Gene expression profiling of malaria parasites reveals common virulence gene expression in adult first-time infected patients and severe cases

Jan Stephan Wichers[1,2,3], Gerry Tonkin-Hill[4], Thorsten Thye[5], Ralf Krumkamp[5,6], Benno Kreuels[7,8], Jan Strauss[1,2,3], Heidrun von Thien[1,2,3], Judith Anna Marie Scholz[1], Helle Smedegaard Hansson[9], Rasmus Weisel Jensen[9], Louise Turner[9], Freia-Raphaella Lorenz[10], Anna Schöllhorn[10], Iris Bruchhaus[1,3], Egbert Tannich[5,6], Rolf Fendel[10,11], Thomas Dan Otto[12], Thomas Lavstsen[9], Tim-Wolf Gilberger[1,2,3], Michael Frank Duffy[13], Anna Bachmann[1,2,3,6,#]

[1] Molecular Biology and Immunology, Bernhard Nocht Institute for Tropical Medicine, 20359 Hamburg, Germany
[2] Centre for Structural Systems Biology, 22607 Hamburg, Germany
[3] Biology Department, University of Hamburg, 22609 Hamburg, Germany
[4] Wellcome Sanger Institute, Hinxton/Cambridge CB10 1SA, UK
[5] Epidemiology and Diagnostics, Bernhard Nocht Institute for Tropical Medicine, 20359 Hamburg, Germany
[6] German Center for Infection Research (DZIF), partner site Hamburg-Borstel-Lübeck-Riems, Germany
[7] Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, 20359 Hamburg, Germany & I. Department of Medicine, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany
[8] Department of Medicine, College of Medicine, Blantyre, Malawi
[9] CMP, University of Copenhagen, 2200 Copenhagen, Denmark
[10] Institute of Tropical Medicine, University of Tübingen, 72074 Tübingen, Germany
[11] German Center for Infection Research (DZIF), partner site Tübingen, Germany
[12] Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow G12 8TA, UK
[13] 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 each parasite a similar array of human endothelial receptor binding phenotypes. Here, we analyzed RNA-seq profiles of parasites isolated from 32 *P. falciparum* infected adult 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. For differential expression analysis of *Pf*EMP1-encoding *var* gene transcripts were *de novo* assembled from RNA-seq data and, in parallel, *var* expressed sequence tags were analyzed and used to predict the encoded domain composition of the transcripts. Both approaches showed in concordance that severe malaria was associated with *Pf*EMP1 containing the endothelial protein C receptor (EPCR)-binding CIDRα1 domain, whereas CD36-binding *Pf*EMP1 was linked to non-severe malaria outcomes. First-time infected adults were more likely to develop severe symptoms and tended to be infected for a longer period. Thus, parasites with more pathogenic *Pf*EMP1 variants are more common in patients with a naïve immune status and/or adverse inflammatory host responses to first infections favors growth of EPCR-binding parasites.

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

Introduction

Despite considerable efforts during recent years to combat malaria, the 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 2019, there were 229 million cases of malaria worldwide, resulting in more than 400,000 deaths (WHO, 2020). Currently, about half of the world's population lives in infection-prone areas, and more than 90% of the malaria deaths occur in Africa. In particular, children under five years of age and pregnant women suffer from severe disease, but adults from areas of lower endemicity and non-immune travelers are also vulnerable to severe malaria. Both, children and adults are affected by cerebral malaria, but the prevalence of different features of severe malaria differs with increasing age. Anemia and convulsions are more frequent in children, jaundice indicative of hepatic dysfunction and oliguric renal failure are the dominant manifestations in adults (Dondorp *et al*, 2008; World Health Organization (WHO), 2014). Moreover, the mortality increases with age (Dondorp *et al*, 2008) and was previously determined as a risk factor for

severe malaria and fatal outcome in non-immune patients, but the causing factors are largely unknown (Schwartz *et al*, 2001).

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 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 (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 DBLα-ζ and pam) and cysteine-rich inter-domain region domains (CIDR; main classes CIDRα-δ and pam) (Rask *et al*, 2010). These domains were initially allocated to subclasses (e.g., the DBLβ subclasses 1 – 13), however, due to frequent recombination between members of the different subclasses, many of these are indistinct and poorly defined (Otto *et al*, 2019).

*Pf*EMP1 molecules have been grouped into four categories (A, B, C and E) depending on the type of N-terminal domains (the *Pf*EMP1 "head structure") as well as the 5' upstream sequence, the chromosomal localization and the direction of transcription of their encoding *var* gene (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 among the different groups (Rask *et al*, 2010). About 20% of *Pf*EMP1 variants belong to group A and are typically longer proteins with a head-structure containing DBLα1 and either an EPCR-binding CIDRα1 domain or a CIDRβ/γ/δ domain of unknown function. Further, the group A includes two conserved, strain-transcendent subfamilies: the *var1* subfamily (previously known as *var1csa*), found in two different variants in the parasite population (3D7- and IT-type) (Otto *et al*, 2019), and the *var3* subfamily, the shortest *var* genes with only two extracellularly exposed domains (DBLα1.3 and DBLε8). Group B and C includes most, about 75% of *Pf*EMP1, and typically have DBLα0-CIDRα2-6 head structures binding CD36, followed by a DBLδ1-CIDRβ/γ domain combination. A particular subset of B-type proteins, also known as group B/A chimeric genes, possess a chimeric DBLα0/1 domain (a.k.a. DBLα2) and an EPCR-binding CIDRα1 domain. Thus, the head structure confers mutually exclusive binding properties, either to EPCR, CD36 or to an unknown receptor via the CIDRβ/γ/δ do-

mains. C-terminally to the head-structure, most group A and some group B and C *Pf*EMP1 have additional DBL domains, of which specific subsets of DBLβ domains of group A and B *Pf*EMP1 bind ICAM-1 (Lennartz *et al*, 2017; Janes *et al*, 2011) and another DBLβ domain, DBLβ12, has been suggested to bind gC1qR (Magallón-Tejada *et al*, 2016). The consistent co-occurrence of specific domain subsets in same *Pf*EMP1 gave rise to the definition of domain cassettes (DC) (Otto *et al*, 2019; Berger *et al*, 2013; Rask *et al*, 2010). The best example of this is the VAR2CSA *Pf*EMP1 (group E, DC2), which binds placental CSA and causes pregnancy-associated malaria (Salanti *et al*, 2004). The VAR2CSA proteins share domain composition, their encoding genes are less diversified than other *var* groups and all parasites possess one or two *var2csa* copies. Another example is the chimeric group B/A *Pf*EMP1 also known as DC8, which includes the DBLα2, specific CIDRα1.1/8 subtypes capable to bind EPCR and typically DBLβ12 domains.

