-exposed patients. The conserved *Plasmodium* protein of unknown function PF3D7_0721100 detected in detergent-resistant fractions of the red blood cell membrane (Yam *et al*, 2013) was also found in a putative *Pf*EMP1 unloading hub together with REX1, MAHRP2, PTP5 and PF3D7_1353100 using protein interaction network analysis (McHugh *et al*, 2020). On the level of *var* gene expression regulation, the chromatin-associated exoribonuclease *Pf*RNase II (PF3D7_0906000) was increased in first-time infected patients (Zhang *et al*, 2014).

Most of the associations with a naïve immune status were also observed in the comparison of severe versus non-severe cases (Table S10). Significantly higher expresses genes included several PHIST and HYP proteins, MSRP5, PfRNase II, PF3D7_0721100 and, additionally, SET1, PTP1, SBP1, PTP5 and PF3D7_1353100. Overall, three proteins defined by McHugh *et al*. as members of the *Pf*EMP1 unloading hub were significantly higher expressed in severe malaria cases (PTP5, PF3D7_0721100 and PF3D7_1353100) and all remaining proteins showed the same trend for upregulation (PIESP2, PfJ23, REX1 and MAHRP2) (McHugh *et al*, 2020).

Furthermore, the KEGG pathway 03410 ‘base excision repair’ facilitating the maintenance of the genome integrity by repairing small bases lesions in the DNA was expressed at significantly higher levels in first-time infected patient samples (Figure 9A, Figure S4). In total, six out of 15 *P. falciparum* genes included into the KEGG pathway were found to be statistically significant enriched upon first-time infection, more precisely the putative endonuclease III (PF3D7_0614800) from the short-patch pathway and the putative A-/G-specific adenine glycosylase (PF3D7_1129500), the putative apurinic/apyrimidinic endonuclease Apn1 (PF3D7_1332600), the proliferating cell nuclear antigens 1 (PF3D7_1361900), the catalytic (PF3D7_1017000) and small (PF3D7_0308000) subunits from the DNA polymerase delta from the long-patch pathway (Figure 9B).

Discussion

Analysis of blood samples from travelers returning to Germany and hospitalized with *P. falciparum* malaria for the first time, allows studies of parasites’ development and gene expression unaffected by host immune responses elicited in previous Plasmodia infections. Here we analyzed gene expression in 32 such patients using direct *ex vivo* RNA-seq to exclude transcriptional adaptation through *in vitro* cultivation. Most of these returning travelers studied were infected with a single or very few parasite genotypes, which likely simplified the analysis of *var* gene expression due to the restricted genomic repertoire. We were able to distinguish

retrospectively between first-time infected and pre-exposed patients by medical history and assessing the presence of antibodies to *P. falciparum* antigens. Eight out of 15 first-time infected patients were classified as severe malaria cases due to impaired organ function according to their medical reports. Despite the relatively low number of patient samples, the RNA-seq approach confirmed previously reported associations between transcripts encoding type A and B EPCR-binding *Pf*EMP1 and infections in naïve hosts and disease severity (Table S6, S7)(Duffy *et al*, 2019; Tonkin-Hill *et al*, 2018; Kessler *et al*, 2017; Bernabeu *et al*, 2016; Jespersen *et al*, 2016; Lavstsen *et al*, 2012).

Overall, there was a high degree of consensus between analyses of the *var* transcriptome data in relation to the different levels of *Pf*EMP1 domain annotation. Stratifying *var* gene expression according to different main and subtype of DBL and CIDR domains, showed only A- and DC8-type *Pf*EMP1 domains, and predominantly those linked to EPCR-binding *Pf*EMP1, to be associated with first-time infections. Conversely, domains typical for CD36-binding *Pf*EMP1 proteins was found at higher levels in malaria-experienced patients. Specifically, expression of *Pf*EMP1 domains included in DC8, DC13 and DC15 as well as all EPCR-binding CIDRα1 domains were associated with first time infections, whereas DBLα0 and CD36-binding CIDRα2-6 domains were linked to pre-exposed individuals. These differences were in large due to the differential expression between the first-time infected patients with more severe symptoms and patients with non-severe malaria. Here, domains of DC8 and DC15 as well as all DBLα1/2 and CIDRα1 domains were associated with severe symptoms, while NTSB, DBLα0, CIDRα2-6 domains including specific subsets of CIDRα2 were linked to non-severe symptoms. These conclusions were closely mirrored in the DBLα tag analysis, and was further corroborated by the differential RNA-seq expression stratified according to the smaller homology blocks, which identified mainly homology blocks of DBLβ1, 3, 5 and 12 DBLβ domains to be associated with first- time infected patients. These DBLβ domains are parts of DCs associated with EPCR-binding, so it is hard to distinguish between co-occurring domains and clear associations. DBLβ domains do not segregate distinctly by sequence similarity into groups reflecting their observed binding to ICAM-1 and gC1qR, (Otto *et al*, 2019). Best defined is ICAM-1 binding DBLβ5 domains found in CD36-binding B-type *Pf*EMP1 (Janes *et al*, 2011; Lennartz *et al*, 2019) and ICAM1-binding of and DBLβ1 and 3 domains found in EPCR-binding type-A *Pf*EMP1 (Lennartz *et al*, 2017). The relative importance of ICAM1 binding to CD36 or EPCR binding *Pf*EMP1 is not well understood, but ICAM-1-binding is believed to contribute to disease severity by either tethering endothelial binding (Bernabeu *et al*, 2019) or initiating or securing endothelial sequestration on inflamed endothelium, which is likely to shed EPCR (Jensen *et al*, 2020)

In addition, three other A-type-associated domains, CIDRγ3, CIDRδ from the DC16 and DBLβ9 from DC5, were found associated with first-time infected patients. DC5 could have

been detected due to its presence C-terminally to some EPCR-binding *Pf*EMP1. However, the CIDRδ domain of DC16 (DBLα1.5/6-CIDRδ1/2) constitute a different subset of A-type *Pf*EMP1, which together with A-type *Pf*EMP1 with CIDRβ2 (found in DC11) or CIDRγ3 domains may be associated with rosetting (Carlson *et al*, 1990; Ghumra *et al*, 2012). Direct evidence that any of these CIDR domains have intrinsic rosetting properties is lacking (Rowe *et al*, 2002). Rather, their association with rosetting may be related to their tandem expression with DBLα1 at the N-terminal head (Ghumra *et al*, 2012). The CIDRδ domain was not associated with severe malaria patient group and the two different DC16-associated DBLα1 domains were found associated with severe and non-severe malaria, respectively. However, CIDRγ3 expression was low, but it was found at higher levels in severe malaria patients.

The DC16 group A signature was not associated with severe disease outcome in previous DBLα-tag studies or qPCR studies by Lavstsen *et al.* (Lavstsen *et al*, 2012) and Bernabeu *et al.* (Bernabeu *et al*, 2016), but DBLα1.5/6 and CIDRδ of DC16 were enriched in cerebral malaria cases with retinopathy in the study of Shabani *et al.* (Shabani *et al*, 2017) and Kessler *et al.* (Kessler *et al*, 2017) using the same qPCR primer set. Also, association of DC11 with severe malaria in Indonesia was found using the same RNA-seq approach as used here (Tonkin-Hill *et al*, 2018). Rosetting is thought to enhance microvascular obstruction but the role of rosetting in severe malaria pathogenesis remains unclear (McQuaid & Rowe, 2020). Together with previous observations, our data suggest that pediatric cerebral malaria infections are dominated by the expansion of parasites expressing EPCR-binding domains accompanied by parasites expressing other group A *Pf*EMP1, possibly rosetting variants.

To the best of our knowledge, this study is the first description of expression differences between the two *var1* alleles, 3D7 and IT. At the transcript level the *var1-IT* allele was found to be enriched in parasites from first-time infected patients; conversely, several transcripts covering almost the full-length protein and in total half of the domains from the *var1*-3D7 allele were increased in pre-exposed and non-severely ill patients. Expression of the *var1* gene was previously observed to be elevated in malaria cases imported to France with an uncomplicated disease phenotype (Argy *et al*, 2017). In general, the *var1* subfamily is ubiquitously transcribed (Winter *et al*, 2003; Duffy *et al*, 2006), atypically late in the cell cycle after transcription of *var* genes encoding the adhesion phenotype (Kyes *et al*, 2003; Duffy *et al*, 2002) and is annotated as a pseudogene in 3D7 due to its premature truncation. Similarly, numerous isolates display frame-shift mutations often in exon 2 in the full gene sequences (Rask *et al*, 2010). However, none of these studies addressed differences in the two *var1*-alleles just recently described by comparing *var* gene sequences from 714 *P. falciparum* genomes (Otto *et al*, 2019) and to date it is unclear if both allele forms fulfill the same function or harbor the same characteristics previously described. Overall, the *var1* gene – and the first 3.2 kb of the 3D7 allele in particular – seems to be under high evolutionary pressure (Otto *et al*, 2019) and the bi-allelic pattern can

be traced back before the split of *P. reichenowi* from *P. praefalciparum* and *P. falciparum* (Otto *et al*, 2018b). Our data indicate that the two alleles, 3D7 and IT, may have different roles during disease, however, this remains to be determined in future studies.

In general, data from immunologically naïve malaria patients are rather limited, restricting our comparison mainly to the severe phenotype described in numerous previous studies. However, a recent study explored the *var* gene expression during infancy in Kenyan children and could correlate the waning of maternal antibodies with increasing transcription of DC8, DC13 and A-type *var* genes in general. After the first year of life the amount of these transcripts decreases with age and acquired immunity (Kivisi *et al*, 2019). A high expression of A-type *var* genes in naïve malaria cases imported to France and an association of DC4, 8 and 13 with disease severity has also been reported (Argy *et al*, 2017). Both qPCR studies are in agreement with the RNA-seq data from our cohort of immunologically malaria-naïve adults, but we could extend the list with DC15 and DC16, that are presumably involved in binding of infected erythrocytes to EPCR and that may also mediate binding to uninfected red blood cells by rosetting.

Overall, most studies – including this one – are looking at differentially expressed genes. From CHMI studies we already know that at the early onset of infection the parasite population expresses a wide range of different *var* genes located in the subtelomeric regions (A- and mainly B-type *var* genes). Since B-types are also highly expressed in pre-exposed cases domains of these *Pf*EMP1s may be missing in pattern of first-time infected patients. Furthermore, EPCR- and ICAM-1 binding and rosetting-mediating variants may confer a parasite growth advantage in malaria naïve hosts, and in some circumstances increase the risk for severe malaria, so that a selection towards these variants may have been already occurred in our patients. This would fit nicely to the observation that the severe patients within the malaria-naïve patient group seem to be infected for a longer time period. Moreover, patients with preformed immunity recognize several CIDRα1 domains capable of EPCR-binding as well as the two atypical CIDR head domains CIDRδ1 and CIDRγ3 more frequently than CIDRα2-6 domains with CD36-binding affinity. This is in agreement with studies from malaria endemic setting indicating that IgG against EPCR-binding domains were acquired first followed by domains with unknown binding phenotypes associated with rosetting and CD36-binding domains. Resulting in the earlier acquisition of antibodies against DBL and CIDR domains of group A and B/A, associated with EPCR binding, than against B- and C-type domains (Obeng-Adjei *et al*, 2020; Cham *et al*, 2009, 2010; Turner *et al*, 2015).

Genes involved in *Pf*EMP1 biology were found to be expressed at lower levels in severe malaria patients (Tonkin-Hill *et al*, 2018). In concordance, GO term analysis revealed a general lower expression of genes involved in antigenic variation or host cell remodeling during first-time infection and severe disease, but the expression analysis of clonally variant genes is

546 complicated by the existence of multiple diverse families that have strain specific members
547 leading to mis-mapping of reads to genes present in the 3D7 strain that do not exist in the
548 clinical samples analyzed. Therefore, manually selected genes encoding regulators of *var*
549 gene expression and *Pf*EMP1 trafficking were additionally inspected for association with a na-
550 ïve immune status and severity. The NAD^+-dependent histone deacetylases Sir2a and Sir2b
551 remove acetyl groups from the N-terminal tails of histone 3 and 4 and are therefore considered
552 as *var* silencing factors. Sir2b regulates the most telomeric B-type *var* genes, Sir2a-regulated
553 *var* genes are of type A, C and E by deacetylation of H3K9ac, H3K14ac and H4K16ac (Duffy
554 *et al*, 2014; Tonkin *et al*, 2009). A down-regulation of Sir2a indicates an elevated expression
555 of these *var* types, which is in concordance with our data on A- and E-type *var* genes found
556 more frequently in first-time infected patients. Contradictory to a previous study showing that
557 the exoribonuclease PfRNase II controls the silencing of A-type *var* genes and was negatively
558 associated with *var*-A expression in severe malaria, PfRNase II expression was significantly
559 enriched in first-time infected and severely ill patients (Zhang *et al*, 2014). Our results also
560 contradicted down-regulation of A- and partially B-type *var* genes by conditional knockout of
561 SWIB (Wang & Zhang, 2020). However, high expression levels of the *var1*-3D7 allele, also an
562 A-type *var* gene, in pre-exposed and non-severe cases may be responsible for these opposite
563 results. On the other hand, expression of genes encoding other factors was also found inverted
564 in comparison to other studies. GBP was found to be highly enriched in severe relative to
565 uncomplicated malaria cases (Lee *et al*, 2018), but is significantly lower expressed in first-time
566 infected patients and has the same trend in severe cases in this study. Contrarily, LyMP, SBP1
567 and SET proteins were expressed on a significantly lower level in severe cases in the study of
568 Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018), but elevated levels were found in our study.
569 A recent study showed that parasites isolated from symptomatic infections were on average
570 younger than blood-circulating parasites from asymptomatic infections presumably due to a
571 more efficient sequestration of parasites in the symptomatic cases (Andrade *et al*, 2020).
572 Based on this and our findings we hypothesize that parasites circulating in severely ill patients
573 are younger due to the expression of EPCR-/ICAM-1-binding *Pf*EMP1 variants, whereas par-
574 asites in non-severe patients may circulate longer due to their expression of *Pf*EMP1 binding
575 to CD36. Although affinity of the CIDR domains to each of these receptors was shown to be
576 similar with median K_d values of 12 nM for CD36 (Hsieh *et al*, 2016) and 16.5 nM for EPCR
577 (Lau *et al*, 2015), recent papers describe a rolling binding phenotype of infected red blood cells
578 over CD36 and a static binding for EPCR and ICAM-1 under flow conditions (Lubiana *et al*,
579 2020; Bernabeu *et al*, 2019; Dasanna *et al*, 2017; Herricks *et al*, 2013). Due to their biconcave
580 shape trophozoites seem to roll faster but less stable by flipping over CD36-expressing cells,
581 whereas schizonts roll over longer distances at different shear stresses applied. This might
582 explain why young trophozoites are found in blood samples from non-severe cases, but not

583 older trophozoites or schizont stages as also previously described (Tonkin-Hill *et al*, 2018).
584 The parasite population in first-time infected individuals may have broader binding potential
585 after liver release as there is no pre-existing immunity to clear previously experienced *Pf*EMP1
586 variants, but during the blood stage infection variants with high-affinity binding to EPCR and
587 ICAM-1 binding are selected which may lead to severe symptoms. This hypothesis is sup-
588 ported by (1) the difference in parasite age between severe and non-severe malaria cases
589 calculated by matching RNA-seq data to a reference data set (López-Barragán *et al*, 2011).
590 (2) This correlates with a higher *var1* expression in parasite from non-severe and pre-exposed
591 patients, which expression is not suppressed in trophozoites (Kyes *et al*, 2003). (3) The ex-
592 pression of EPCR- & ICAM-1-binding variants in parasites from severe and first-time infected
593 patients is significantly increased, conversely transcripts of CD36-binding variants are found
594 more frequently in parasites from non-severe and pre-exposed patients. For parasites survival
595 and transmission, it may be highly beneficial to have more less virulent *Pf*EMP1 variants able
596 to bind CD36. This interaction may not, or is less likely to, lead to obstruction of blood flow,
597 inflammation and organ failure at least of the brain, where CD36 is nearly absent.
598
599 **Material and methods**
600
601 ***Ethics statement***
602 The study was conducted according to the principles of the Declaration of Helsinki in its 6th
603 revision as well as International Conference on Harmonization–Good Clinical Practice (ICH-
604 GCP) guidelines. All patients, aged 19 to 70 years, provided written informed consent for this
605 study, which was approved by the Ethical Review Board of the Medical Association of Ham-
606 burg (reference number PV3828).
607
608 ***Blood sampling and processing***
609 Blood samples from *P. falciparum* malaria patients collected either at the diagnostic unit of the
610 Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany, or at the Medical Clinic
611 and Polyclinic of University Clinic Hamburg-Eppendorf, Hamburg, Germany, were used in this
612 study. EDTA blood samples (1–30 mL) were obtained from the patients. The plasma was sep-
613 arated by centrifugation and immediately stored at -20°C. Erythrocytes were isolated by Ficoll
614 gradient centrifugation followed by filtration through Plasmodipur filters (EuroProxima) to clear
615 the remaining granulocytes. An aliquot of red blood cells (about 50–100 µl) was separated and
616 further processed for gDNA purification. At least 400 µl of purified erythrocytes were rapidly
617 lysed in 5 volumes pre-warmed TRIzol (ThermoFisher Scientific) and stored at -80°C until fur-
618 ther processing (*ex vivo* samples, n = 32).
619

620 ***Serological assays***

621 *Luminex assay*

622 The Luminex assay was conducted as previously described using the same plex of antigens
623 tested (Bachmann *et al*, 2019). In brief, plasma samples from patients were screened for indi-
624 vidual recognition of 19 different CIDRα1, 12 CIDRα2–6, three CIDRδ1 domains and a single
625 CIDRγ3 domain as well as of the controls AMA1, MSP1, CSP, VAR2CSA (VAR2), tetanus
626 toxin (TetTox) and BSA. The data are shown as mean fluorescence intensities (MFI) allowing
627 comparison between different plasma samples, but not between different antigens. Alterna-
628 tively, the breadth of antibody recognition (%) was calculated using MFI values from Danish
629 controls plus two standard deviations (SD) as cut off.