Due to the sequence diversity of *var* genes, studies of *var* expression in patients have relied on analysis of DBLα expressed sequence tags (EST) (Warimwe *et al*, 2009, 2012) informing on relative distribution of different *var* transcripts and qPCR primer sets covering some but not all subsets of DBL and CIDR domains (Lavstsen *et al*, 2012; Mkumbaye *et al*, 2017). So far, only very few studies (Tonkin-Hill *et al*, 2018; Andrade *et al*, 2020; Duffy *et al*, 2016; Kamaliddin *et al*, 2019) have used the RNA-seq technology to quantify assembled *var* transcripts *in vivo.* Moreover, most studies have focused on the role of *Pf*EMP1 in severe pediatric malaria. Consensus from these studies is, that severe malaria in children is associated with expression of *Pf*EMP1 with EPCR-binding CIDRα1 domains (Jespersen *et al*, 2016; Kessler *et al*, 2017; Storm *et al*, 2019; Shabani *et al*, 2017; Mkumbaye *et al*, 2017; Magallón-Tejada *et al*, 2016), but elevated expression of dual EPCR and ICAM-1-binding *Pf*EMP1 (Lennartz *et al*, 2017) and the group A associated DC5 and DC6 have also been associated with severe disease outcome (Magallón-Tejada *et al*, 2016; Avril *et al*, 2013, 2012; Claessens *et al*, 2012; Lavstsen *et al*, 2012; Duffy *et al*, 2019). Less effort has been put into understanding the role of *Pf*EMP1 in relation to severe disease in adults, and its different symptomatology and higher fatality rate. Two gene expression studies from regions of unstable transmission in India showed elevated expression of EPCR-binding variants (DC8, DC13) and DC6 (Bernabeu *et al*, 2016; Subudhi *et al*, 2015), but also of transcripts encoding B- and C-type *Pf*EMP1 in severe cases (Subudhi *et al*, 2015).

In this study, we applied an improved genome-wide expression profiling approach using RNA-seq to study gene expression, in particular *var* gene expression, in *P. falciparum* parasites from hospitalized adult travelers and combined it with a novel prediction analysis of *var* transcripts from DBLα EST. Individuals were clustered into i) first-time infected (n=15) and ii) pre-exposed (n=17) individuals on the basis of serological data or into iii) severe (n=8) and iv) non-severe (n=24) cases according to medical reports. Our multi-dimensional analysis

revealed 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 variant were expressed at higher levels in pre-exposed patients and non-severe cases. Interestingly, severe complications occurred only in the group of first-time infected patients who also tended to be infected for a longer period, indicating that severity of infection in adults is dependent on duration of infection, host immunity and parasite virulence gene expression.

Results

Cohort characterization

This study is based on a cohort of 32 adult malaria patients hospitalized in Hamburg, Germany. All patients had fever indicative of symptomatic malaria. MSP1 genotyping estimated a low number of different parasite genotypes present in the patients (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 infections (Drakeley *et al*, 2005). Principal component analysis (PCA) 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 parasitophorous vacuolar membrane-enclosed merozoite structures (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 ap-

proach to build models predictive of individual's protective status. A multidimensional scaling plot was used to visualize cluster allocation confirming the classification of patient #21 as being non-immune (Figure 1E). The patient #26, positioned at the borderline to pre-exposed patients, was grouped into the naïve cluster in accordance with the Luminex data and the patient statement that this potentially pre-exposed patient returned from his first trip to Africa. The calculated variable importance highlighted the relevance of the mADRB assay, the ELISA and the Luminex to allocate patients into cluster (Figure 1F).

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 broad range of parasite antigens, especially typical parasite blood stage markers, including MSP1, MSP2, MSP4, MSP10, EBA175, REX1, and AMA1 (Figure 1D, upper panel). Markers for a recent infection, MSP1, MSP4, GLURP and 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).

Eight patients from the malaria-naïve group were considered as having severe malaria based on the predefined criteria. The remaining 24 cases were assigned to the non-severe malaria group (Figure 1G, Table S2). Comparing the IgG antibody response of severe and non-severe cases within the previously malaria-naïve group, elevated antibody levels were found in the severe subgroup. The highest fold change was observed for 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 (Figure 2A). Transcriptome libraries were sequenced for all 32 patient samples (NCBI BioProject ID: PRJNA679547). 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) for the 30 samples included in the *de novo* assembly (Table S3). Variation in parasite ages – defined as the progression of the intraerythrocytic development cycle measured

in hours post invasion – 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 or in median age (Table 1, Figure S2). However, a small, statistically not significant bias (p=0.17) 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) (Table 1, Figure S2D). None of the samples revealed high proportions of late trophozoites (all <3%), schizonts (0%) or gametocytes (all <6%) (Figure S2A, B). The estimated proportions were used to control for differences in parasite stage between samples by including them as covariates in the regression analysis of differential core gene expression (Figure 2A, Figure S3).

Genome-wide analysis of differential gene expression

Global gene expression analysis according to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) identified 420 genes to be higher and 236 to be lower expressed ($p \leq 0.05$) in first-time infected patients, together corresponding to 11.3% of all *P. falciparum* genes (Table S4). Similar, 362 genes were significantly higher and 219 genes lower expressed in severe cases (Table S5). A gene set enrichment analysis (GSEA) of the differentially expressed genes using Gene Ontology (GO) terms and KEGG pathway annotations showed that the KEGG pathway 03410 'base excision repair' facilitating the maintenance of genome integrity by repairing small bases lesions in the DNA was expressed at significantly higher levels in first-time infected patient samples (Figure 2B, Figure S4). In total, six out of 15 *P. falciparum* genes included in this KEGG pathway were found to be statistically significant enriched upon first-time infection, including 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 2C). Additionally, a significantly lower expression level for genes associated with several GO terms involved in antigenic variation and host cell remodeling was found in first-time infected patients (Figure 2B, Table S4) and severe cases (Table S5).

As variant surface antigens like *var*, *rif* and *stevor* are largely clone-specific, analysis of reads from the clinical isolates mapping to homologous regions in 3D7 genes would be distorted and flawed. Therefore, we analyzed *var* gene expression by first *de novo* assembling *var* genes from the RNA-seq data and subsequently analyzing the expression of specific *var* gene subsets or according to the domains encoded. In addition, we manually screened dif-

252 ferentially expressed genes known to be involved in *var* gene regulation or correct display of
253 *Pf*EMP1 at the host cell surface (Table S4, S5).
254
255 ***Differential* var *gene expression***
256 To correlate individual *var* genes or common *var* gene-encoded traits (Figure 3) with a naïve
257 immune status or disease severity, differential *var* transcript levels were analyzed as in Ton-
258 kin-Hill *et al.* (Tonkin-Hill *et al*, 2018)(Figure 2A, Figure S3). *Var* transcripts were assembled
259 from each patient sample separately, and annotated. In total, 6,441 contigs with over 500 bp-
260 length were generated with an N50 of 2,302 bp and a maximum length of 10,412 bp (Data
261 S1). A median of 200 contigs (IQR: 137–279) with >500 bp was assembled per sample. One
262 or more DBL or CIDR domains could be annotated to 5,488 of the contigs, whereas the re-
263 maining contigs could only be annotated by these domains smaller building blocks, the so-
264 called homology blocks defined by Rask *et al.* (Rask *et al*, 2010) (Table S6, S7).
265
266 *Differential* var *transcript levels*
267 We first looked for highly similar transcripts present in multiple samples. The Salmon RNA-
268 seq quantification pipeline (Patro *et al*, 2017), which identifies equivalence classes allowing
269 reads to contribute to the expression estimates of multiple transcripts, was used to estimate
270 expression levels for each transcript. Due to the high diversity in *var* genes, mainly assem-
271 bled transcripts of the strain-transcendent variants *var1*, *var2csa* and *var3* were found to be
272 differentially expressed. Notably, the *var1-IT* variant was expressed at higher levels in para-
273 sites from first-time infected patients, whereas the *var1-3D7* variant was expressed at higher
274 levels in parasites from pre-exposed and non-severe patients (Figure 4A, B, Figure 5A, B,
275 Table S9, S10). This was confirmed by mapping normalized reads from all patients to both
276 *var1* variants as well *as var2csa* (Figure S5). Beyond the conserved variants, several *var*
277 fragments from B- or C-type *var* genes were associated with a naïve immune status and
278 three transcripts from A, DC8 and B-type *var* genes as well as *var2csa* were linked to severe
279 malaria patients (Figure 4, 5, Table S9, S10).
280
281 *Differential* var *domain transcript levels*
282 To assess differential expression of specific *var* domains, read counts corresponding to do-
283 mains with the same classification were calculated. This showed that different EPCR-binding
284 CIDRα1 domain variants and other domains found in DCs with CIDRα1 domains were ex-
285 pressed at significantly higher levels in first-time infected patients (Figure 4C–E, Table S9).
286 Specifically, besides domains from DC8 (DBLα2, CIDRα1.1, DBLβ12) and DC13 (DBLα1.7,
287 CIDRα1.4), the CIDRα1.7 and DBLα1.2 (DC15) were increased upon infection of malaria-
288 naïve individuals. The DBLα1.2 domain was in all of the 32 gene assemblies flanked by an

EPCR-binding CIDRα1 domain, 56% of these were a CIDRα1.5 domain (Table S6, S7). In addition, parasites from first-time infected patients showed a significantly higher level of transcripts encoding the CIDRδ1 domain of DC16 (DBLα1.5/6-CIDRδ1/2), and the DBLβ6 domain (Figure 4D–E). The DBLβ6 is associated with A-type *var* genes and can be found adjacent to DC15 and DC16 (Otto *et al*, 2019) (Table S6, S7). In general, domains associated with the same domain cassette showed the same trend even if some domains did not reach statistical significance set at $p < 0.05$ (Table S9).

Domains found expressed at lower levels in malaria-naïve included group B and C N-terminal head structure domains NTSB, DBLα0.13/22/23 and CD36-binding CIDR domains (CIDRα2.8/9,6) as well as the C-terminal CIDRγ11 domain and domains of the *var1-3D7* variant (DBLα1.4, DBLγ15, DBLε5) (Figure 4C–E).

When comparing severe to the non-severe cases, domains of DC8 (DBLα2, CIDRα1.1, DBLβ12), DC15 (DBLα1.2) and DC16 (DBLα1.6) and the A-type linked DBLβ6 were found associated with severe disease. Domain types expressed at significantly higher levels in non-severe cases included N-terminal head structure domains, DBLα0.23, CIDRα2.4/9 from group B and C *Pf*EMP1, DC16 (DBLα1.5) and the CIDRα1.3 domain from the *var1-3D7* allele (Figure 5C–E, Table S10).

As DBL and CIDR sub-classes are poorly defined (Otto *et al*, 2019) and different domain subclasses confer the same binding phenotype (Higgins & Carrington, 2014), 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 4F, 5F). This showed significant differences for domains associated with EPCR- or CD36-binding *Pf*EMP1. As expected, domains associated with EPCR-binding as well as the CIDRγ3 domain were expressed at higher levels in naïve and more severe cases, whereas domains associated with the CD36-binding phenotype were higher expressed in pre-exposed and non-severe patients.

Differential var *homology block transcript levels*

*Pf*EMP1 domains have been described as composed of 628 homology blocks (Rask *et al*, 2010). Homology block expression levels were obtained by aggregating read counts for each block after first identifying all occurrences of the block within the assembled contigs. Transcripts encoding blocks number 255, 584 and 614, all typically located within DBLβ domains of DC8 and CIDRα1-containing type A *Pf*EMP1 (Figure 4G, H, Table S9), number 557, located in the interdomain 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. 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 (Figure 4G, H, Table S9). No homology blocks were associated with severe cases, but two blocks, 591 and 559, found in group B *Pf*EMP1 were found to be lower expressed in severe malaria cases (Figure 5G, H, Table S10).

Var expression profiling by DBLα-tag sequencing

To supplement the RNA-seq analysis with an orthogonal analysis, we conducted deep sequencing of RT-PCR-amplified DBLα ESTs from 30 of the patient samples (Lavstsen *et al*, 2012) (Figure 2A). Between 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) ratio of 0.4 (CI of 95%: -2.5–3.3) determined by Bland-Altman plotting (Figure S6). Around 82.6% (median) of all unique DBLα-tag sequences detected with >10 reads (92.9% of all DBLα-tag sequences) were found in the RNA-seq approach; and 81.8% of the upper 75th percentile of RNA-seq contigs (spanning across the DBLα-tag region) were found by the DBLα-tag approach.

Using the Varia tool (Mackenzie *et al*, 2020) the domain composition of the *var* genes from which the DBLα-tag sequences originated were predicted. The tool searches an extended varDB containing >200K annotated *var* genes for genes with near identical DBLα-tag sequences and returns the consensus domain annotation among the hit sequences. A partial domain annotation was made for ~85% of the DBLα-tag sequences (Figure 2A, Table S11). In line with the RNA-seq data, this analysis showed that DBLα1 and DBLα2 sequences were enriched in first-time infected and severe malaria patients. Conversely, a significant higher proportion of DBLα0 sequences was found in pre-exposed individuals and mild cases (Figure 6A, 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 was observed in first-time infected patients and severe cases (Figure 6A, B). Prediction of the NTS and CIDR domains flanking the DBLα domain showed a significantly 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 6A, B). Subsequent 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 S7). Assessing expression in relation to domain subtype, CIDRα1.1/5, DBLβ12, DBLγ2/12, DBLα2, DBLα1.2/2

and DBLδ5 (together with DBLγ12 components of the DC5) were found associated with severe malaria, while CIDRα3.1/4, DBLα0.12/16, and DBLδ1 associated with non-severe cases (Figure S7).

Overall, these data corroborate 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. In general, 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 with significant differences for CIDRα2–6, CIDRδ1 and CIDRγ3, but not for EPCR-binding CIDRα1 domains (Figure 7A, B).

The data was also analyzed using the average MFI reactivity (plus two standard deviations) of a Danish control cohort as a cut off for seropositivity to calculate the coverage of antigen recognition (Table S1) (Cham *et al*, 2010). In this analysis, almost half of the tested antigens were recognized by pre-exposed (median: 47.2%) and non-severe patients (median: 44.4%), but only 1/4 of the antigens were recognized by first-time infected patients (median: 25.0%) and 1/20 by severely ill patients (median: 5.6%). *Pf*EMP1 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 7C, Table S1).

Taken together, this analysis indicates that higher levels of antibodies against severe malaria-associated EPCR-binding variants are present in pre-exposed and non-severe cases, which might have selected against parasites expressing CIDRα1 domains during the current infection.