630

631 *Merozoite-triggered antibody-dependent respiratory burst (mADRB)*

632 The assay to determine the mADRB activity of the patients was set up as described before
633 (Kapelski *et al*, 2014). Polymorphonuclear neutrophil granulocytes (PMNs) from one healthy
634 volunteer were isolated by a combination of dextran-sedimentation and Ficoll-gradient centrif-
635 ugation. Meanwhile, 1.25 x 10^6 merozoites were incubated with 50 µl of 1:5 diluted plasma
636 (decomplemented) for 2 h. The opsonized merozoites were pelleted (20 min, 1500 g), re-sus-
637 pended in 25 µl HBSS and then transferred to a previously blocked well of an opaque 96 well
638 high-binding plate (Greiner Bio-One). Chemiluminescence was detected in HBSS using 83.3
639 µM luminol and 1.5 x 10^5 PMNs at 37°C for 1 h to characterize the PMN response, with read-
640 ings taken at 2 min intervals using a multiplate reader (CLARIOstar, BMG Labtech). PMNs
641 were added in the dark, immediately before readings were initiated.

642

643 *ELISA*

644 Antibody reactivity against parasitophorous vacuolar membrane-enclosed merozoite struc-
645 tures (PEMS) was estimated by ELISA. PEMS were isolated as described before (Llewellyn *et*
646 *al*, 2015). For the ELISA, 0.625 x 10^5 PEMS were coated on the ELISA plates in PBS. Plates
647 were blocked using 1% Casein (Thermo Scientific #37528) and incubated for 2 h at 37°C. After
648 washing using PBS/0.1% Tween, plasma samples were added at two-fold dilutions of 1:200 to
649 1:12800 in PBS/0.1% Casein. The samples were incubated for 2 h at room temperature (RT).
650 IgG was quantified using HRP-conjugated goat anti-human IgG at a dilution of 1:20,000 and
651 incubated for 1 h. For the color reaction, 50 µl of TMB substrate was used and stopped by
652 adding 1 M HCl after 20 min. Absorbance was quantified at 450 nm using a multiplate reader
653 (CLARIOstar, BMG Labtech).

654

655

656

Protein microarray

Microarrays were produced at the University of California Irvine, Irvine, California, USA (Doolan *et al*, 2008). In total, 262 *P. falciparum* proteins representing 228 unique antigens were expressed using an *E. coli* lysate *in vitro* expression system and spotted on a 16-pad ONCYTE AVID slide. The selected *P. falciparum* antigens are known to frequently provide a positive signal when tested with sera from individuals with sterile and naturally acquired immunity against the parasite (Obiero *et al*, 2019; Dent *et al*, 2016; Doolan *et al*, 2008; Felgner *et al*, 2013). For the detection of binding antibodies, secondary IgG antibody (goat anti-human IgG QDot™800, Grace Bio-Labs #110635), secondary IgM antibody (biotin-SP-conjugated goat anti-human IgM, Jackson ImmunoResearch #109-065-043) and Qdot™585 Streptavidin Conjugate (Invitrogen #Q10111MP) were used (Taghavian *et al*, 2018).

Study serum samples as well as the European control serum were diluted 1:50 in 0.05X Super G Blocking Buffer (Grace Bio-Labs, Inc.) containing 10% *E. coli* lysate (GenScript, Piscataway, NJ) and incubated for 30 minutes on a shaker at RT. Meanwhile, microarray slides were rehydrated using 0.05X Super G Blocking buffer at RT. Rehydration buffer was subsequently removed and samples added onto the slides. Arrays were incubated overnight at 4°C on a shaker (180 rpm). Serum samples were removed the following day and microarrays were washed using 1X TBST buffer (Grace Bio-Labs, Inc.). Secondary antibodies were then applied at a dilution of 1:200 and incubated for two hours at RT on the shaker, followed by another washing step and a one-hour incubation in a 1:250 dilution of Qdot™585 Streptavidin Conjugate. After a final washing step, slides were dried by centrifugation at 500 g for 10 minutes. Slide images were taken using the ArrayCAM® Imaging System (Grace Bio-Labs) and the ArrayCAM 400-S Microarray Imager Software.

Microarray data were analyzed in R statistical software package version 3.6.2. All images were manually checked for any noise signal. Each antigen spot signal was corrected for local background reactivity by applying a normal-exponential convolution model (McGee & Chen, 2006) using the RMA-75 algorithm for parameter estimation (available in the LIMMA package v3.28.14) (Silver *et al*, 2009). Data was log2-transformed and further normalized by subtraction of the median signal intensity of mock expression spots on the particular array to correct for background activity of antibodies binding to *E. coli* lysate. After log2 transformation data approached normal distribution. Differential antibody levels (protein array signal) in the different patient groups were determined by Welch-corrected Student's t-test. Antigens with $p<0.05$ and a fold change >2 of mean signal intensities were defined as differentially recognized between the tested sample groups. Volcano plots were generated using the PAA package (Turewicz *et al*, 2016) and GraphPad Prism 8. Individual antibody breadths were defined as number of seropositive features with signal intensities exceeding an antigen-specific threshold set at six standard deviations above the mean intensity in negative control samples.

694

695 *Unsupervised random forest model*

696 An unsupervised random forest (RF) model, a machine learning method based on multiple
697 classification and regression trees, was calculated to estimate proximity between patients.
698 Variable importance was calculated, which shows the decrease in prediction accuracy if values
699 of a variable are permuted randomly. The *k*-medoids clustering method was applied on the
700 proximity matrix to group patients according to their serological profile. Input data for random
701 forest were Luminex measurements for MSP1, AMA1 and CSP reduced by principal compo-
702 nent analysis (PCA; first principal component selected), mADRB, ELISA, and antibody breadth
703 of IgG and IgM determined by protein microarray were used to fit the RF model. Multidimen-
704 sional scaling was used to display patient cluster. All analyses were done with R (4.02) using
705 the packages randomForest (4.6-14) to run RF models and cluster (2.1.0) for k-medoids clus-
706 tering.

707

708 ***DNA purification and MSP1 genotyping***

709 Genomic DNA was isolated using the QIAamp DNA Mini Kit (Qiagen) according to the manu-
710 facturer's protocol. To assess the number of *P. falciparum* genotypes present in the patient
711 isolates, MSP1 genotyping was carried out as described elsewhere (Robert *et al*, 1996).

712

713 ***RNA extraction, RNA-seq library preparation, and sequencing***

714 TRIzol samples were thawed, mixed rigorously with 0.2 volumes of cold chloroform and incu-
715 bated for 3 min at room temperature. After centrifugation for 30 min at 4°C and maximum
716 speed, the supernatant was carefully transferred to a new tube and mixed with an equal vol-
717 ume of 70% ethanol. Afterwards the manufacturer's instruction from the RNeasy MinElute Kit
718 (Qiagen) were followed with DNase digestion (DNase I, Qiagen) for 30 min on column. Elution
719 of the RNA was carried out in 14 µl. Human globin mRNA was depleted from all samples
720 except from samples #1 and #2 using the GLOBINclear kit (ThermoFisher Scientific). The
721 quality of the RNA was assessed using the Agilent 6000 Pico kit with the Bioanalyzer 2100
722 (Agilent) (Figure S5), the RNA quantity using the Qubit RNA HA assay kit and a Qubit 3.0
723 fluorometer (ThermoFisher Scientific). Upon arrival at BGI Genomics Co. (Hong Kong), the
724 RNA quality of each sample was double-checked before sequencing. The median RIN value
725 over all *ex vivo* samples was 6.75 (IQR: 5.93–7.40) (Figure S5), although this measurement
726 has only limited significance for samples containing RNA of two species. Customized library
727 construction in accordance to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) including amplification
728 with KAPA polymerase and HiSeq 2500 100 bp paired-end sequencing was also performed
729 by BGI Genomics Co. (Hong Kong).

730

731 ***RNA-seq read mapping and data analysis***

732 Var *gene assembly*

733 *Var* genes were assembled using the pipeline described in Tonkin-Hill *et al.* (Tonkin-Hill *et al*,
734 2018). Briefly, non-*var* reads were first filtered out by removing reads that aligned to *H. sapi-*
735 *ens*, *P. vivax* or non-*var P. falciparum*. Assembly of the remaining reads was then performed
736 using a pipeline combining SOAPdenovo-Trans and Cap3 (Xie *et al*, 2014; Huang & Madan,
737 1999; Liao *et al*, 2013). Finally, contaminants were removed from the resulting contigs and
738 they were then translated into the correct reading frame. Reads were mapped to the contigs
739 using BWA-MEM (Li, 2013) and RPKM values were calculated for each *var* transcript to com-
740 pare individual transcript levels in each patient. Although transcripts might be differentially cov-
741 ered by RNA-seq due to their variable GC content, this seems not to be an issue between *var*
742 genes (Tonkin-Hill *et al*, 2018).

743

744 Var *transcript differential expression*

745 Expression for the assembled *var* genes was quantified using Salmon v0.14.1 (Patro *et al*,
746 2017) for 531 transcripts with five read counts in at least 3 patient isolates. Both the naïve and
747 pre-exposed groups as well as the severe and non-severe groups were compared. The com-
748 bined set of all *de novo* assembled transcripts was used as a reference in addition to the coding
749 regions of the 3D7 reference genome. The Salmon algorithm identifies equivalence sets be-
750 tween transcripts allowing a single read to support the expression of multiple transcripts. This
751 enables it to account for the redundancy present in our dataset. To confirm the suitability of
752 this approach we also ran the Corset algorithm as used in Tonkin-Hill *et al.*, (Tonkin-Hill *et al*,
753 2018; Davidson & Oshlack, 2014). Unlike Salmon which attempts to quantify the expression
754 of transcripts themselves, Corset copes with the redundancy present in *de novo* transcriptome
755 assemblies by clustering similar transcripts together using both the sequence identify of the
756 transcripts as well as multi-mapping read alignments. Of the transcripts identified using the
757 Salmon analysis 5/15 in the naïve versus pre-exposed and 4/13 in the severe versus non-
758 severe were identified in the significant clusters produced using Corset. As the two algorithms
759 take very different approaches and as Salmon is quantifying transcripts rather than the 'gene'
760 like clusters of Corset this represents a fairly reasonable level of concordance between the
761 two methods. In both the Salmon and Corset pipelines differential expression analysis of the
762 resulting *var* expression values was performed using DESeq2 v1.26 (Love *et al*, 2014). The
763 Benjamini-Hochberg method was used to control for multiple testing (Benjamini & Hochberg,
764 1995).

765 To check differential expression of the conserved *var* gene variants *var1*-3D7, *var1*-IT and
766 *var2csa* raw reads were mapped with BWA-MEM (AS score >110) to the reference genes from
767 the 3D7 and the IT strains. The mapped raw read counts (bam files) were normalized with the

number of 3D7 mappable reads in each isolate using bamCoverage by introducing a scaling factor to generate bigwig files displayed in Artemis (Carver *et al*, 2012).

Var *domain and segment differential expression*

Differential expression analysis at the domain and segment level was performed using a similar approach to that described previously (Tonkin-Hill *et al*, 2018). Initially, the domain families and homology blocks defined in Rask *et al.* were annotated to the assembled transcripts using HMMER v3.1b2 (Rask *et al*, 2010; Eddy, 2011). Domains and segments previously identified to be significantly associated with severe disease in Tonkin-Hill *et al.,* 2018 were also annotated by single pairwise comparison in the assembled transcripts using USEARCH v11.0.667 (Tonkin-Hill *et al*, 2018; Edgar, 2010). Overall, 336 contigs (5.22% of all contigs >500 bp) possess partial domains in an unusual order, e.g. an NTS in an internal region or a tandem arrangement of two DBLα or CIDRα domains. This might be caused by *de novo* assembly errors, which is challenging from transcriptome data. Therefore, in both cases the domain or segment with the most significant alignment was taken as the best annotation for each region of the assembled *var* transcripts (E-value cutoff of $1e^{-}8$), once with the additional requirement that at least 60% of the domain was found. The expression at each of these annotations was then quantified using featureCounts v1.6.4 before the counts were aggregated to give a total for each domain and segment family in each sample. Finally, similar to the transcript level analysis, DESeq2 was used to test for differences in expression levels of both domain and segment families in the naïve versus pre-exposed groups as well as the severe versus non-severe groups. Again, more than five read counts in at least three patient isolates were required for inclusion into differential expression analysis.

Differential expression of core genes

Differential gene expression analysis of *P. falciparum* core genes was done in accordance to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) by applying the script as given in the Github repository (https://github.com/gtonkinhill/falciparum_transcriptome_manuscript/tree/master/all_gene_analysis). In brief, subread-align v1.4.6 (Liao *et al*, 2013) were used to align the reads to the *H. sapiens* and *P. falciparum* reference genomes. Read counts for each gene were obtained with FeatureCounts v1.20.2 (Liao *et al*, 2014). To account for parasite life cycle, each sample is considered as a composition of six parasite life cycle stages excluding the ookinete stage (López-Barragán *et al*, 2011). Unwanted variations were determined with the 'RUV' (Remove Unwanted Variation) algorithm implemented in the R package ruv v0.9.6 (Gagnon-Bartsch & Speed, 2012) adjusting for systematic errors of unknown origin by using the genes with the 1009 lowest p-values as controls as described in (Vignali *et al*, 2011). The gene counts and

estimated ring-stage factor, and factors of unwanted variation were then used as input for the Limma /Voom (Law *et al*, 2014; Smyth, 2005) differential analysis pipeline.

Functional enrichment analysis of differentially expressed core genes

Genes that were identified as significantly differentially expressed (defined as -1<logFC>1, p<0.05) during prior differential gene expression analysis were used for functional enrichment analysis using the R package gprofiler2 (Kolberg *et al*, 2020). Enrichment analysis was performed on multiple input lists containing genes expressed significantly higher (logFC > 1, P < 0.05) and lower (logFC < - 1, P < 0.05) between different patient cohorts. All *var* genes were excluded from the enrichment analysis. For custom visualization of results, gene set data sources available for *P. falciparum* were downloaded from gprofiler (Raudvere *et al*, 2019). Pathway data available in the KEGG database (https://www.kegg.jp/kegg/) was accessed via the KEGG API using KEGGREST (Tenenbaum, 2020) to supplement gprofiler data sources and build a custom data source in Gene Matrix Transposed file format (*.gmt) for subsequent visualization. Functional enrichment results were then output to a Generic Enrichment Map (GEM) for visualization using the Cytoscape EnrichmentMap app (Merico *et al*, 2010) and RCy3 (Gustavsen *et al*, 2019). Bar plots of differential gene expression values for genes of selected KEGG pathways were generated using ggplot2 (Wickham, 2016) and enriched KEGG pathways were visualized using KEGGprofile (Zhao S, Guo Y, 2020).