Discussion

Non-immune travelers and adults from areas of unstable malaria transmission are prone to severe malaria. Currently, scarce information on the *Pf*EMP1-mediated pathogenicity re-

399 sponsible for the different symptomatology in comparison to pediatric severe malaria and the
400 higher fatality rate in adults is available. Here we present the first in-depth gene expression
401 analysis of 32 *ex vivo* blood samples from adult travelers using RNA-seq and expressed se-
402 quence tag analyses. Despite the relatively low number of patient samples recruited in 5
403 years, our data confirmed previously reported associations between transcripts encoding
404 type A and B EPCR-binding *Pf*EMP1 and infections in naïve hosts and disease severity (Ta-
405 ble S6, S7) (Duffy *et al*, 2019; Tonkin-Hill *et al*, 2018; Kessler *et al*, 2017; Bernabeu *et al*,
406 2016; Jespersen *et al*, 2016; Lavstsen *et al*, 2012). Our results further suggests that parasite
407 interaction with EPCR is linked to severe disease in children as well as in adults. However,
408 since CIDRα1-containing *Pf*EMP1s possess multiple binding traits (Lennartz *et al*, 2017;
409 Magallón-Tejada *et al*, 2016), co-interaction with other receptors may further increase the
410 risk for severe malaria.

411 Overall, there was a high degree of consensus between observations made on *var* expres-
412 sion analyzed at different levels of *Pf*EMP1 domain annotation. Stratifying *var* gene expres-
413 sion according to different main and subtype of DBL and CIDR domains, showed only A- and
414 DC8-type *Pf*EMP1 domains, and predominantly those linked to EPCR-binding *Pf*EMP1, to be
415 associated with first-time infections. Conversely, domains typical for CD36-binding *Pf*EMP1
416 proteins were found at higher levels in malaria-experienced adults. Specifically, expression of
417 *Pf*EMP1 domains included in DC8, DC13 and DC15 as well as all EPCR-binding CIDRα1
418 domains were associated with first-time adult infections, whereas DBLα0 and CD36-binding
419 CIDRα2-6 domains were linked to pre-exposed individuals. These differences were largely
420 due to the differential expression between the first-time infected patients with more severe
421 symptoms and patients with non-severe malaria. Here, domains of DC8 and DC15 as well as
422 all DBLα1/2 and CIDRα1 domains were associated with severe symptoms, while NTSB,
423 DBLα0, CIDRα2-6 domains including specific subsets of CIDRα2 were linked to non-severe
424 symptoms. These conclusions were closely mirrored in the DBLα-tag analysis, and was fur-
425 ther corroborated by the differential RNA-seq expression stratified according to the smaller
426 homology blocks, which identified mainly homology blocks of DBLβ1, 3, 5 and 12 to be asso-
427 ciated with first-time infected patients. These DBLβ domains are parts of DCs associated
428 with EPCR-binding, so it is hard to distinguish between co-occurring domains and clear as-
429 sociations.

430 In addition, three other group A *Pf*EMP1-associated domains, CIDRγ3, CIDRδ from the
431 DC16, DBLβ9 from DC5, and DBLβ6 were found associated with first-time infected patients.
432 DBLβ9 and DBLβ6 could have been detected due to its presence C-terminally to some
433 EPCR-binding or A-type *Pf*EMP1. However, the CIDRδ domain of DC16 (DBLα1.5/6-
434 CIDRδ1/2) constitute a different subset of A-type *Pf*EMP1, which together with CIDRβ2 and
435 CIDRγ3-containing group A *Pf*EMP1 (found in DC11) may be associated with rosetting (Carl-

436 son *et al*, 1990; Ghumra *et al*, 2012). Direct evidence that any of these CIDR domains have
437 intrinsic rosetting properties is lacking (Rowe *et al*, 2002). Rather, their association with
438 rosetting may be related to their tandem expression with DBLα1 at the N-terminal head
439 (Ghumra *et al*, 2012). The DC16 group A signature was not associated with severe disease
440 outcome in previous DBLα-tag studies or qPCR studies by Lavstsen *et al.* (Lavstsen *et al*,
441 2012) and Bernabeu *et al.* (Bernabeu *et al*, 2016), but DBLα1.5/6 and CIDRδ of DC16 were
442 enriched in cerebral malaria cases with retinopathy in the study of Shabani *et al.* (Shabani *et*
443 *al*, 2017) and Kessler *et al.* (Kessler *et al*, 2017) using the same qPCR primer set. Also, as-
444 sociation of DC11 with severe malaria in Indonesia was found using the same RNA-seq ap-
445 proach as used here (Tonkin-Hill *et al*, 2018). Rosetting is thought to enhance microvascular
446 obstruction but its role in severe malaria pathogenesis remains unclear (McQuaid & Rowe,
447 2020). However, together with previous observations our data suggest that pediatric cerebral
448 malaria infections are dominated by the expansion of parasites expressing EPCR-binding
449 domains accompanied by parasites expressing other group A *Pf*EMP1, possibly acting as
450 rosetting variants. This was consistent with pre-exposed patients more frequent recognition
451 of CIDRα1 domains than CIDRα2-6 domains, indicating that IgG against EPCR-binding
452 CIDR domains were acquired before IgG to other CIDR domains, as has been observed for
453 malaria endemic populations (Obeng-Adjei *et al*, 2020; Cham *et al*, 2009, 2010; Turner *et al*,
454 2015).

455 Data from adult cohorts are rather limited restricting our comparison mainly to three *var* gene
456 expression studies based on qPCR or a custom cross-strain microarray (Argy *et al*, 2017;
457 Bernabeu *et al*, 2016; Subudhi *et al*, 2015), but the majority of cases from the Indonesian
458 RNA-seq study are also adults (Tonkin-Hill *et al*, 2018). A high expression of A- and B-type
459 *var* genes and an association of DC4, 8 and 13 with disease severity has been reported for
460 malaria cases imported to France (Argy *et al*, 2017) and parasites from severe Indian adults
461 show an elevated expression of DC6 and 8 (Bernabeu *et al*, 2016) and DC13 (Subudhi *et al*,
462 2015). In the severe cases from Indonesia mainly the DCs 4, 8 and 11 were found on an
463 elevated level (Tonkin-Hill *et al*, 2018). All studies are in agreement with the expression data
464 from our cohort of adult travelers, although we here in addition found a higher expression of
465 DC15 (EPCR binding) and DC16 (putative rosetting variants) *var* genes in malaria-naïve pa-
466 tients.

467 Measurements of the total abundances of the different *var* groups are challenging (e.g., due
468 differences in the coverage or sensitivity of qPCR or DBLα-tag primer pairs targeting the dif-
469 ferent *var* groups), but there is cumulative evidence that CIDRα1 containing transcripts are
470 dominating the infection in children with severe malaria, severe anemia and cerebral malaria
471 and transcripts with CIDRα2-6 domains are most abundantly expressed during uncomplicat-
472 ed malaria (Jespersen *et al*, 2016; Duffy *et al*, 2019; Warimwe *et al*, 2009, 2012). Although

the median expression of CIDRα2-6 is lower in first-time infected and severe cases compared to pre-exposed and non-severe cases, in most of our adult patients CD36-binding *var* transcripts appear to dominate the expression pattern. This is in concordance with all three other adult studies also indicating a substantial expression of B- and C-type variants associated with binding of CD36 (Argy *et al*, 2017; Bernabeu *et al*, 2016; Subudhi *et al*, 2015) and Subudhi *et al.* even show an association with complicated adult malaria (Subudhi *et al*, 2015). Maybe parasite binding to CD36 is specifically enhanced in adult severe malaria cases compared to children, which is interesting due to their different disease symptomatology (Dondorp *et al*, 2008; Schwartz *et al*, 2001). Alternatively, our adult cohort differs not only in age but also in terms of disease severity from pedatric cohorts, and less sick patients may simply have a less dominant expression of EPCR-binding variants. However, for the parasite's survival and transmission, it may be highly beneficial to express more of the less virulent *Pf*EMP1 variants able to bind CD36. This interaction may not, or is less likely to, result in obstruction of blood flow, inflammation and organ failure at least of the brain, where CD36 is nearly absent (Turner *et al*, 1994).

To the best of our knowledge, this study is the first description of expression differences between the two *var1* variants, 3D7 and IT. The *var1-IT* variant was found enriched in parasites from first-time infected patients, whereas several transcripts of the *var1*-3D7 variant 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; Dahlbäck *et al*, 2007) and is annotated as a pseudogene in 3D7 due to a premature stop codon. 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* variants that were recently identified by comparing *var* gene sequences from 714 *P. falciparum* genomes (Otto *et al*, 2019), and to date it is still unclear if both variants fulfill the same function or have the same characteristics previously described. Overall, the *var1* gene – and the first 3.2 kb of the 3D7 variant in particular – seems to be under high evolutionary pressure (Otto *et al*, 2019) and both variants 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 variants, VAR1-3D7 and VAR1-IT, may have different roles during disease, however, this remains to be determined in future studies.

In summary, our data show a significant increase in transcripts encoding EPCR-binding and other A-type variants in parasites from severe and first-time infected patients, conversely transcripts of CD36-binding variants are found more frequently in parasites from non-severe

and pre-exposed patients. Since CD36-binding variants are still overrepresented in all groups of adult malaria patients we postulate that the parasite population in first-time infected individuals may have broad binding potential after liver release as there is no pre-existing immunity to clear previously experienced *Pf*EMP1 variants. During the blood stage infection selection towards EPCR-binding and other A-type variants, which may confer a parasite growth advantage and also increase the risk for severe malaria, may already have occurred in our adult severe malaria patients indicated by the longer period of infection.

Material and methods

Sample collection and ethics statement

The study was conducted according to the principles of the Declaration of Helsinki in its 6th revision as well as International Conference on Harmonization–Good Clinical Practice (ICH-GCP) guidelines. All 32 patients were treated as in- or outpatients in Hamburg, Germany. Patients were either seen in the outpatient clinic of the University Medical Center Hamburg-Eppendorf (UKE) at the Bernhard Nocht Institute for Tropical Medicine, treated as inpatients at the UKE or at the Bundeswehrkrankenhaus Hamburg. Blood samples for this analysis were collected after patients were informed about the aims and risks of the study and signed an informed consent form for voluntary blood draw (n=21). In the remaining cases, no designated blood samples were drawn, instead remains from diagnostic blood samples were used (n=11). The study was approved by the relevant ethics committee (Ethical Review Board of the Medical Association of Hamburg, reference numbers PV3828 and PV4539).

Blood sampling and processing

EDTA blood samples (1–30 mL) were obtained from the adult patients. The plasma was separated by centrifugation and immediately stored at -20°C. Erythrocytes were isolated by Ficoll gradient centrifugation followed by filtration through Plasmodipur filters (EuroProxima) to clear the remaining granulocytes. An aliquot of erythrocytes (about 50–100 µl) was separated and further processed for gDNA purification. At least 400 µl of purified erythrocytes were rapidly lysed in 5 volumes pre-warmed TRIzol (ThermoFisher Scientific) and stored at -80°C until further processing.

Serological assays

Luminex assay

The Luminex assay was conducted as previously described using the same plex of antigens tested (Bachmann *et al*, 2019). In brief, plasma samples from patients were screened for individual recognition of 19 different CIDRα1, 12 CIDRα2–6, three CIDRδ1 domains and a

single CIDRγ3 domain as well as of the controls AMA1, MSP1, CSP, VAR2CSA (VAR2), tetanus toxin (TetTox) and BSA. The data are shown as mean fluorescence intensities (MFI) allowing comparison between different plasma samples, but not between different antigens. Alternatively, the breadth of antibody recognition (%) was calculated using MFI values from Danish controls plus two standard deviations (SD) as cut off.

Merozoite-triggered antibody-dependent respiratory burst (mADRB)

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

ELISA

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

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 im-

584 munity against the parasite (Obiero *et al*, 2019; Dent *et al*, 2016; Doolan *et al*, 2008; Felgner
585 *et al*, 2013). For the detection of binding antibodies, secondary IgG antibody (goat anti-
586 human IgG QDot™800, Grace Bio-Labs #110635), secondary IgM antibody (biotin-SP-
587 conjugated goat anti-human IgM, Jackson ImmunoResearch #109-065-043) and Qdot™585
588 Streptavidin Conjugate (Invitrogen #Q10111MP) were used (Taghavian *et al*, 2018).

589 Study serum samples as well as the positive and European control sera were diluted 1:50 in
590 0.05X Super G Blocking Buffer (Grace Bio-Labs, Inc.) containing 10% *E. coli* lysate
591 (GenScript, Piscataway, NJ) and incubated for 30 minutes on a shaker at RT. Meanwhile,
592 microarray slides were rehydrated using 0.05X Super G Blocking buffer at RT. Rehydration
593 buffer was subsequently removed and samples added onto the slides. Arrays were incubated
594 overnight at 4°C on a shaker (180 rpm). Serum samples were removed the following day and
595 microarrays were washed using 1X TBST buffer (Grace Bio-Labs, Inc.). Secondary antibod-
596 ies were then applied at a dilution of 1:200 and incubated for two hours at RT on the shaker,
597 followed by another washing step and a one-hour incubation in a 1:250 dilution of Qdot™585
598 Streptavidin Conjugate. After a final washing step, slides were dried by centrifugation at 500
599 g for 10 minutes. Slide images were taken using the ArrayCAM® Imaging System (Grace
600 Bio-Labs) and the ArrayCAM 400-S Microarray Imager Software.

601 Microarray data were analyzed in R statistical software package version 3.6.2. All images
602 were manually checked for any noise signal. Each antigen spot signal was corrected for local
603 background reactivity by applying a normal-exponential convolution model (McGee & Chen,
604 2006) using the RMA-75 algorithm for parameter estimation (available in the LIMMA package
605 v3.28.14) (Silver *et al*, 2009). Data was log2-transformed and further normalized by subtrac-
606 tion of the median signal intensity of mock expression spots on the particular array to correct
607 for background activity of antibodies binding to *E. coli* lysate. After log2 transformation data
608 approached normal distribution. Differential antibody levels (protein array signal) in the differ-
609 ent patient groups were determined by Welch-corrected Student's t-test. Antigens with
610 $p<0.05$ and a fold change >2 of mean signal intensities were defined as differentially recog-
611 nized between the tested sample groups. Volcano plots were generated using the PAA
612 package (Turewicz *et al*, 2016) and GraphPad Prism 8. Individual antibody breadths were
613 defined as number of seropositive features with signal intensities exceeding an antigen-
614 specific threshold set at six standard deviations above the mean intensity in negative control
615 samples.
616
617 *Unsupervised random forest model*
618 An unsupervised random forest (RF) model, a machine learning method based on multiple
619 classification and regression trees, was calculated to estimate proximity between patients.
620 Variable importance was calculated, which shows the decrease in prediction accuracy if val-

ues of a variable are permuted randomly. The *k*-medoids clustering method was applied on the proximity matrix to group patients according to their serological profile. Input data for random forest were Luminex measurements for MSP1, AMA1 and CSP reduced by principal component analysis (PCA; first principal component selected), mADRB, ELISA, and antibody breadth of IgG and IgM determined by protein microarray were used to fit the RF model. Multidimensional scaling was used to display patient cluster. All analyses were done with R (4.02) using the packages randomForest (4.6-14) to run RF models and cluster (2.1.0) for k-medoids clustering.