DBLα-tag sequencing

For DBLα-tag PCR the forward primer varF_dg2 (5'-tcgtcggcagcgtcagatgtgtataagaga-cagGCAMGMAGTTTYGCNGATATWGG-3') and the reverse primer brlong2 (5'-gtctcgtgggctcggagatgtgtataagagacagTCTTCDSYCCATTCVTCRAACCA-3') were used resulting in an amplicon size of 350-500 bp (median 422 bp) plus the 67 bp overhang (small type). Template cDNA (1 µl) was mixed with 5x KAPA HiFi buffer, 0.3 µM of each dNTP, 2 µM of each primer and 0.5 U KAPA HiFi Hotstart Polymerase in a final reaction volume of 25 µl. Reaction mixtures were incubated at 95°C for 2 min and then subjected to 35 cycles of 98°C for 20 s, 54°C for 30 s and 68°C for 75 s with a final elongation step at 72°C for 2 min. For the first 5 cycles cooling from denaturation temperature is performed to 65°C at a maximal ramp of 3°C per second, then cooled to 54°C with a 0.5°C per second ramp. Heating from annealing temperature to elongation temperature was performed with 1°C per second, all other steps with a ramp of 3°C per second. Agarose gel images taken afterwards showed clean amplicons. The DBLα-tag primers contain an overhang, which was used to conduct a second indexing PCR reaction using sample-specific indexing primers as described in Nag *et al.* (Nag *et al*, 2017). The overhang sequence also serves as annealing site for Illumina sequencing primers and indexing primers include individual 8-base combinations and adapter sequences that will

allow the final PCR product to bind in MiSeq Illumina sequencing flow cells. Indexing PCR reactions were performed with a final primer concentration of 0.065 µM and 1 µl of first PCR amplicon in a final volume of 20 µl; and by following steps: Heat activation at 95 °C, 15 min, 20 cycles of 95 °C for 20 s, 60 °C for 1 min and 72 °C for 1 min, and one final elongation step at 72 °C for 10 min. Indexing PCR amplicons were pooled (4 µl of each) and purified using AM-Pure XP beads (Beckman Coulter, California, United States) according to manufacturer's protocol, using 200 µl pooled PCR product and 0.6 x PCR-pool volume of beads, to eliminate primer dimers. The purified PCR pool were analyzed on agarose gels and Agilent 2100 Bioanalyser to verify elimination of primer dimers, and correct amplicon sizes. Concentration of purified PCR pools was measured by Nanodrop2000 (Thermo Fisher Scientific, Waltham, MA, USA) and an aliquot adjusted to 4 nM concentration was pooled with other unrelated DNA material and added to an Illumina MiSeq instrument for paired end 300 bp reads using a MiSeq v3 flow cell.

DBLα-tag sequence analysis

The paired-end DBLα-tag sequences were identified and partitioned into correct sample origin based on unique index sequences. Each indexed raw sequence-pair were then processed through the *Galaxy* webtool (usegalaxy.eu). Read quality checks was first performed with *FastQC* to ensure a good NGS run (sufficient base quality, read length, duplication etc.). Next, the sequences were trimmed by the Trimmomatic application, with a four base sliding window approach and a *Phred* quality score above 20 to ensure high sequence quality output. The trimmed sequences were then paired and converted, following analysis using the Varia tool for quantification and prediction of the domain composition of the full-length *var* sequences from which the DBLα-tag originated (Mackenzie *et al*, 2020). In brief, Varia clusters DBLα-tags with 99% sequence identity using *Vsearch* program (v2.14.2), and each unique tag is used to search a database consisting of roughly 235,000 annotated *var* genes for near identical *var* sequences (95% identity over 200 nucleotides). The domain composition of all "hit" sequences is checked for conflicting annotations and the most likely domain composition is retuned. The tool validation indicated prediction of correct domain compositions for around 85% of randomly selected *var* tags, with higher hit rate and accuracy of the N-terminal domains. An average of 2,223.70 reads per patient sample was obtained and clusters consisting of less than 10 reads were excluded from the analysis. The raw Varia output file is given in Table S8. The proportion of transcripts encoding a given *Pf*EMP1 domain type or subtype was calculated for each patient. These expression levels were used to first test the hypothesis that N-terminal domain types associated with EPCR are found more frequently in first-time infections or upon severity of disease, while those associated with non-EPCR binding were associated with pre-exposed or mild cases. Secondly, quantile regression was used to calculate median differences (with

95%-confidence intervals) in expression levels for all main domain classes and subtypes between severity and exposure groups. All analyses were done with R (4.02) using the package quantreg (5.73) for quantile regression.

For the comparison of both approaches, DBLα-tag sequencing and RNA-seq, only RNA-seq contigs spanning the whole DBLα-tag region were considered. All conserved variants, the subfamilies *var1*, *var2csa* and *var3*, detected by RNA-seq were omitted form analysis since they were not properly amplified by the DBLα-tag primers. To scan for the occurrence of DBLα-tag sequences within the contigs assembled from the RNA-seq data we applied BLAST (basic local alignment search tool) v2.9.0 software (Altschul *et al*, 1990). Therefore, we created a BLAST database from the RNA-seq assemblies and screened for the occurrence of those DBLα-tag sequence with more than 97% percent sequence identity using the "megablast option".

Calculation of the proportion of RNA-seq data covered by DBLα-tag was done with the upper 75th percentile based on total RPKM values determined for each patient. Vice versa, only DBLα-tag clusters with more than 10 reads were considered and percent coverage of reads and clusters calculated for each individual patient.

For all samples the agreement between the two molecular methods DBLα-tag sequencing and RNA-seq was analyzed with a Bland-Altman plot, each individually and summarized. The ratio between %-transformed measurements are plotted on the y-axis and the mean of the respective DBLα-tag and RNA-seq results are plotted on the x-axis. The bias and the 95% limits of agreement were calculated using GraphPad Prism 8.4.2.

Acknowledgments

We thank all the patients providing an extra blood sample for our research purposes. Furthermore, we thank Jürgen May for critical reading of the manuscript and Tobias Spielmann for helpful discussions. We thank Marlene Danner Dalgaard, Kathrine Hald Langhoff and Sif Ravn Søeborg technical assistance with DBLα-tag sequencing.

References

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410

Andrade CM, Fleckenstein H, Thomson-Luque R, Doumbo S, Lima NF, Anderson C, Hibbert J, Hopp CS, Tran TM, Li S, *et al* (2020) Increased circulation time of Plasmodium falciparum underlies persistent asymptomatic infection in the dry season. *Nat Med*

Argy N, Kendjo E, Augé-Courtoi C, Cojean S, Clain J, Houzé P, Thellier M, Hubert V,

915 Deloron P & Houzé S (2017) Influence of host factors and parasite biomass on the
916 severity of imported Plasmodium falciparum malaria. *PLoS One* 12: e0175328
917 Avril M, Brazier AJ, Melcher M, Sampath S & Smith JD (2013) DC8 and DC13 var Genes
918 Associated with Severe Malaria Bind Avidly to Diverse Endothelial Cells. *PLoS Pathog* 9
919 Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL, Sullivan DJ, Bull PC, Stins
920 MF & Smith JD (2012) A restricted subset of var genes mediates adherence of
921 Plasmodium falciparum-infected erythrocytes to brain endothelial cells. *Proc Natl Acad*
922 *Sci U S A* 109: E1782–E1790
923 Bachmann A, Bruske E, Krumkamp R, Turner L, Wichers JS, Petter M, Held J, Duffy MF,
924 Sim BKL, Hoffman SL, *et al* (2019) Controlled human malaria infection with Plasmodium
925 falciparum demonstrates impact of naturally acquired immunity on virulence gene
926 expression. *PLOS Pathog* 15: e1007906
927 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and
928 Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 57: 289–300
929 Berger SS, Turner L, Wang CW, Petersen JEV, Kraft M, Lusingu JPA, Mmbando B,
930 Marquard AM, Bengtsson DBAC, Hviid L, *et al* (2013) Plasmodium falciparum
931 Expressing Domain Cassette 5 Type PfEMP1 (DC5-PfEMP1) Bind PECAM1. *PLoS One*
932 8: 69117
933 Bernabeu M, Danziger SA, Avril M, Vaz M, Babar PH, Brazier AJ, Herricks T, Maki JN,
934 Pereira L, Mascarenhas A, *et al* (2016) Severe adult malaria is associated with specific
935 PfEMP1 adhesion types and high parasite biomass. *Proc Natl Acad Sci U S A* 113:
936 E3270–E3279
937 Bernabeu M, Gunnarsson C, Vishnyakova M, Howard CC, Nagao RJ, Avril M, Taylor TE,
938 Seydel KB, Zheng Y & Smith JD (2019) Binding Heterogeneity of Plasmodium
939 falciparum to Engineered 3D Brain Microvessels Is Mediated by EPCR and ICAM-1.
940 *MBio* 10: e00420-19
941 Borrmann S (2020) Mapping of safe and early chemo-attenuated live Plasmodium falciparum
942 immunization identifies immune signature of vaccine efficacy. *bioRxiv*:
943 2020.09.14.296152
944 Carlson J, Helmby H, Wahlgren M, Carlson J, Helmby H, Wahlgren M, Hill A, Brewster D &
945 Greenwood BM (1990) Human cerebral malaria: association with erythrocyte rosetting
946 and lack of anti-rosetting antibodies. *Lancet* 336: 1457–1460
947 Carver T, Harris SR, Berriman M, Parkhill J & McQuillan JA (2012) Artemis: An integrated
948 platform for visualization and analysis of high-throughput sequence-based experimental
949 data. *Bioinformatics* 28: 464–469
950 Chaiyaroj SC, Coppel RL, Magowan C & Brown G V. (1994) A Plasmodium falciparum
951 isolate with a chromosome 9 deletion expresses a trypsin-resistant cytoadherence

952 molecule. *Mol Biochem Parasitol* 67: 21–30
953 Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, Lavstsen T, Hviid L,
954 Duffy PE & Theander TG (2010) Hierarchical, domain type-specific acquisition of
955 antibodies to Plasmodium falciparum erythrocyte membrane protein 1 in Tanzanian
956 children. *Infect Immun* 78: 4653–4659
957 Cham GKK, Turner L, Lusingu J, Vestergaard L, Mmbando BP, Kurtis JD, Jensen ATR,
958 Salanti A, Lavstsen T & Theander TG (2009) Sequential, Ordered Acquisition of
959 Antibodies to Plasmodium falciparum Erythrocyte Membrane Protein 1 Domains . *J*
960 *Immunol* 183: 3356–3363
961 Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, Bull PC, Mok S,
962 Gupta AP, Wang CW, *et al* (2012) A subset of group A-like var genes encodes the
963 malaria parasite ligands for binding to human brain endothelial cells. *Proc Natl Acad Sci*
964 *U S A* 109: E1772
965 Dasanna AK, Lansche C, Lanzer M & Schwarz US (2017) Rolling Adhesion of Schizont
966 Stage Malaria-Infected Red Blood Cells in Shear Flow. *Biophys J* 112: 1908–1919
967 Davidson NM & Oshlack A (2014) Corset: Enabling differential gene expression analysis for
968 de novo assembled transcriptomes. *Genome Biol* 15: 410
969 Day KP, Karamalis F, Thompson J, Barnes DA, Peterson C, Brown H, Brown G V. & Kemp
970 DJ (1993) Genes necessary for expression of a virulence determinant and for
971 transmission of Plasmodium falciparum are located on a 0.3-megabase region of
972 chromosome 9. *Proc Natl Acad Sci U S A* 90: 8292–8296
973 Dent AE, Malhotra I, Wang X, Babineau D, Yeo KT, Anderson T, Kimmel RJ, Angov E, Lanar
974 DE, Narum D, *et al* (2016) Contrasting Patterns of Serologic and Functional Antibody
975 Dynamics to Plasmodium falciparum Antigens in a Kenyan Birth Cohort. *Clin Vaccine*
976 *Immunol* 23: 104–116
977 Doolan DL, Mu Y, Unal B, Sundaresh S, Hirst S, Valdez C, Randall A, Molina D, Liang X,
978 Freilich DA, *et al* (2008) Profiling humoral immune responses to P. falciparum infection
979 with protein microarrays. *Proteomics* 8: 4680–4694
980 Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, Malima R,
981 Lusingu J, Manjurano A, Nkya WMM, *et al* (2005) Estimating medium- and long-term
982 trends in malaria transmission by using serological markers of malaria exposure. *Proc*
983 *Natl Acad Sci U S A* 102: 5108–5113
984 Duffy F, Bernabeu M, Babar PH, Kessler A, Wang CW, Vaz M, Chery L, Mandala WL,
985 Rogerson SJ, Taylor TE, *et al* (2019) Meta-analysis of Plasmodium falciparum var
986 Signatures Contributing to Severe Malaria in African Children and Indian Adults. *MBio*
987 10: e00217-19
988 Duffy MF, Brown G V., Basuki W, Krejany EO, Noviyanti R, Cowman AF & Reeder JC (2002)

Transcription of multiple var genes by individual, trophozoite-stage Plasmodium falciparum cells expressing a chondroitin sulphate A binding phenotype. *Mol Microbiol* 43: 1285–1293

Duffy MF, Caragounis A, Noviyanti R, Kyriacou HM, Choong EK, Boysen K, Healer J, Rowe JA, Molyneux ME, Brown G V., *et al* (2006) Transcribed var genes associated with placental malaria in Malawian women. *Infect Immun* 74: 4875–4883

Duffy MF, Selvarajah SA, Josling GA & Petter M (2014) Epigenetic regulation of the plasmodium falciparum genome. *Brief Funct Genomics* 13: 203–216

Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7: 1002195

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461

Falk N, Kaestli M, Qi W, Ott M, Baea K, Cortés A & Beck HP (2009) Analysis of plasmodium falciparum var genes expressed in children from Papua New Guinea. *J Infect Dis* 200: 347–356

Felgner PL, Roestenberg M, Liang L, Hung C, Jain A, Pablo J, Nakajima-Sasaki R, Molina D, Teelen K, Hermsen CC, *et al* (2013) Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Sci Rep* 3

Gagnon-Bartsch JA & Speed TP (2012) Using control genes to correct for unwanted variation in microarray data. *Biostatistics* 13: 539–552

Ghumra A, Semblat J-P, Ataide R, Kifude C, Adams Y, Claessens A, Anong DN, Bull PC, Fennell C, Arman M, *et al* (2012) Induction of Strain-Transcending Antibodies Against Group A PfEMP1 Surface Antigens from Virulent Malaria Parasites. *PLoS Pathog* 8: e1002665

Gustavsen JA, Pai S, Isserlin R, Demchak B & Pico AR (2019) RCy3: Network biology using Cytoscape from within R. *F1000Research* 8: 1774

Heiber A, Kruse F, Pick C, Grüring C, Flemming S, Oberli A, Schoeler H, Retzlaff S, Mesén-Ramírez P, Hiss JA, *et al* (2013) Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in Plasmodium falciparum protein export. *PLoS Pathog* 9: e1003546

Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, Mayanja-Kizza H, Ssewanyana I, Kamya MR, Beeson JG, *et al* (2015) Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proc Natl Acad Sci U S A* 112: E4438–E4447

Herricks T, Avril M, Janes J, Smith JD & Rathod PK (2013) Clonal variants of Plasmodium falciparum exhibit a narrow range of rolling velocities to host receptor CD36 under dynamic flow conditions. *Eukaryot Cell* 12: 1490–1498

1026 Van Den Hoogen LL, Walk J, Oulton T, Reuling IJ, Reiling L, Beeson JG, Coppel RL, Singh
1027 SK, Draper SJ, Bousema T, *et al* (2019) Antibody responses to antigenic targets of
1028 recent exposure are associated with low-density parasitemia in controlled human
1029 plasmodium falciparuminfections. *Front Microbiol* 10
1030 Hsieh FL, Turner L, Bolla JR, Robinson C V., Lavstsen T & Higgins MK (2016) The structural
1031 basis for CD36 binding by the malaria parasite. *Nat Commun* 7
1032 Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:
1033 868–877
1034 Janes JH, Wang CP, Levin-Edens E, Vigan-Womas I, Guillotte M, Melcher M, Mercereau-
1035 Puijalon O & Smith JD (2011) Investigating the host binding signature on the
1036 Plasmodium falciparum PfEMP1 protein family. *PLoS Pathog* 7
1037 Jensen AR, Adams Y & Hviid L (2020) Cerebral Plasmodium falciparum malaria: The role of
1038 PfEMP1 in its pathogenesis and immunity, and PfEMP1-based vaccines to prevent it.
1039 *Immunol Rev* 293: 230–252 doi:10.1111/imr.12807
1040 Jensen ATR, Magistrado P, Sharp S, Joergensen L, Lavstsen T, Chiucchiuini A, Salanti A,
1041 Vestergaard LS, Lusingu JP, Hermsen R, *et al* (2004) Plasmodium falciparum
1042 Associated with Severe Childhood Malaria Preferentially Expresses PfEMP1 Encoded
1043 by Group A var Genes. *J Exp Med* 199: 1179–1190
1044 Jespersen JS, Wang CW, Mkumbaye SI, Minja DT, Petersen B, Turner L, Petersen JE,
1045 Lusingu JP, Theander TG & Lavstsen T (2016) *Plasmodium falciparum var* genes
1046 expressed in children with severe malaria encode CIDRα1 domains. *EMBO Mol Med* 8:
1047 839–850
1048 Kapelski S, Klockenbring T, Fischer R, Barth S & Fendel R (2014) Assessment of the
1049 neutrophilic antibody-dependent respiratory burst (ADRB) response to Plasmodium
1050 falciparum . *J Leukoc Biol* 96: 1131–1142
1051 Kessler A, Dankwa S, Bernabeu M, Harawa V, Danziger SA, Duffy F, Kampondeni SD,
1052 Potchen MJ, Dambrauskas N, Vigdorovich V, *et al* (2017) Linking EPCR-Binding
1053 PfEMP1 to Brain Swelling in Pediatric Cerebral Malaria. *Cell Host Microbe* 22: 601-
1054 614.e5
1055 Kirchgatter K & Del Portillo HA (2002) Association of severe noncerebral Plasmodium
1056 falciparum malaria in Brazil with expressed PfEMP1 DBL1α sequences lacking cysteine
1057 residues. *Mol Med* 8: 16–23
1058 Kivisi CA, Muthui M, Hunt M, Fegan G, Otto TD, Githinji G, Warimwe GM, Rance R, Marsh
1059 K, Bull PC, *et al* (2019) Exploring Plasmodium falciparum Var Gene Expression to
1060 Assess Host Selection Pressure on Parasites During Infancy. *Front Immunol* 10: 2328
1061 Kolberg L, Raudvere U, Kuzmin I, Vilo J & Peterson H (2020) gprofiler2 -- an R package for
1062 gene list functional enrichment analysis and namespace conversion toolset g:Profiler.