Patient classification according to severity

Severity was defined in line with the WHO criteria for severe malaria in adults (World Health Organization (WHO), 2014). Patients were considered as having severe malaria if they showed signs of impaired organ function (e.g., jaundice, renal failure, cerebral manifestations) or had extremely high parasitemia (>10%). In addition, patients #1 and #26 were included into the severe group due to circulating schizonts indicative of a very high sequestering parasite biomass associated with severity (Bernabeu *et al*, 2016) (Table S2).

DNA purification and MSP1 genotyping

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

RNA extraction, RNA-seq library preparation, and sequencing

TRIzol samples were thawed, mixed rigorously with 0.2 volumes of cold chloroform and incubated for 3 min at room temperature. After centrifugation for 30 min at 4°C and maximum speed, the supernatant was carefully transferred to a new tube and mixed with an equal volume of 70% ethanol. Afterwards the manufacturer's instruction from the RNeasy MinElute Kit (Qiagen) were followed with DNase digestion (DNase I, Qiagen) for 30 min on column. Elution of the RNA was carried out in 14 µl. Human globin mRNA was depleted from all samples except from samples #1 and #2 using the GLOBINclear kit (ThermoFisher Scientific). The quality of the RNA was assessed using the Agilent 6000 Pico kit with the Bioanalyzer 2100 (Agilent) (Figure S5), the RNA quantity using the Qubit RNA HA assay kit and a Qubit 3.0 fluorometer (ThermoFisher Scientific). Upon arrival at BGI Genomics Co. (Hong Kong), the RNA quality of each sample was double-checked before sequencing. The median RIN value over all *ex vivo* samples was 6.75 (IQR: 5.93–7.40) (Figure S8), although this measurement has only limited significance for samples containing RNA of two species. Customized library construction in accordance to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) including amplification

with KAPA polymerase and HiSeq 2500 100 bp paired-end sequencing was also performed by BGI Genomics Co. (Hong Kong).

RNA-seq read mapping and data analysis

Differential expression of core genes

Differential gene expression analysis of *P. falciparum* core genes was done in accordance with Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) using the scripts available 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).

Var *gene assembly*

Samples from patient #1 and #2 not subjected to globin-mRNA depletion due to their low RNA content after multiple rounds of DNase treatment showed low percentages of *P. falciparum* specific reads (12.4% and 15.68%) (Table S2). Consequently, less than one million *P. falciparum* reads were obtained for each of these samples and they were omitted from assembly due to low coverage, but included in the differential gene expression analysis.

Var genes were assembled using the pipeline described in Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018). The separate assembly approach was chosen since it reduces the risk for generating false chimeric genes and results in longer contigs compared to the combined all sample assembly approach. Briefly, non-*var* reads were first filtered out by removing reads that aligned to *H. sapiens*, *P. vivax* or non-*var P. falciparum*. Assembly of the remaining reads was then performed using a pipeline combining SOAPdenovo-Trans and Cap3 (Xie *et al*, 2014; Huang & Madan, 1999; Liao *et al*, 2013). Finally, contaminants were removed from the resulting contigs and they were then translated into the correct reading frame. Reads were mapped to the contigs using BWA-MEM (Li, 2013) and RPKM values were calculated for each *var* transcript to compare individual transcript levels in each patient. Although transcripts might be differentially covered by RNA-seq due to their variable GC content, this seems not to be an issue between *var* genes (Tonkin-Hill *et al*, 2018).

Var *transcript differential expression*

Expression for the assembled *var* genes was quantified using Salmon v0.14.1 (Patro *et al*, 2017) for 531 transcripts with five read counts in at least 3 patient isolates. Both the naïve and pre-exposed groups as well as the severe and non-severe groups were compared. The combined set of all *de novo* assembled transcripts was used as a reference. As the RNA-seq reads from each sample were assembled independently it is possible for a highly similar transcript to be present multiple times in the combined set of transcripts from all samples. The Salmon algorithm identifies equivalence sets between transcripts allowing a single read to support the expression of multiple transcripts. As a result, Salmon accounts for the redundancy present in our whole set of *var* gene contigs from all separate sample-specific assemblies. To confirm the suitability of this approach we also ran the Corset algorithm as used in Tonkin-Hill *et al.*, (Tonkin-Hill *et al*, 2018; Davidson & Oshlack, 2014). Unlike Salmon which attempts to quantify the expression of transcripts themselves, Corset copes with the redundancy present in *de novo* transcriptome assemblies by clustering similar transcripts together using both the sequence identity of the transcripts as well as multi-mapping read alignments. Of the transcripts identified using the Salmon analysis 5/15 in the naïve versus pre-exposed and 4/13 in the severe versus non-severe were identified in the significant clusters produced using Corset. As the two algorithms take very different approaches and as Salmon is quanti-

732 fying transcripts rather than the 'gene' like clusters of Corset this represents a fairly reasona-
733 ble level of concordance between the two methods. However, due to the high diversity in *var*
734 genes, both of these approaches are only able to identify significant associations between
735 transcripts and phenotypes when there is sufficient similarity within the associated sequenc-
736 es. In both the Salmon and Corset pipelines differential expression analysis of the resulting
737 *var* expression values was performed using DESeq2 v1.26 (Love *et al*, 2014). The
738 Benjamini-Hochberg method was used to control for multiple testing (Benjamini & Hochberg,
739 1995).

740 To check differential expression of the conserved *var* gene variants *var1*-3D7, *var1*-IT and
741 *var2csa* raw reads were mapped with BWA-MEM (AS score >110) to the reference genes
742 from the 3D7 and the IT strains. The mapped raw read counts (bam files) were normalized
743 with the number of 3D7 mappable reads in each isolate using bamCoverage by introducing a
744 scaling factor to generate bigwig files displayed in Artemis (Carver *et al*, 2012).
745
746 Var *domain and homology block differential expression*
747 Differential expression analysis at the domain and homology level was performed using a
748 similar approach to that described previously (Tonkin-Hill *et al*, 2018). Initially, the domain
749 families and homology blocks defined in Rask *et al.* were annotated to the assembled tran-
750 scripts using HMMER v3.1b2 (Rask *et al*, 2010; Eddy, 2011). Domains and homology blocks
751 previously identified to be significantly associated with severe disease in Tonkin-Hill *et al.,*
752 2018 were also annotated by single pairwise comparison in the assembled transcripts using
753 USEARCH v11.0.667 (Tonkin-Hill *et al*, 2018; Edgar, 2010). Overall, 336 contigs (5.22% of
754 all contigs >500 bp) possess partial domains in an unusual order, e.g., an NTS in an internal
755 region or a tandem arrangement of two DBLα or CIDRα domains. This might be caused by
756 *de novo* assembly errors, which is challenging from transcriptome data. Therefore, in both
757 cases the domain or homology block with the most significant alignment was taken as the
758 best annotation for each region of the assembled *var* transcripts (E-value cutoff of $1e^{-8}$). The
759 expression at each of these annotations was then quantified using featureCounts v1.6.4 be-
760 fore the counts were aggregated to give a total for each domain and homology block family in
761 each sample. Finally, similar to the transcript level analysis, DESeq2 was used to test for
762 differences in expression levels of both domain and homology block families in the naïve
763 versus pre-exposed groups as well as the severe versus non-severe groups. Again, more
764 than five read counts in at least three patient isolates were required for inclusion into differen-
765 tial expression analysis.
766
767 ***DBLα-tag sequencing***

For DBLα-tag PCR the forward primer varF_dg2 (5'-tcgtcggcagcgtcagatgtgtataagagacagGCAMGMAGTTTYGCNGATATWGG-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 AMPure 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

805 window approach and a *Phred* quality score above 20 to ensure high sequence quality out-
806 put. The trimmed sequences were then paired and converted, following analysis using the
807 Varia tool for quantification and prediction of the domain composition of the full-length *var*
808 sequences from which the DBLα-tag originated (Mackenzie *et al*, 2020). In brief, Varia clus-
809 ters DBLα-tags with 99% sequence identity using *Vsearch* program (v2.14.2), and each
810 unique tag is used to search a database consisting of roughly 235,000 annotated *var* genes
811 for near identical *var* sequences (95% identity over 200 nucleotides). The domain composi-
812 tion of all "hit" sequences is checked for conflicting annotations and the most likely domain
813 composition is retuned. The tool validation indicated prediction of correct domain composi-
814 tions for around 85% of randomly selected *var* tags, with higher hit rate and accuracy of the
815 N-terminal domains. An average of 2,223.70 reads per patient sample was obtained and
816 clusters consisting of less than 10 reads were excluded from the analysis. The raw Varia
817 output file is given in Table S10. The proportion of transcripts encoding a given *Pf*EMP1 do-
818 main type or subtype was calculated for each patient. These expression levels were used to
819 first test the hypothesis that N-terminal domain types associated with EPCR are found more
820 frequently in first-time infections or upon severity of disease, while those associated with
821 non-EPCR binding were associated with pre-exposed or mild cases. Secondly, quantile re-
822 gression was used to calculate median differences (with 95%-confidence intervals) in ex-
823 pression levels for all main domain classes and subtypes between severity and exposure
824 groups. All analyses were done with R (4.02) using the package quantreg (5.73) for quantile
825 regression.

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

835 Calculation of the proportion of RNA-seq data covered by DBLα-tag was done with the upper
836 75th percentile based on total RPKM values determined for each patient. Vice versa, only
837 DBLα-tag clusters with more than 10 reads were considered and percent coverage of reads
838 and clusters calculated for each individual patient.

839 For all samples the agreement between the two molecular methods DBLα-tag sequencing
840 and RNA-seq was analyzed with a Bland-Altman plot, each individually and summarized.
841 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 who provided an extra blood sample for our research purposes. We would also like to thank the staff of the I. Medical Department at the UKE and of the Bundeswehrkrankenhaus for identifying patients for the study and specifically Maria Mackroth, Julian Schulze zur Wiesch, Johannes Jochum and Thierry Rolling for assisting in the recruitment of patients. 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.

This work was supported by funding from DFG BA5213/3-1 (JSW, AB), the TTU Malaria of the DZIF network (RK, ET, RF, AB), the Partnership of Universität Hamburg and DESY (PIER) project ID PIF-2018-87 (JS, TG), the State Graduate Funding Program Scholarship (HmbNFG) of the University of Hamburg (JAMS), the National Health Medical Research Council of Australia (MD), the Wellcome Trust grant 104111/Z/14/ZR (TO), the KFJ FONDEN, Læge Sofus Carl Emil Friis og hustru Olga Doris Friis' Legat, Lundbeck Foundation (R344-2020-934) and the Danish Council for Independent Research (LT, RWJ, LT).

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, Deloron P & Houzé S (2017) Influence of host factors and parasite biomass on the severity of imported Plasmodium falciparum malaria. *PLoS One* 12: e0175328

Avril M, Brazier AJ, Melcher M, Sampath S & Smith JD (2013) DC8 and DC13 var Genes Associated with Severe Malaria Bind Avidly to Diverse Endothelial Cells. *PLoS Pathog* 9

Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL, Sullivan DJ, Bull PC, Stins MF & Smith JD (2012) A restricted subset of var genes mediates adherence of Plasmodium falciparum-infected erythrocytes to brain endothelial cells. *Proc Natl Acad Sci U S A* 109: E1782–E1790

879 Bachmann A, Bruske E, Krumkamp R, Turner L, Wichers JS, Petter M, Held J, Duffy MF,
880 Sim BKL, Hoffman SL, *et al* (2019) Controlled human malaria infection with Plasmodium
881 falciparum demonstrates impact of naturally acquired immunity on virulence gene ex-
882 pression. *PLOS Pathog* 15: e1007906
883 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and
884 Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 57: 289–300
885 Berger SS, Turner L, Wang CW, Petersen JEV, Kraft M, Lusingu JPA, Mmbando B,
886 Marquard AM, Bengtsson DBAC, Hviid L, *et al* (2013) Plasmodium falciparum Express-
887 ing Domain Cassette 5 Type PfEMP1 (DC5-PfEMP1) Bind PECAM1. *PLoS One* 8:
888 69117
889 Bernabeu M, Danziger SA, Avril M, Vaz M, Babar PH, Brazier AJ, Herricks T, Maki JN,
890 Pereira L, Mascarenhas A, *et al* (2016) Severe adult malaria is associated with specific
891 PfEMP1 adhesion types and high parasite biomass. *Proc Natl Acad Sci U S A* 113:
892 E3270–E3279
893 Borrmann S (2020) Mapping of safe and early chemo-attenuated live Plasmodium falciparum
894 immunization identifies immune signature of vaccine efficacy. *bioRxiv*:
895 2020.09.14.296152
896 Carlson J, Helmby H, Wahlgren M, Carlson J, Helmby H, Wahlgren M, Hill A, Brewster D &
897 Greenwood BM (1990) Human cerebral malaria: association with erythrocyte rosetting
898 and lack of anti-rosetting antibodies. *Lancet* 336: 1457–1460
899 Carver T, Harris SR, Berriman M, Parkhill J & McQuillan JA (2012) Artemis: An integrated
900 platform for visualization and analysis of high-throughput sequence-based experimental
901 data. *Bioinformatics* 28: 464–469
902 Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, Lavstsen T, Hviid L,
903 Duffy PE & Theander TG (2010) Hierarchical, domain type-specific acquisition of anti-
904 bodies to Plasmodium falciparum erythrocyte membrane protein 1 in Tanzanian chil-
905 dren. *Infect Immun* 78: 4653–4659
906 Cham GKK, Turner L, Lusingu J, Vestergaard L, Mmbando BP, Kurtis JD, Jensen ATR,
907 Salanti A, Lavstsen T & Theander TG (2009) Sequential, Ordered Acquisition of Anti-
908 bodies to Plasmodium falciparum Erythrocyte Membrane Protein 1 Domains . *J Immu-*
909 *nol* 183: 3356–3363
910 Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, Bull PC, Mok S,
911 Gupta AP, Wang CW, *et al* (2012) A subset of group A-like var genes encodes the ma-
912 laria parasite ligands for binding to human brain endothelial cells. *Proc Natl Acad Sci U*
913 *S A* 109: E1772
914 Dahlbäck M, Lavstsen T, Salanti A, Hviid L, Arnot DE, Theander TG & Nielsen MA (2007)
915 Changes in var gene mRNA levels during erythrocytic development in two phenotypi-

cally distinct Plasmodium falciparum parasites. *Malar J* 6: 78

Davidson NM & Oshlack A (2014) Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol* 15: 410