F1000Research 9: 709

Kraemer SM & Smith JD (2003) Evidence for the importance of genetic structuring to the structural and functional specialization of the Plasmodium falciparum var gene family. *Mol Microbiol* 50: 1527–1538

Kyes SA, Christodoulou Z, Raza A, Horrocks P, Pinches R, Rowe JA & Newbold CI (2003) A well-conserved Plasmodium falciparum var gene shows an unusual stage-specific transcript pattern. *Mol Microbiol* 48: 1339–1348

Kyes SA, Kraemer SM & Smith JD (2007) Antigenic variation in Plasmodium falciparum: Gene organization and regulation of the var multigene family. *Eukaryot Cell* 6: 1511–1520 doi:10.1128/EC.00173-07

Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, Koné AK, Doumbo OK, Plowe C V. & Rowe JA (2006) Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol* 150: 211–218

Lau CKY, Turner L, Jespersen JS, Lowe ED, Petersen B, Wang CW, Petersen JEV, Lusingu J, Theander TG, Lavstsen T, *et al* (2015) Structural conservation despite huge sequence diversity allows EPCR binding by the pfemp1 family implicated in severe childhood malaria. *Cell Host Microbe* 17: 118–129

Lavstsen T, Salanti A, Jensen ATR, Arnot DE & Theander TG (2003) Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J* 2: 1–14

Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, Wang CW, Berger SS, Baraka V, Marquard AM, *et al* (2012) Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc Natl Acad Sci U S A* 109: E1791–E1800

Law CW, Chen Y, Shi W & Smyth GK (2014) Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15

Lee HJ, Georgiadou A, Walther M, Nwakanma D, Stewart LB, Levin M, Otto TD, Conway DJ, Coin LJ & Cunnington AJ (2018) Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci Transl Med* 10: eaar3619

Lennartz F, Adams Y, Bengtsson A, Olsen RW, Turner L, Ndam NT, Ecklu-Mensah G, Moussiliou A, Ofori MF, Gamain B, *et al* (2017) Structure-Guided Identification of a Family of Dual Receptor-Binding PfEMP1 that Is Associated with Cerebral Malaria. *Cell Host Microbe* 21: 403–414

Lennartz F, Smith C, Craig AG & Higgins MK (2019) Structural insights into diverse modes of ICAM-1 binding by Plasmodium falciparum-infected erythrocytes. *Proc Natl Acad Sci U*

S A 116: 20124–20134

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

Liao Y, Smyth GK & Shi W (2013) The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41

Liao Y, Smyth GK & Shi W (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930

Llewellyn D, Miura K, Fay MP, Williams AR, Murungi LM, Shi J, Hodgson SH, Douglas AD, Osier FH, Fairhurst RM, *et al* (2015) Standardization of the antibody-dependent respiratory burst assay with human neutrophils and Plasmodium falciparum malaria. *Sci Rep* 5

López-Barragán MJ, Lemieux J, Quiñones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K & Su X (2011) Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC Genomics* 12: 587

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Lubiana P, Bouws P, Roth LK, Dörpinghaus M, Rehn T, Brehmer J, Wichers JS, Bachmann A, Höhn K, Roeder T, *et al* (2020) Adhesion between P. falciparum infected erythrocytes and human endothelial receptors follows alternative binding dynamics under flow and febrile conditions. *Sci Rep* 10: 4548

Mackenzie G, Jensen R, Lavstsen T & Otto T (2020) Varia: Prediction of variable genes. (https://github.com/GCJMackenzie/Varia)

Magallón-Tejada A, Machevo S, Cisteró P, Lavstsen T, Aide P, Rubio M, Jiménez A, Turner L, Valmaseda A, Gupta H, *et al* (2016) Cytoadhesion to gC1qR through Plasmodium falciparum Erythrocyte Membrane Protein 1 in Severe Malaria. *PLOS Pathog* 12: e1006011

Maier AG, Rug M, O'Neill MT, Brown M, Chakravorty S, Szestak T, Chesson J, Wu Y, Hughes K, Coppel RL, *et al* (2008) Exported Proteins Required for Virulence and Rigidity of Plasmodium falciparum-Infected Human Erythrocytes. *Cell* 134: 48–61

McGee M & Chen Z (2006) Parameter estimation for the exponential-normal convolution model for background correction of affymetrix GeneChip data. *Stat Appl Genet Mol Biol* 5

McHugh E, Carmo OMS, Blanch A, Looker O, Liu B, Tiash S, Andrew D, Batinovic S, Low AJY, Cho HJ, *et al* (2020) Role of plasmodium falciparum protein GEXP07 in Maurer's Cleft morphology, Knob architecture, and P. Falciparum EMP1 trafficking. *MBio* 11

McQuaid F & Rowe JA (2020) Rosetting revisited: A critical look at the evidence for host erythrocyte receptors in Plasmodium falciparum rosetting. *Parasitology* 147: 1–11

1137 doi:10.1017/S0031182019001288
1138 Merico D, Isserlin R, Stueker O, Emili A & Bader GD (2010) Enrichment Map: A Network-
1139 Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* 5:
1140 e13984
1141 Mkumbaye SI, Wang CW, Lyimo E, Jespersen JS, Manjurano A, Mosha J, Kavishe RA,
1142 Mwakalinga SB, Minja DTR, Lusingu JP, *et al* (2017) The severity of *Plasmodium*
1143 *falciparum* infection is associated with transcript levels of *var* genes encoding EPCR-
1144 binding PfEMP1. *Infect Immun*: IAI.00841-16
1145 Nacer A, Roux E, Pomel S, Scheidig-Benatar C, Sakamoto H, Lafont F, Scherf A & Mattei D
1146 (2011) clag9 is not essential for PfEMP1 surface expression in non-cytoadherent
1147 Plasmodium falciparum parasites with a chromosome 9 deletion. *PLoS One* 6
1148 Nag S, Dalgaard MD, Kofoed PE, Ursing J, Crespo M, Andersen LOB, Aarestrup FM, Lund
1149 O & Alifrangis M (2017) High throughput resistance profiling of Plasmodium falciparum
1150 infections based on custom dual indexing and Illumina next generation sequencing-
1151 technology. *Sci Rep* 7: 1–13
1152 Nunes MC, Goldring JPD, Doerig C & Scherf A (2007) A novel protein kinase family in
1153 Plasmodium falciparum is differentially transcribed and secreted to various cellular
1154 compartments of the host cell. *Mol Microbiol* 63: 391–403
1155 Obeng-Adjei N, Larremore DB, Turner L, Ongoiba A, Li S, Doumbo S, Yazew TB, Kayentao
1156 K, Miller LH, Traore B, *et al* (2020) Longitudinal analysis of naturally acquired PfEMP1
1157 CIDR domain variant antibodies identifies associations with malaria protection. *JCI*
1158 *Insight*
1159 Oberli A, Slater LM, Cutts E, Brand F, Mundwiler-Pachlatko E, Rusch S, Masik MFG, Erat
1160 MC, Beck HP & Vakonakis I (2014) A Plasmodium falciparum PHIST protein binds the
1161 virulence factor PfEMP1 and comigrates to knobs on the host cell surface. *FASEB J* 28:
1162 4420–4433
1163 Oberli A, Zurbrügg L, Rusch S, Brand F, Butler ME, Day JL, Cutts EE, Lavstsen T,
1164 Vakonakis I & Beck HP (2016) Plasmodium falciparum Plasmodium helical interspersed
1165 subtelomeric proteins contribute to cytoadherence and anchor P. falciparum erythrocyte
1166 membrane protein 1 to the host cell cytoskeleton. *Cell Microbiol* 18: 1415–1428
1167 Obiero JM, Campo JJ, Scholzen A, Randall A, Bijker EM, Roestenberg M, Hermsen CC,
1168 Teng A, Jain A, Davies DH, *et al* (2019) Antibody Biomarkers Associated with Sterile
1169 Protection Induced by Controlled Human Malaria Infection under Chloroquine
1170 Prophylaxis. *mSphere* 4
1171 Otto TD (2019) ThomasDOtto/varDB: First release of varDB.
1172 Otto TD, Assefa SA, Böhme U, Sanders MJ, Kwiatkowski D, Berriman M, Newbold C &
1173 Newbold C (2019) Evolutionary analysis of the most polymorphic gene family in

falciparum malaria. *Wellcome Open Res* 4: 193

Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, Bull PC, Pearson RD, Abdi A, Dimonte S, *et al* (2018a) Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome open Res* 3: 52

Otto TD, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, Oyola SO, Okouga AP, Boundenga L, Willaume E, *et al* (2018b) Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol* 3: 687–697

Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417–419

Perkins M (1988) Stage-dependent processing and localization of a Plasmodium falciparum protein of 130,000 molecular weight. *Exp Parasitol* 65: 61–68

Rask TS, Hansen DA, Theander TG, Gorm Pedersen A & Lavstsen T (2010) Plasmodium falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and Conquer. *PLoS Comput Biol* 6: e1000933

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H & Vilo J (2019) G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47: W191–W198

Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, Sarthou JL & Mercereau-Puijalon O (1996) Extensive genetic diversity of Plasmodium falciparum isolates collected from patients with severe malaria in Dakar, Senegal. *Trans R Soc Trop Med Hyg* 90: 704–711

Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen ATR, Müller D, Theander T & Beck HP (2006) Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum infection in Tanzanian children. *Infect Immun* 74: 3904–3911

Rowe JA, Claessens A, Corrigan RA & Arman M (2009) Adhesion of Plasmodium falciparum-infected erythrocytes to human cells: Molecular mechanisms and therapeutic implications. *Expert Rev Mol Med* 11: e16 doi:10.1016/0925-4005(95)85135-6

Rowe JA, Obiero J, Marsh K & Raza A (2002) Short report: Positive correlation between rosetting and parasitemia in Plasmodium falciparum clinical isolates. *Am J Trop Med Hyg* 66: 458–460

Saul A (1999) The role of variant surface antigens on malaria-infected red blood cells. *Parasitol Today* 15: 455–7

Shabani E, Hanisch B, Opoka RO, Lavstsen T & John CC (2017) Plasmodium falciparum EPCR-binding PfEMP1 expression increases with malaria disease severity and is

elevated in retinopathy negative cerebral malaria. *BMC Med* 15: 183

Silver JD, Ritchie ME & Smyth GK (2009) Microarray background correction: Maximum likelihood estimation for the normal-exponential convolution. *Biostatistics* 10: 352–363

Smyth GK (2005) limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* pp 397–420. Springer-Verlag

Spielmann T, Hawthorne PL, Bixon MWA, Hannemann M, Klotz K, Kemp DJ, Klonis N, Tilley L, Trenholme KR & Gardiner DL (2006) A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in Plasmodium falciparum codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. *Mol Biol Cell* 17: 3613–3624

Spycher C, Klonis N, Spielmann T, Kump E, Steiger S, Tilley L & Beck H-P (2003) MAHRP-1, a novel Plasmodium falciparum histidine-rich protein, binds ferriprotoporphyrin IX and localizes to the Maurer's clefts. *J Biol Chem* 278: 35373–83

Spycher C, Rug M, Klonis N, Ferguson DJP, Cowman AF, Beck H-P & Tilley L (2006) Genesis of and Trafficking to the Maurer's Clefts of Plasmodium falciparum-Infected Erythrocytes. *Mol Cell Biol* 26: 4074–4085

Spycher C, Rug M, Pachlatko E, Hanssen E, Ferguson D, Cowman AF, Tilley L & Beck H (2008) The Maurer ' s cleft protein MAHRP1 is essential for trafficking of PfEMP1 to the surface of Plasmodium falciparum-infected erythrocytes. *Mol Microbiol* 68: 1300–1314

Storm J, Jespersen JS, Seydel KB, Szestak T, Mbewe M, Chisala N V, Phula P, Wang CW, Taylor TE, Moxon CA, *et al* (2019) Cerebral malaria is associated with differential cytoadherence to brain endothelial cells. *EMBO Mol Med* 11

Taghavian O, Jain A, Joyner CJ, Ketchum S, Nakajima R, Jasinskas A, Liang L, Fong R, King C, Greenhouse B, *et al* (2018) Antibody Profiling by Proteome Microarray with Multiplex Isotype Detection Reveals Overlap between Human and Aotus nancymaae Controlled Malaria Infections. *Proteomics* 18

Tenenbaum D (2020) KEGGREST: Client-side REST access to KEGG. R package version 1.30.0.

Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, Marfurt J, Cobbold SA, Rambhatla JS, McConville MJ, *et al* (2018) The Plasmodium falciparum transcriptome in severe malaria reveals altered expression of genes involved in important processes including surface antigen–encoding var genes. *PLoS Biol* 16: e2004328

Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, Da Silva LM, Scherf A, Ivens A, *et al* (2009) Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. *PLoS Biol* 7: 0771–0788

Turewicz M, Ahrens M, May C, Marcus K & Eisenacher M (2016) PAA: An R/bioconductor

1248 package for biomarker discovery with protein microarrays. *Bioinformatics* 32: 1577–
1249 1579
1250 Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JE V., Avril M, Brazier AJ, Freeth J,
1251 Jespersen JS, Nielsen MA, *et al* (2013) Severe malaria is associated with parasite
1252 binding to endothelial protein C receptor. *Nature* 498: 502–505
1253 Turner L, Lavstsen T, Mmbando BP, Wang CW, Magistrado PA, Vestergaard LS, Ishengoma
1254 DS, Minja DTR, Lusingu JP & Theander TG (2015) IgG antibodies to endothelial protein
1255 C receptor-binding cysteine-rich interdomain region domains of Plasmodium falciparum
1256 erythrocyte membrane protein 1 are acquired early in life in individuals exposed to
1257 malaria. *Infect Immun* 83: 3096–3103
1258 Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, Bouzek H, Moon W, Babak
1259 T, Fried M, *et al* (2011) NSR-seq transcriptional profiling enables identification of a gene
1260 signature of Plasmodium falciparum parasites infecting children. *J Clin Invest* 121:
1261 1119–1129
1262 Wang WF & Zhang YL (2020) PfSWIB, a potential chromatin regulator for var gene
1263 regulation and parasite development in Plasmodium falciparum. *Parasites and Vectors*
1264 13: 48
1265 Warimwe GM, Keane TM, Fegan G, Musyoki JN, Newton CRJC, Pain A, Berriman M, Marsh
1266 K & Bull PC (2009) Plasmodium falciparum var gene expression is modified by host
1267 immunity. *Proc Natl Acad Sci U S A* 106: 21801–21806
1268 WHO (2019) WHO World Malaria Report 2019
1269 Wickham H (2016) Ggplot2 : elegant graphics for data analysis Springer
1270 Winter G, Chen Q, Flick K, Kremsner P, Fernandez V & Wahlgren M (2003) The 3D7var5.2
1271 (varCOMMON) type var gene family is commonly expressed in non-placental
1272 Plasmodium falciparum malaria. *Mol Biochem Parasitol* 127: 179–191
1273 Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, *et al* (2014)
1274 SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads.
1275 *Bioinformatics* 30: 1660–1666
1276 Yam XY, Birago C, Fratini F, Di Girolamo F, Raggi C, Sargiacomo M, Bachi A, Berry L, Fall
1277 G, Currà C, *et al* (2013) Proteomic analysis of detergent-Resistant membrane
1278 microdomains in trophozoite blood stage of the human malaria parasite plasmodium
1279 falciparum. *Mol Cell Proteomics* 12: 3948–3961
1280 Zhang Q, Siegel TN, Martins RM, Wang F, Cao J, Gao Q, Cheng X, Jiang L, Hon CC,
1281 Scheidig-Benatar C, *et al* (2014) Exonuclease-mediated degradation of nascent RNA
1282 silences genes linked to severe malaria. *Nature* 513: 431–435
1283 Zhao S, Guo Y SY (2020) KEGGprofile: An annotation and visualization package for multi-
1284 types and multi-groups expression data in KEGG pathway. R package version 1.32.0.

Figure legends

Figure 1: Subgrouping of patients into first-time infected and pre-exposed individuals based on antibody levels against *P. falciparum*. In order to further characterize the patient cohort plasma samples (n=32) were subjected to Luminex analysis with the *P. falciparum* antigens AMA1, MSP1, CSP known to induce a strong antibody response in humans. With exception of patient #21 unsupervised clustering of the PCA-reduced data clearly discriminates between first-time infected (naïve) and pre-exposed patients with higher antibody levels against tested *P. falciparum* antigens and also assigns plasma samples from patients with unknown immune status into naïve and pre-exposed clusters (**A**). Classification of patient #21 into the naïve subgroup was confirmed using different serological assays assessing antibody levels against *P. falciparum* on different levels: a merozoite-directed antibody-dependent respiratory burst (mADRB) assay (Kapelski *et al*, 2014) (**B**), a PEMS-specific ELISA (**C**) and a 262-feature protein microarray covering 228 well-known *P. falciparum* antigens detecting reactivity with individual antigens and the antibody breadth of IgG (upper panel) and IgM (lower panel) (**D**). The boxes represent medians with IQR; the whiskers depict minimum and maximum values (range) with outliers located outside the whiskers. Serological assays revealed significant differences between patient groups (Mann Whitney U test). Reactivity of patient plasma IgG and IgM with individual antigens in the protein microarray is presented as volcano plot highlighting the significant hits in red. Box plots represent antibody breadths by summarizing the number of recognized antigens out of 262 features tested. Data from all assays were used for an unsupervised random forest approach (**E**). The variable importance plot of the random forest model shows the decrease in prediction accuracy if values of a variable are permuted randomly. The decrease in accuracy was determined for each serological assay indicating that the mADRB, ELISA and Luminex assays are most relevant in the prediction of patient clusters (**F**). Patients with known immune status based on medical reports were marked in all plots with filled circles in blue (naïve) and grey (pre-exposed), samples from patients with unknown immune status are shown as open circles. ELISA: Enzyme-linked immunosorbent assay, IQR: interquartile range, PCA: Principal component analysis

Figure 2: Estimated stage proportions for each sample. Patient samples consist of a combination of different parasite stages. To estimate the proportion of different life cycle stages in each sample a constrained linear model was fit using data from López-Barragán *et al.* (López-Barragán *et al*, 2011). The proportions of rings (8 hpi), early trophozoites (19 hpi), late trophozoites (30 hpi), schizonts (42 hpi) and gametocytes stages shown in the columns of the bar plots must add to 1 for each sample. Shown are the comparisons between first-time infected

(naïve; blue) and pre-exposed samples (grey) (**A**) and severe (red) and non-severe cases (grey) (**B**). A bias towards the early trophozoite appears in the non-severe malaria sample group, which was confirmed by calculating the age in hours post infection (hpi) for each parasite sample. The boxes represent medians with IQR; the whiskers depict minimum and maximum values (range) with outliers located outside the whiskers (**C**, **D**). IQR: interquartile range

Figure 3: Analysis of RNA-seq data at the level of *var* gene transcripts using the separate assembly approach. RNA-seq reads of each patient sample were matched to *de novo* assembled *var* contigs with varying length, domain and homology block composition. Shown are significant differently expressed *var* gene contigs with an adjusted p-value of <0.05 in first-time infected (blue) and pre-exposed patient samples (grey) (**A**, **B**) as well as severe (red) and non-severe cases (grey) (**C**, **D**). Data are displayed as heatmaps showing expression levels in log transformed normalized Salmon read counts for each individual sample (**A**, **C**) or as box plot with median log transformed normalized Salmon read counts and interquartile range (IQR) for each group of samples (**B**, **D**). Normalized Salmon read counts for all assembled transcripts are available in Table S4.

Figure 4: Analysis of RNA-seq data via *de novo* assembly at the level of *var* gene domains. RNA-seq reads of each patient sample were matched to *de novo* assembled *var* contigs with varying length, domain and homology block composition. Shown are significantly differently expressed *Pf*EMP1 domain subfamilies from Rask *et al.* (Rask *et al*, 2010) with an adjusted p-value of <0.05 in first-time infected (blue) and pre-exposed patient samples (grey) (**A**, **B**) as well as severe (red) and non-severe cases (grey) (**C**, **D**) using HMMER3 models. The N-terminal head structure (NTS-DBLα -CIDRα/β/γ/δ) confers a mutually exclusive binding phenotype either to EPCR-, CD36-, CSA- or an unknown receptor. Expression values of the N-terminal domains were summarized for each patient and differences in the distribution among patient groups were tested using the Mann-Whitney U test (**E**, **F**). Data are displayed as heatmaps showing expression levels in log transcripts per million (TPM) for each individual sample (**A**, **C**) or as box plot with median log TPM and interquartile range (IQR) for each group of samples (**B**, **D, E, F**).

Figure 5: Analysis of RNA-seq data via *de novo* assembly at the level of *var* gene homology blocks. RNA-seq reads of each patient sample were matched to *de novo* assembled *var* contigs with varying length, domain and homology block composition. Shown are significantly differently expressed homology blocks from Rask *et al.* (Rask *et al*, 2010) with an adjusted p-value of <0.05 in first-time infected (blue) and pre-exposed patient samples (grey) (**A**, **B**) as well as severe (red) and non-severe cases (grey) (**C**, **D**). Data are displayed as heatmaps showing expression levels in log transcripts per million (TPM) for each individual sample (**A**,

C) or as box plot with median log TPM and interquartile range (IQR) for each group of samples (**B**, **D**).

Figure 6: Summary of *Pf*EMP1 transcripts, domains, and homology blocks that were found more or less frequently in malaria-naïve and severely ill patients. A schematic presentation of typical *Pf*EMP1 domain compositions (**A**). The N-terminal head structure confers mutually exclusive receptor binding phenotypes: EPCR (yellow: CIDRα1.1/4-8), CD36 (salmon CIDRα2-6), CSA (VAR2CSA) and yet unknown phenotypes (orange: CIDRβ/γ/δ, red: CIDRα1.2/3 from VAR1, VAR3). Group A includes the conserved subfamilies VAR1 and VAR3, EPCR binding variants and those with unknown binding phenotypes sometimes associated with rosetting. Group B *Pf*EMP1 can have EPCR-binding capacities, but most variants share a four-domain structure with group C-type variants capable of CD36-binding. Dual binder can be found within group A and B with an DBLβ domain responsible for ICAM-1- (DBLβ1/3/5) or gC1qr-binding (DBLβ12). Inter-strain conserved tandem arrangements of domains, so called domain cassettes (DC), can be found within all groups as selectively indicated.
Transcripts, domains and homology blocks according to Rask *et al.* (Rask *et al*, 2010) found significant differently expressed (p-value <0.05) between patient groups of both comparisons: first-time infected (blue) versus pre-exposed (black) cases and severe (red) versus non-severe (black) cases (**B**).
ATS: acidic terminal sequence, CIDR: cysteine-rich interdomain region, CSA: chondroitin sulphate A, DBL: Duffy binding-like, DC: domain cassette, EPCR: endothelial protein C receptor, gC1qr: receptor for complement component C1q, ICAM-1: intercellular adhesion molecule 1, NTS: N-terminal segment, TM: transmembrane domain

Figure 7: Verification of RNA-seq results using DBLα-tag sequencing. Amplified DBLα-tag sequences were blasted against the ~2,400 genomes on varDB (Otto, 2019) to obtain subclassification into DBLα0/1/2 and prediction of adjacent head-structure NTS and CIDR domains and their related binding phenotype. Proportion of each NTS and DBLα subclass as well as CIDR domains grouped according to binding phenotype (CIDRα1.1/4-8: EPCR-binding, CIDRα2-6: CD36-binding, CIDRβ/γ/δ: unknown binding phenotype/rosetting) was calculated and shown separately on the left, number of total reads and individual sequence cluster with n ≥10 sequences are shown on the right. Differences in the distribution among first-time infected (blue) and pre-exposed individuals (grey) (**A**) as well as severe (red) and non-severe cases (grey) (**B**) were tested using the Mann-Whitney U test. The boxes represent medians with IQR; the whiskers depict minimum and maximum values (range) with outliers located outside the whiskers. Quantile regression was applied to look for differences between patient groups on the level of domain main classes (left) and subdomains (right). Shown are median

differences with 95%-confidence intervals of domains with values unequal 0. Domains with positive values tend to be higher expressed in naïve (**C**) and severe patients (**D**).

Figure 8: Correlation of *var* gene expression with antibody levels against head structure CIDR domains. Patient plasma samples (n=32) were subjected to Luminex analysis with 35 *Pf*EMP1 head structure CIDR domains. The panel includes EPCR-binding CIDRα1 domains (n = 19), CD36-binding CIDRα2–6 domains (n = 12) and CIDR domains with unknown binding phenotype (CIDRγ3: n = 1, CIDRδ1: n = 3) as well as the minimal binding region of VAR2CSA (VAR2). Box plots showing mean fluorescence intensities (MFI) extending from the 25th to the 75th percentiles with a line at the median indicate higher reactivity of the pre-exposed (**A**) and non-severe cases (**B**) with all *Pf*EMP1 domains tested. Significant differences were observed for recognition of CIDRα2–6, CIDRδ1 and CIDRγ3; VAR2CSA recognition differed only between severe and non-severe cases (Mann Whitney U test). Furthermore, the breadth of IgG recognition (%) of CIDR domains for the different patient groups was calculated and shown as a heat map (**C**).

Figure 9: Differential all gene expression analysis. Gene set enrichment analysis (GSEA) of GO terms and KEGG pathways indicate gene sets deregulated in first-time infected malaria patients. GO terms related to antigenic variation and host cell remodeling are significantly down-regulated, only the KEGG pathway 03410 'base excision repair' shows a significant up-regulation in malaria-naïve patients (**A**). Log fold changes (logFC) for the 15 *P. falciparum* genes assigned to the KEGG pathway 03410 'base excision repair' are plotted with the six significant hits marked with * for p<0.05 and ** for p<0.01 (**B**).

Table 1: Characteristics and classification of malaria patients.

Table 2: Patient groups data.

Table 3: *var* transcripts up- and down-regulated in first-time infected patients.

Table 4: *var* transcripts up- and down-regulated in severe cases.

Table 5: *var* domains defined by Rask *et al.* (Rask *et al*, 2010) up- and down-regulated in first-time infected patients.

Table 6: *var* domains defined by Rask *et al.* . (Rask *et al*, 2010) up- and down-regulated in severe cases.

Table 7: *var* blocks defined by Rask *et al.* . (Rask *et al*, 2010) up- and down-regulated in first-time infected patients.

Table 8: *var* blocks defined by Rask *et al.* . (Rask *et al*, 2010) down-regulated in severe cases.

Supplement

Supplement figure 1: Early immune response in mild and severe malaria within the naïve patient cluster. Antibody reactivity against individual antigens within the three subgroups 'naïve with mild symptoms', 'naïve with severe symptoms' and 'pre-exposed with mild symptoms'. Sera from all volunteers were assessed on protein microarrays and data normalized to control spots containing no antigen (no DNA control spots). Median reactivity of the mild infected malaria-naïve, severely infected malaria-naïve as well as the mild infected with pre-exposure to malaria are represented as bar-charts. IgG data is given for all 262 *P. falciparum* proteins spotted on the microarray representing 228 unique antigens (**A**). To estimate differences in immune response in mild and severe malaria within the malaria-naïve population, normalized IgG (**B**) and IgM (**C**) antibody responses were compared in the two subpopulations. Differentially recognized antigens ($p < 0.05$ and fold change >2) are depicted in red.

Supplement figure 2: Differential expression of the *var1*-allele forms and *var2csa* between patient groups. RNA-seq reads from each patient were normalized against the number of mappable reads to the 3D7 genome and aligned to the *var1*-3D7 and *var1*-IT allele forms as well as *var2csa*. The resulting bigwig files were displayed in Artemis (Carver *et al*, 2012). Individual samples are colored according to the patient group: first-time infected in blue (A), severe in red (B) and the respective pre-exposed or non-severe samples in grey.

Supplement figure 3: Comparison of the DBLα-tag sequencing with RNA-seq analysis. DBLα-tag sequencing and RNA-seq data compared in Bland-Altman plots for all patients summarized (**A**) and for each individual patient (**B**), where the mean log expression of each gene is indicated on the X-axis and the log ratio between normalized DBLα-tag (% of reads) and RNA-seq values (% of RPKM from all contigs containing both DBLα-tag primer binding sites) on the y-axis. The mean (equal to bias) of all ratios (line) and the confidence interval (CI) of 95% (dotted lines) are indicated. Data points with negative values for one of the approaches are displayed in dependence of their mean log expression on top (DBLα-tag sequence clusters

not detected by RNA-seq) or bottom (RNA-seq contigs not found within DBLα-tag sequence cluster) of the graph.

Supplement figure 4: The base excision repair (KEGG:03410) in *P. falciparum*. Orthologues present in *P. falciparum* are indicated by gene IDs, log fold changes (logFC) are indicated by color code (red: up-regulated, blue: down-regulated) (**A**). Summary of logFC in gene expression in first-time infected relative to pre-exposed patients and p-values for the logFC.

Supplement figure 5: RNA quality. The Bioanalyzer automated RNA electrophoresis system was used to characterize the total RNA quality prior library synthesis. The calculated RIN value is provided, although this measurement is questionable for samples from mixed species. From the four rRNA peaks visible in all samples, the inner peaks represent *P. falciparum* 18S and 28S rRNA, the outer peaks are of human origin.

Data S1: Sequences of assembled *var* contigs from all patient isolates.

Supplement table 1: Data from Luminex, mADRB, ELISA and protein microarray. Sero-prevalence of head-structure CIDR domains determined by applying a cut off from Danish controls (mean + 2 STD) to the Luminex data.

Supplement table 2: Raw read counts by sample for *H. sapiens*, *P. falciparum*, *var* exon 1 and percentage of reads that mapped either to *P. falciparum* or *var* exon 1 as well as the number of assembled *var* contigs >500 bp in length.

Supplement table 3: Features of the assembled *var* fragments annotated in accordance with Rask *et al.* . (Rask *et al*, 2010) and Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018). The reading frame used for translation is given after the contig ID, the position of each annotation is provided by starting and ending amino acid followed by the p-value from the blast search against the respective database. For annotations in accordance with Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) either the short ID or 'NA' (not applicable) is listed at the end. Short IDs are only available for significant differently expressed domains and blocks between severe and non-severe cases (Tonkin-Hill *et al*, 2018).

Supplement table 4: Summary of *var* gene fragments assembled for each patient isolate showing length, raw read counts, RPKM, blast hits, domain and block annotations in accordance with Rask *et al.* . (Rask *et al*, 2010). The RPKM for the contigs was calculated as number of mapped reads and normalized by the number of mapped reads against all

transcript in each isolate, respectively. Therefore, RPKM expression values are only valid to compare within a single sample since RNA-seq reads were mapped only to the contigs of the respective patient isolate using BWA-MEM (Li, 2013). Further, the amount of blast hits with 500 bp or 80% of overlap against the ~2400 samples from varDB (Otto, 2019) with an identity cutoff of 98%. Further hits of 1 kb (>98% identity) against the *var* genes from the 15 reference genomes (Otto *et al*, 2018a) are listed. The last two column show the annotations from Rask *et al.* (Rask *et al*, 2010) associated to each contig.

Supplement table 5: Log transformed normalized Salmon read counts for assembled *var* transcripts, TPM for collapsed domains and homology blocks from each patient isolate. Normalized counts and TPM values calculated for transcripts, domains and blocks with expression in at least three patient isolates with more than five read counts.

Supplement table 6: Differently expressed *var* transcripts, domains and homology blocks between first-time infected and pre-exposed patient samples.

Supplement table 7: Differently expressed *var* transcripts, domains and homology blocks between severe and non-severe patient samples.

Supplement table 8: Data from DBLα-tag sequencing.

Supplement table 9: Differentially expressed genes excluding *var* genes (all gene analysis) between first-time infected and pre-exposed patient samples.

Supplement table 10: Differentially expressed genes excluding *var* genes (all gene analysis) between severe and non-severe patient samples.

Figure 1

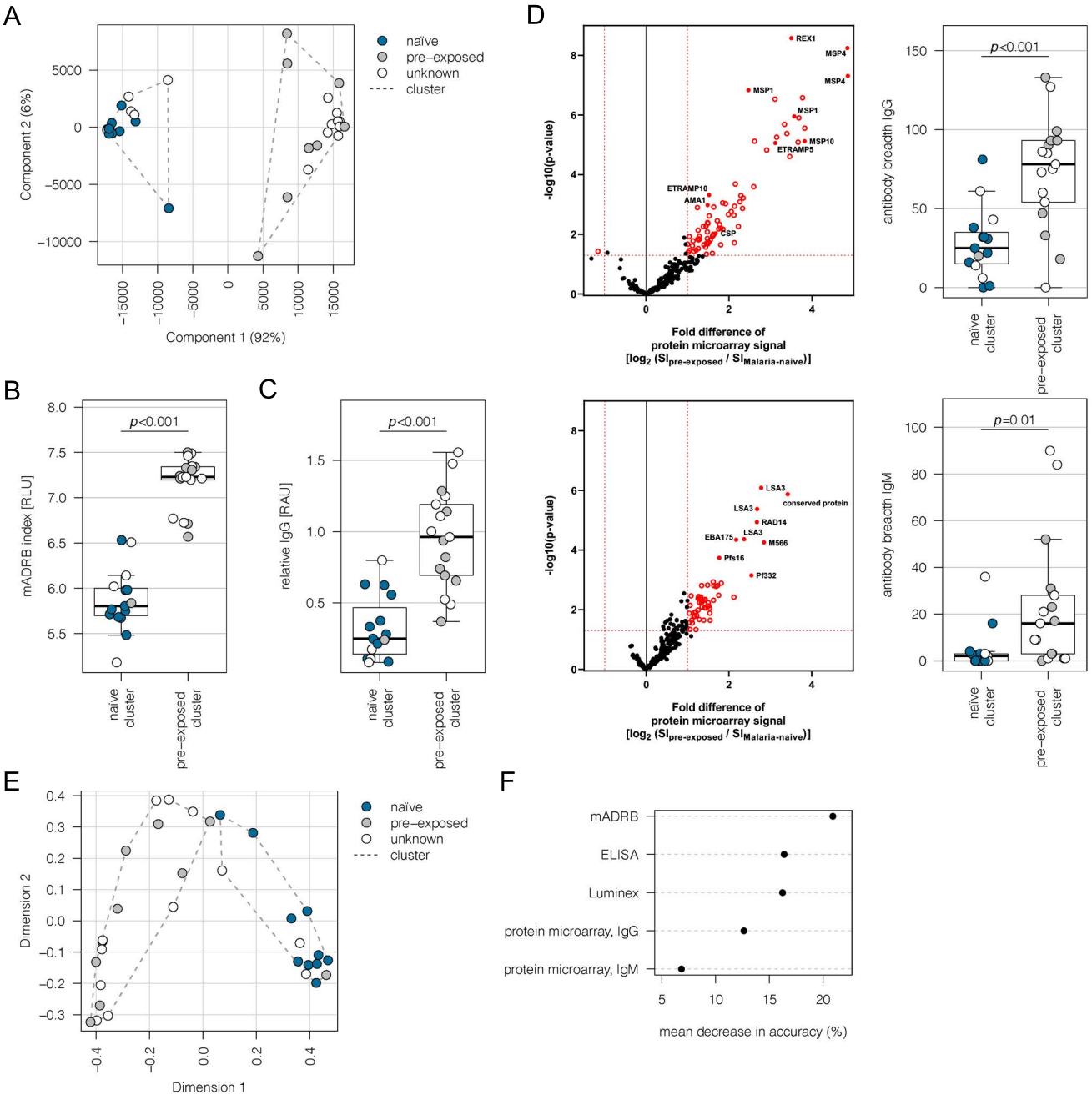


Figure 2

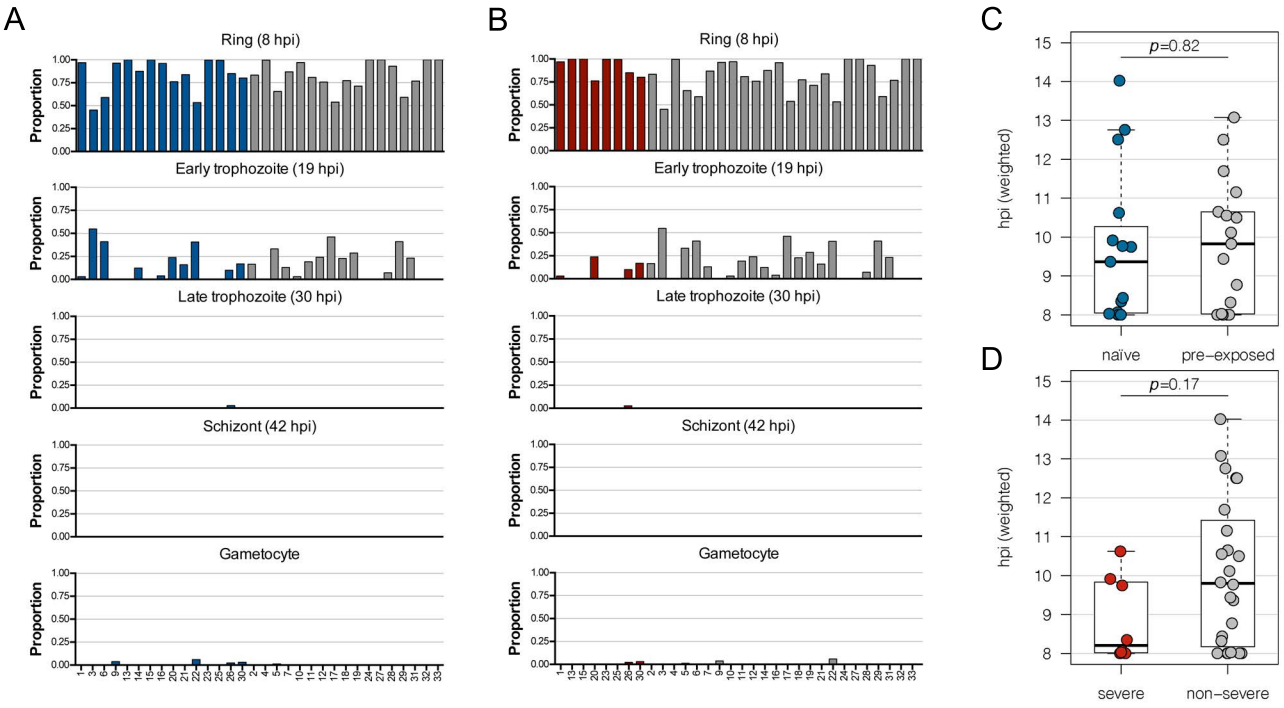


Figure 3

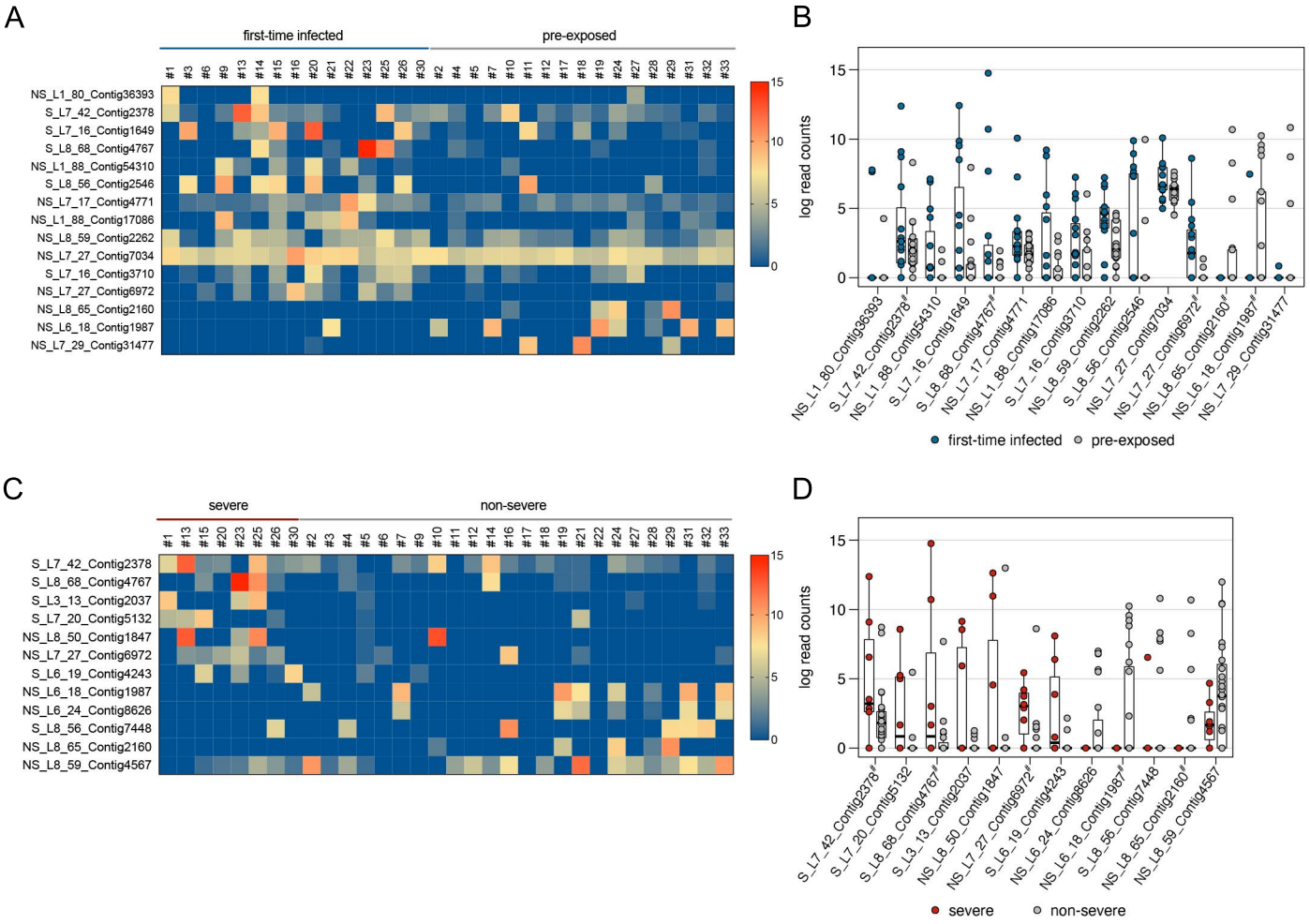


Figure 4

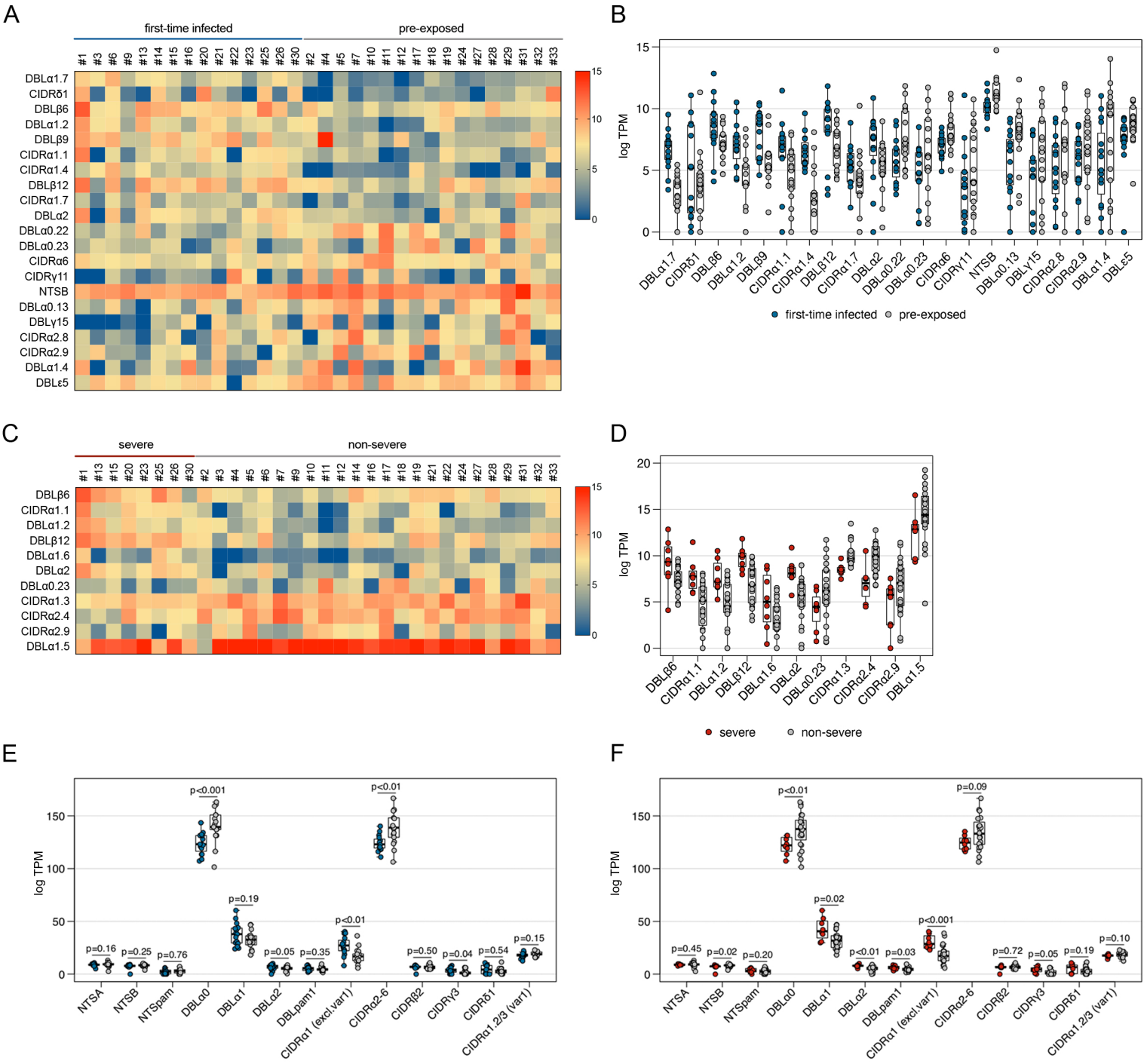


Figure 5

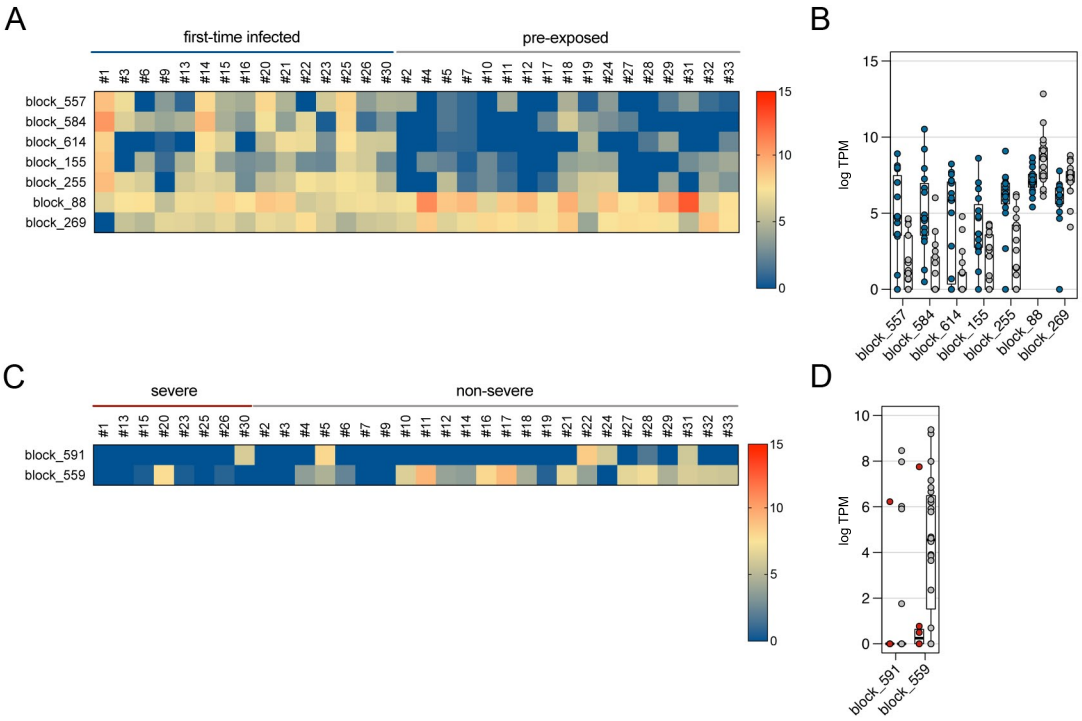
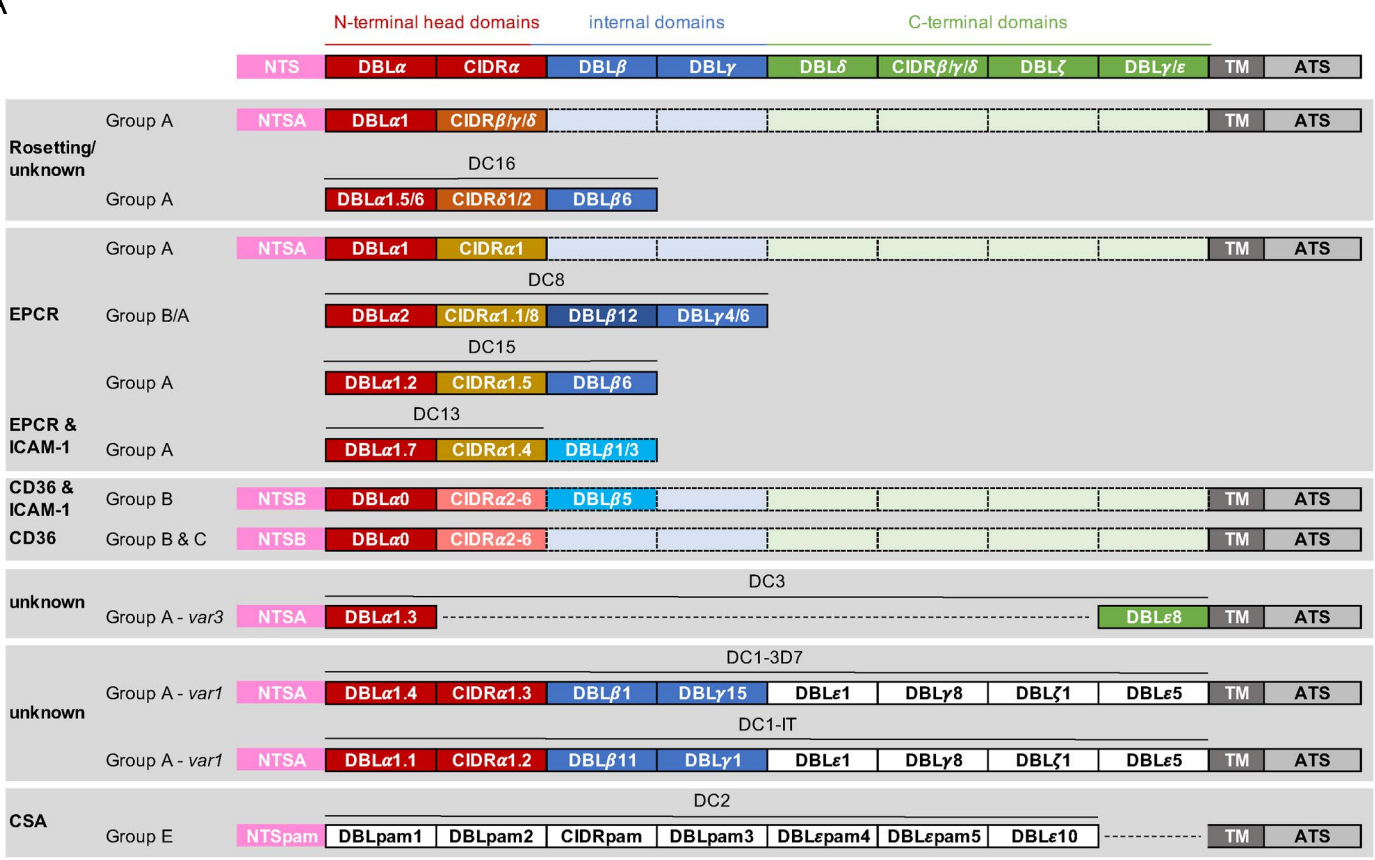


Figure 6

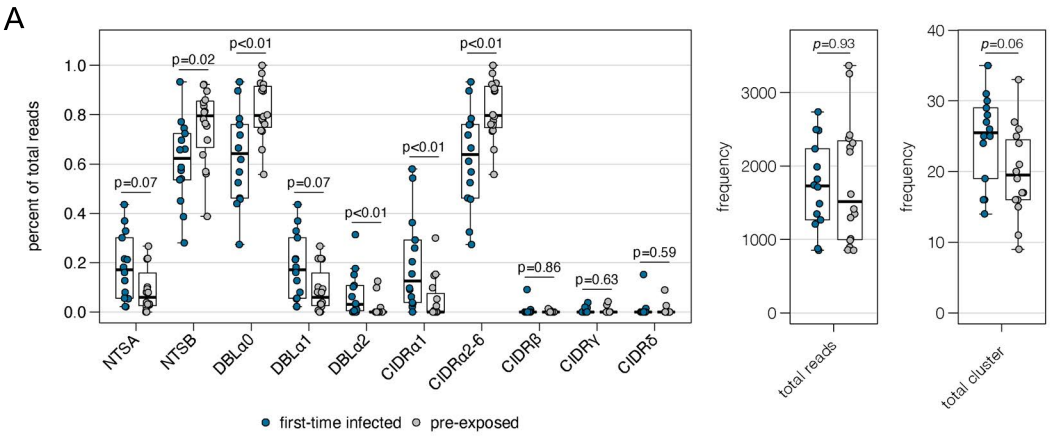
A

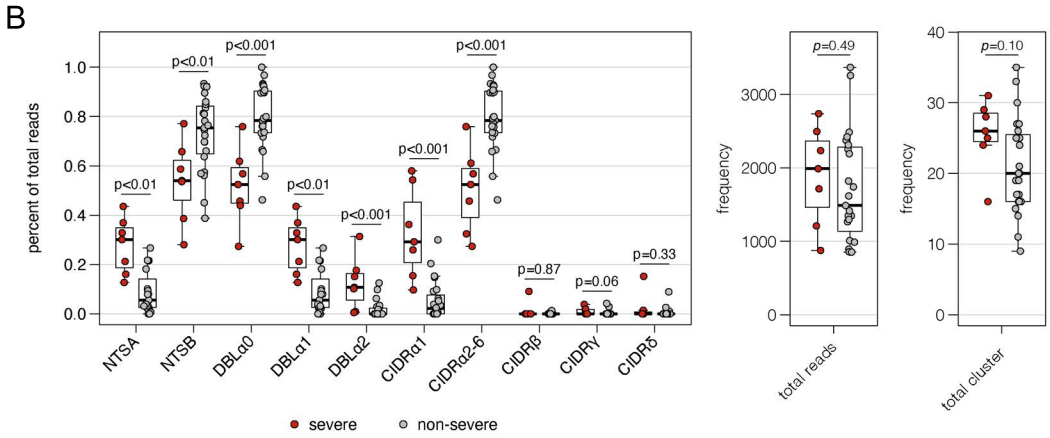


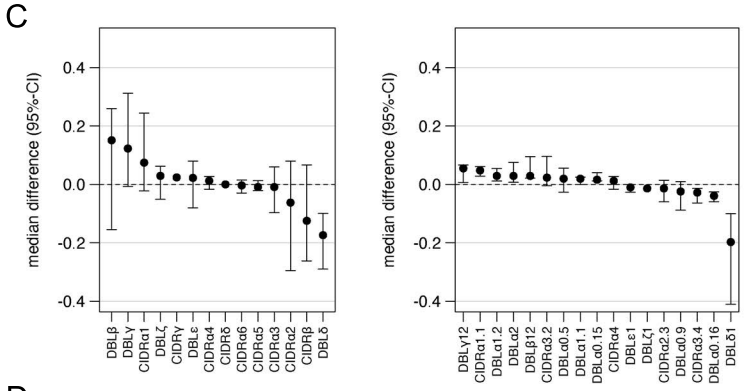
B

Binding phenotype	DC	first-time infected – pre-exposed: Transcripts	Domains	Homology blocks	severe – non-severe: Transcripts	Domains	Homology blocks
Rosetting/ unknown	DC16		CIDRδ1 DBLβ6	155 (NTSA)		DBLα1.6, DBLα1.5 DBLβ6	
EPCR	DC8, DC15		DBLα1.2, DBLα2 CIDRα1.1, CIDRα1.7 DBLβ6, DBLβ12	155 (NTSA)	S_L7_20_Contig5132	DBLα1.2, DBLα2 CIDRα1.1 DBLβ6, DBLβ12	
EPCR & ICAM-1	DC13		DBLα1.7 CIDRα1.4	584 (DBLβ3)			
CD36 & ICAM-1		S_L8_68_Contig4767			S_L8_68_Contig4767 S_L3_13_Contig2037		
CD36		NS_L1_88_Contig17086 NS_L7_29_Contig31477	NTSB DBLα0.13, DBLα0.22, DBLα0.23 CIDRα2.8, CIDRα2.9, CIDRα6	88 (DBLα0) 69 (ATSB)		DBLα0.23 CIDRα2.4, CIDRα2.9	591 (CIDRβ/γ-ATSB)
unknown	DC3	S_L7_16_Contig3710		155 (NTSA)			
unknown	DC1-3D7	NS_L8_65_Contig2160 NS_L6_18_Contig1987	DBLα1.4 DBLγ15 DBLε5		NS_L6_24_Contig8626 NS_L6_18_Contig1987 S_L8_56_Contig7448 NS_L8_65_Contig2160 NS_L8_59_Contig4567	CIDRα1.3	
unknown	DC1-IT	S_L7_16_Contig1649 S_L8_56_Contig2546 NS_L1_80_Contig36393		155 (NTSA)			
CSA	DC2	S_L7_42_Contig2378, NS_L8_59_Contig2262			S_L7_42_Contig2378 NS_L8_50_Contig1847		
others		NS_L1_88_Contig54310 NS_L7_17_Contig4771 NS_L7_27_Contig7034	DBLβ9 CIDRγ11	557 (DBLβ-DBLγ) 614 (DBLβ12/3) 255 (DBLβ12/3/1/5)	S_L6_19_Contig4243		559 (DBLζ6)

Figure 7







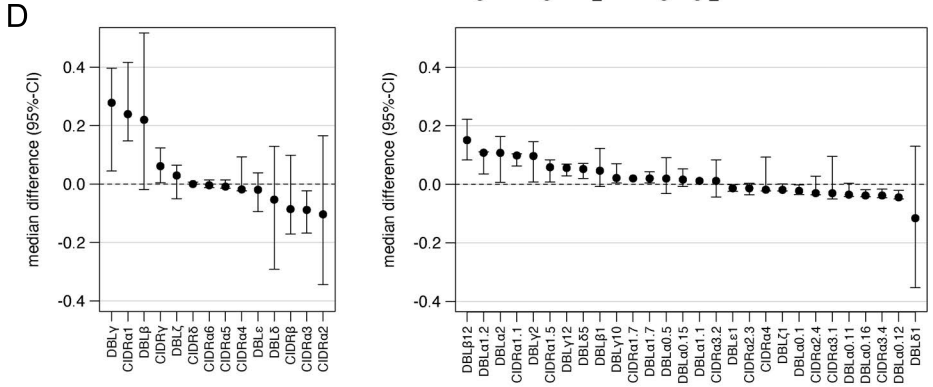


Figure 8

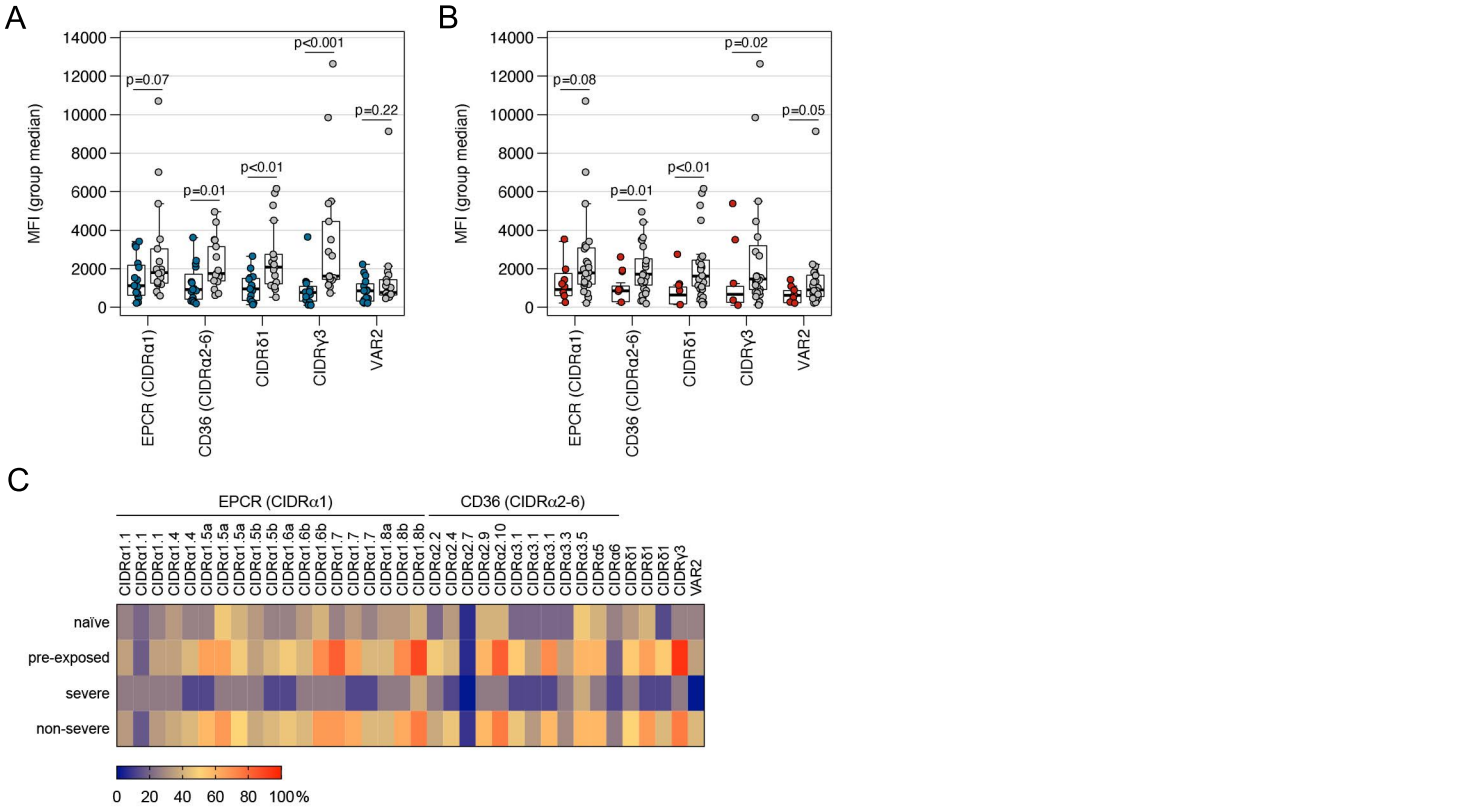
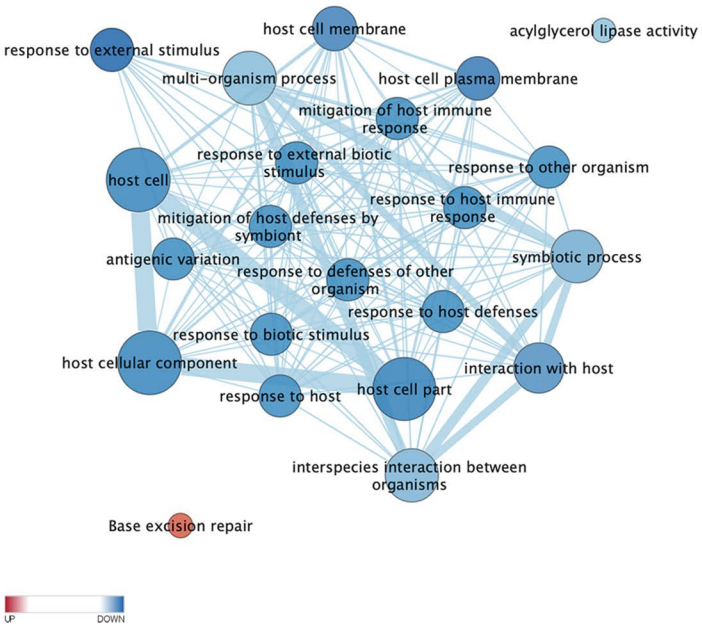


Figure 9

A



B

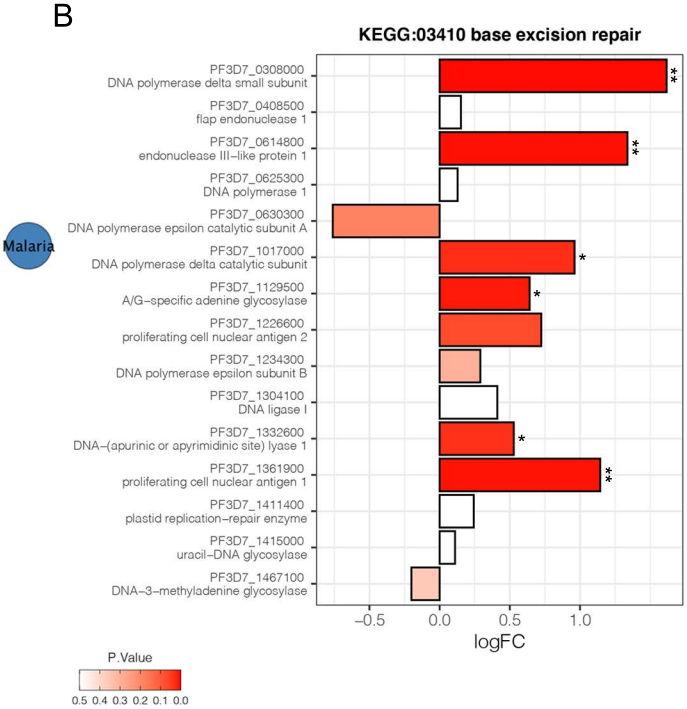


Table 1: Characteristics of each patient.

ID #	Parasitemia	Symptoms	Classification
1	1%	High fever (40,2°C), liver dysfunction (transaminases ↑↑), developed 7% parasitemia with circulating schizonts upon hospitalization	Naïve, severe
2	1%	Fever (38.5°)	Pre-exposed, non-severe
3	>5%	Fever, chill	Naïve, non-severe
4	1.5%	Fever, chill, headache, joint pain	Pre-exposed, non-severe
5	1.5%	Fever, chill	Pre-exposed, non-severe
6	0.5%	Fever (39°C), diarrhea, headache, liver dysfunction (transaminases ↑↑)	Naïve, non-severe
7	2.5%	Fever, 'malaria-associated symptoms'	Pre-exposed, non-severe
9	7%	Fever, chill, joint pain, headache, liver dysfunction (transaminases ↑↑)	Naïve, non-severe
10	2%	Fever (39°C)	Pre-exposed, non-severe
11	3%	Fever, chill	Pre-exposed, non-severe
12	0.5%	Fever, chill, joint pain, headache, diarrhea, abdominal pain	Pre-exposed, non-severe
13	35%	Fever, multiple organ failure, metabolic acidosis, death	Naïve, severe
14	<1%	Fever, diarrhea	Naïve, non-severe
15	35%	Fever, cerebral and liver dysfunction (transaminases ↑↑)	Naïve, severe
16	8%	Fever, chill, headache	Naïve, non-severe
17	2.5%	Fever, diarrhea, headache	Pre-exposed, non-severe
18	7%	Fever, diarrhea, headache, abdominal pain	Pre-exposed, non-severe
19	0.8%	No clinical data	Pre-exposed, non-severe
20	40%	Fever, diarrhea, cerebral dysfunction, renal failure (creatinine ↑↑↑), anemia	Naïve, severe
21	7%	Fever, headache	Naïve, non-severe
22	3%	Fever (up to 41°C), headache, abdominal cramps	Naïve, non-severe
23	12%	Fever, headache, multiple organ dysfunction (renal: creatinine ↑↑, liver: transaminases ↑, cerebrum: aphasia, ataxia)	Naïve, severe
24	3%	Fever, diarrhea	Pre-exposed, non-severe
25	48%	Fever, multiple organ failure, liver dysfunction (transaminases ↑), lactate acidosis	Naïve, severe
26	7%	Fever, headache, cerebral dysfunction	Naïve, severe
27	0.2%	Fever, diarrhea	Pre-exposed, non-severe
28	3.5%	Fever, chill	Pre-exposed, non-severe
29	8%	Fever, headache, diarrhea	Pre-exposed, non-severe
30	11%	Fever, headache, diarrhea, cerebral dysfunction, metabolic acidosis	Naïve, severe
31	1%	No clinical data	Pre-exposed, non-severe
32	2%	Fever, chill, diarrhea, headache	Pre-exposed, non-severe
33	3.5%	Fever, diarrhea, abdominal pain	Pre-exposed, non-severe

Table 2: Patient data.

	First-time infected (naïve) (n=15)	Pre-exposed (n=17)	Severe malaria (n=8)	Non-severe malaria (n=24)
Female sex [n (%)]	6 (40%)	3 (18%)	3 (38%)	6 (25%)
Patient age in years [median (IQR)]	34 (26–53)	38 (31–45)	47 (27–59)	35 (31–46)
Hb g/dl [median (IQR)]*	13.1 (12.1–14.6)	12.2 (11.8–13.1)	12.1 (11.6–13.0)	13.2 (12.0–14.3)
Parasitemia % [median (IQR)]	7.0 (4.0–23.5)	2.0 (1.0–3.0)	23.5 (10.0–36.3)	2.5 (1.0–3.9)
MSP1 [n (%)]	1: 10 (66%) 2: 1 (7%) 3: 3 (20%) 4: 1 (7%)	1: 12 (71%) 2: 3 (18%) 3: 2 (12%) 4: 0 (0%)	1: 5 (63%) 2: 1 (13%) 3: 1 (13%) 4: 1 (13%)	1: 17 (71%) 2: 3 (13%) 3: 4 (17%) 4: 0 (0%)
Total reads [median (IQR)]	41,341,958 (37,804,417–43,659,324)	41,259,082 (36,921,362–43,904,892)	42,458,431 (38,520,154–49,561,881)	41,050,568 (36,920,201–44,030,863)
P. falciparum reads [median (IQR)]	35,940,843 (34,099,395–39,090,313)	37,065,150 (28,707,096–38,070,441)	37,980,501 (35,195,959–45,563,701)	35,559,157 (29,711,534–37,774,576)
Number of assembled *var* contigs (>500 bp) [median (IQR)]	220.5 (169.3–320.8)	165.5 (121.3–251.5)	292 (210–404)	174 (121–259)
Parasite age [median (IQR)]	9.4 (8.0–10.3)	9.8 (8.0–10.6)	8.2 (8.0–9.8)	9.8 (8.2–11.4)

Sides of infection: Ghana (n=10), Nigeria (n=6), Guinea (n=3), Tansania (n=2), Kongo (n=2), other African countries (n=10), Germany (n=1), unknown (n=1)

* n=21

Table 3: *Var* transcripts up- and downregulated in first-time infected patients.

	Transcript	Length (bp)	Log2 fold change	p_{adj}	Comment	Associated *var* group	Rask domain composition
↑ naïve	NS_L1_80_Contig36393	2,536	23.04	2.28e-11	*var1*	A	DBLγ8-DBLζ1-DBLε5
	S_L7_42_Contig2378#	8,504	4.38	0.0009	*var2csa*	E	NTSpam-DBLpam1-DBLpam2-CIDRpam-DBLpam3-DBLεpam4-DBLεpam5-DBLε10
	NS_L1_88_Contig54310	2,975	6.31	0.0027		B	DBLδ1-CIDRβ1
	S_L7_16_Contig1649	2,244	6.31	0.0050	*var1-IT*	A	NTSA-DBLα1.1-CIDRα1.2
	S_L8_68_Contig4767#	6,006	8.45	0.0068	ICAM-1 binding?	B	DBLβ5-DBLβ5-DBLδ1-CIDRβ1
	NS_L7_17_Contig4771	2,561	2.24	0.0243		B	DBLδ1-CIDRγ5
	NS_L1_88_Contig17086	3,568	5.13	0.0243	CD36-binding	B	CIDRα3.4-DBLδ1-CIDRβ5
	S_L7_16_Contig3710	3,037	3.74	0.0319	DC3, *var3*	A	DBLα1.3-DBLε8
	NS_L8_59_Contig2262	5,189	2.15	0.0323	*var2csa*	E	DBLpam3-DBLεpam4-DBLεpam5-DBLε10
	S_L8_56_Contig2546	804	6.58	0.0339	*var1-IT*	A	NTSA-DBLα1.1
	NS_L7_27_Contig7034	1,410	1.26	0.0366		B	CIDRβ5
	NS_L7_27_Contig6972#	724	4.50	0.0403			not applicable
↓ naïve	NS_L8_65_Contig2160#	7,254	-22.85	2.28e-11	*var1-3D7*	A	NTSA-DBLα1.4-CIDRα1.3-DBLβ1-DBLγ15-DBLε1-DBLγ8
	NS_L6_18_Contig1987#	549	-9.63	0.0027	*var1-3D7*	A	NTSA-DBLα1.4
	NS_L7_29_Contig31477	4,497	-8.92	0.0479	CD36-binding	B/C	DBLα0.20-CIDRα3.1-DBLδ1-CIDRβ6

also found in severe cases up-/downregulated

Table 4: *Var* transcripts up- and downregulated in severe cases.

	Transcript	Length (bp)	Log2 fold change	p_{adj}	Comment	Associated *var* group	Rask domain composition
↑ severe	S_L7_42_Contig2378#	8,504	4.79	4.61e-5	*var2csa*	E	NTSpam-DBLpam1-DBLpam2-CIDRpam-DBLpam3-DBLεpam4-DBLεpam5-DBLε10
	S_L7_20_Contig5132	6,276	10.72	0.0007	EPCR- & gC1qR binding	A	DBLα1.2-CIDRα1.6-DBLβ12-DBLγ6-DBLδ1
	S_L8_68_Contig4767#	6,006	7,85	0.0076	ICAM-1 binding?	B	DBLβ5-DBLβ5-DBLδ1-CIDRβ1
	S_L3_13_Contig2037	1,807	8,75	0.0234	ICAM-1 binding?	A/B	DBLβ5
	NS_L8_50_Contig1847	8,254	12.91	0.0319	*var2csa*	E	NTSpam-DBLpam1-DBLpam2-CIDRpam-DBLpam3-DBLεpam4-DBLεpam5-DBLε10
	NS_L7_27_Contig6972#	724	5.09	0.0347			not applicable
	S_L6_19_Contig4243	5,681	7.64	0.0364		A/B	DBLγ17-DBLγ10-DBLδ4-CIDRδ1
↓ severe	NS_L6_24_Contig8626	1,175	-20.35	9.20e-17	*var1-3D7*	A	NTSA-DBLα1.4
	NS_L6_18_Contig1987#	549	-25.45	7.13e-15	*var1-3D7*	A	NTSA-DBLα1.4
	S_L8_56_Contig7448	3,570	-23.55	1.00e-8	*var1-3D7*	A	NTSA-DBLα1.4-CIDRα1.3-DBLβ1
	NS_L8_65_Contig2160#	7,254	-22.46	1.03e-8	*var1-3D7*	A	NTSA-DBLα1.4-CIDRα1.3-DBLβ1-DBLγ15-DBLε1-DBLγ8
	NS_L8_59_Contig4567	10,201	-5.04	0.0234	*var1-3D7*	A	NTSA-DBLα1.4-CIDRα1.3-DBLβ1-DBLγ15-DBLε1-DBLγ8-DBLζ1-DBLε5

also found in naïve cases up-/downregulated

Table 5: *Var* domains up- and downregulated in first-time infected patients.

	Rask domain	Log2 fold change	p_{adj}	Comment	Associated *var* group
↑ naïve	DBLα1.7	3.52	6.36e-8	DC13	A
	CIDRδ1	3.95	0.0003	DC16, rosetting	A
	DBLβ6#	2.45	0.0016	DC15, DC16	A
	DBLα1.2#	2.30	0.0092	DC15	A
	DBLβ9	2.13	0.0092	DC5, PECAM1-binding	A
	CIDRα1.1#	2.38	0.0229	DC8, EPCR-binding	B
	CIDRα1.4	2.28	0.0298	DC13, EPCR-binding	A
	DBLβ12#	1.77	0.0381	DC8, gC1qr-binding	B
	CIDRα1.7	1.64	0.0425	EPCR-binding	A
	DBLα2#	1.86	0.0425	DC8	B
↓ naïve	DBLα0.22	-2.54	0.0035		B, C
	DBLα0.23#	-3.03	0.0035		B, C
	CIDRα6	-1.68	0.0092	DC22, CD36-binding	B, C
	CIDRγ11	-3.11	0.0092		B, C
	NTSB	-1.55	0.0105		B, C
	DBLα0.13	-2.28	0.0216		B
	DBLγ15	-2.77	0.0381	DC1, *var1-3D7*	A, *var1*
	CIDRα2.8	-2.41	0.0425	CD36-binding	B
	CIDRα2.9#	-2.15	0.0425	CD36-binding	B
	DBLα1.4	-2.58	0.0425	DC1/4, *var1-3D7*	A, *var1*
	DBLε5	-1.35	0.0499	DC1, *var1*	A, *var1*

also found in severe cases up-/downregulated

Table 6: *Var* domains up- and downregulated in severe cases.

	Rask domain	Log2 fold change	p_{adj}	Comment	Associated *var* group
↑ severe	DBLβ6#	2.87	0.001	DC15, DC16	A
	CIDRα1.1#	3.02	0.009	DC8, EPCR-binding	B
	DBLα1.2#	2.49	0.022	DC15/(DC8)	A
	DBLβ12#	2.19	0.023	DC8, gC1qr-binding	B
	DBLα1.6	2.89	0.023	DC16	A
	DBLα2#	2.23	0.042	DC8	B
↓ severe	DBLα0.23#	-3.61	0.006		B, C
	CIDRα1.3	-2.09	0.006	DC1, *var1-3D7*	A, *var1*
	CIDRα2.4	-2.31	0.009	CD36-binding	B
	CIDRα2.9#	-2.74	0.039	CD36-binding	B
	DBLα1.5	-2.64	0.042	DC16, rosetting	A

also found in naïve cases up-/downregulated

Table 7: *Var* blocks up- and downregulated in first-time infected patients.

	Rask block	Log2 fold change	p_{adj}	Comment	Associated *var* group	Frequence on VarDom 1.0 server
↑ naïve	557	4.33	0.0003	interdomain (DBLβ-DBLγ)	B	5
	584	4,90	0.0003	DBLβ3	A	5
	614	4.59	0.0174	DBLβ (6x DBLβ12, 1x DBLβ3)	A, B/A	7
	155	3.03	0.0427	NTSA	A	37
	255	2.79	0.0427	DBLβ (8x DBLβ12, 6x DBLβ3, 3x DBLβ1, 1x DBLβ5)	A, B/A	18
↓ naïve	88	-1.41	0.0427	DBLα0	B, C	75
	269	-1.26	0.0446	ATSB	B, C	16

Table 8: *Var* blocks downregulated in severe cases.

	Rask block	Log2 fold change	p_{adj}	Comment	Associated *var* group	Frequence on VarDom 1.0 server
↓ severe	591	-23.93	3.90E-12	Interdomain (CIDRβ/γ-ATSB)	B	5
	559	-4.62	0.03999292	DBLζ6	B	5