Dent AE, Malhotra I, Wang X, Babineau D, Yeo KT, Anderson T, Kimmel RJ, Angov E, Lanar DE, Narum D, *et al* (2016) Contrasting Patterns of Serologic and Functional Antibody Dynamics to Plasmodium falciparum Antigens in a Kenyan Birth Cohort. *Clin Vaccine Immunol* 23: 104–116

Dondorp AM, Lee SJ, Faiz MA, Mishra S, Price R, Tjitra E, Than M, Htut Y, Mohanty S, Yunus E Bin, *et al* (2008) The relationship between age and the manifestations of and mortality associated with severe malaria. *Clin Infect Dis* 47: 151–157

Doolan DL, Mu Y, Unal B, Sundaresh S, Hirst S, Valdez C, Randall A, Molina D, Liang X, Freilich DA, *et al* (2008) Profiling humoral immune responses to P. falciparum infection with protein microarrays. *Proteomics* 8: 4680–4694

Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya WMM, *et al* (2005) Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc Natl Acad Sci U S A* 102: 5108–5113

Duffy F, Bernabeu M, Babar PH, Kessler A, Wang CW, Vaz M, Chery L, Mandala WL, Rogerson SJ, Taylor TE, *et al* (2019) Meta-analysis of Plasmodium falciparum var Signatures Contributing to Severe Malaria in African Children and Indian Adults. *MBio* 10: e00217-19

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, Noviyanti R, Tsuboi T, Feng ZP, Trianty L, Sebayang BF, Takashima E, Sumardy F, Lampah DA, Turner L, *et al* (2016) Differences in PfEMP1s recognized by antibodies from patients with uncomplicated or severe malaria. *Malar J* 15

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

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.

953 *Sci Rep* 3
954 Gagnon-Bartsch JA & Speed TP (2012) Using control genes to correct for unwanted varia-
955 tion in microarray data. *Biostatistics* 13: 539–552
956 Ghumra A, Semblat J-P, Ataide R, Kifude C, Adams Y, Claessens A, Anong DN, Bull PC,
957 Fennell C, Arman M, *et al* (2012) Induction of Strain-Transcending Antibodies Against
958 Group A PfEMP1 Surface Antigens from Virulent Malaria Parasites. *PLoS Pathog* 8:
959 e1002665
960 Gustavsen JA, Pai S, Isserlin R, Demchak B & Pico AR (2019) RCy3: Network biology using
961 Cytoscape from within R. *F1000Research* 8: 1774
962 Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, Mayanja-Kizza H,
963 Ssewanyana I, Kamya MR, Beeson JG, *et al* (2015) Novel serologic biomarkers provide
964 accurate estimates of recent Plasmodium falciparum exposure for individuals and com-
965 munities. *Proc Natl Acad Sci U S A* 112: E4438–E4447
966 Higgins MK & Carrington M (2014) Sequence variation and structural conservation allows
967 development of novel function and immune evasion in parasite surface protein families.
968 *Protein Sci* 23: 354–365
969 Van Den Hoogen LL, Walk J, Oulton T, Reuling IJ, Reiling L, Beeson JG, Coppel RL, Singh
970 SK, Draper SJ, Bousema T, *et al* (2019) Antibody responses to antigenic targets of re-
971 cent exposure are associated with low-density parasitemia in controlled human plasmo-
972 dium falciparuminfections. *Front Microbiol* 10
973 Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:
974 868–877
975 Janes JH, Wang CP, Levin-Edens E, Vigan-Womas I, Guillotte M, Melcher M, Mercereau-
976 Puijalon O & Smith JD (2011) Investigating the host binding signature on the Plasmo-
977 dium falciparum PfEMP1 protein family. *PLoS Pathog* 7
978 Jespersen JS, Wang CW, Mkumbaye SI, Minja DT, Petersen B, Turner L, Petersen JE, Lus-
979 ingu JP, Theander TG & Lavstsen T (2016) Plasmodium falciparum var genes ex-
980 pressed in children with severe malaria encode CIDR α1 domains. *EMBO Mol Med* 8:
981 839–850
982 Kamaliddin C, Rombaut D, Guillochon E, Royo J, Ezinmegnon S, Agbota G, Huguet S,
983 Guemouri S, Peirera C, Coppée R, *et al* (2019) From genomic to LC-MS/MS evidence:
984 Analysis of PfEMP1 in Benin malaria cases. *PLoS One* 14: e0218012
985 Kapelski S, Klockenbring T, Fischer R, Barth S & Fendel R (2014) Assessment of the neu-
986 trophilic antibody-dependent respiratory burst (ADRB) response to Plasmodium falcipa-
987 rum . *J Leukoc Biol* 96: 1131–1142
988 Kessler A, Dankwa S, Bernabeu M, Harawa V, Danziger SA, Duffy F, Kampondeni SD,
989 Potchen MJ, Dambrauskas N, Vigdorovich V, *et al* (2017) Linking EPCR-Binding

PfEMP1 to Brain Swelling in Pediatric Cerebral Malaria. *Cell Host Microbe* 22: 601-614.e5

Kolberg L, Raudvere U, Kuzmin I, Vilo J & Peterson H (2020) gprofiler2 -- an R package for 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

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: E1791-800

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

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

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

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-

1027 Mury C, Zhao K & Su X (2011) Directional gene expression and antisense transcripts in
1028 sexual and asexual stages of Plasmodium falciparum. *BMC Genomics* 12: 587
1029 Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion
1030 for RNA-seq data with DESeq2. *Genome Biol* 15: 550
1031 Mackenzie G, Jensen RW, Lavstsen T & Otto TD (2020) Varia: Prediction, analysis and
1032 visualisation of variable genes. *bioRxiv*: 2020.12.15.422815
1033 Magallón-Tejada A, Machevo S, Cisteró P, Lavstsen T, Aide P, Rubio M, Jiménez A, Turner
1034 L, Valmaseda A, Gupta H, *et al* (2016) Cytoadhesion to gC1qR through Plasmodium
1035 falciparum Erythrocyte Membrane Protein 1 in Severe Malaria. *PLOS Pathog* 12:
1036 e1006011
1037 McGee M & Chen Z (2006) Parameter estimation for the exponential-normal convolution
1038 model for background correction of affymetrix GeneChip data. *Stat Appl Genet Mol Biol*
1039 5
1040 McQuaid F & Rowe JA (2020) Rosetting revisited: A critical look at the evidence for host
1041 erythrocyte receptors in Plasmodium falciparum rosetting. *Parasitology* 147: 1–11
1042 doi:10.1017/S0031182019001288 [PREPRINT]
1043 Merico D, Isserlin R, Stueker O, Emili A & Bader GD (2010) Enrichment Map: A Network-
1044 Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* 5:
1045 e13984
1046 Mkumbaye SI, Wang CW, Lyimo E, Jespersen JS, Manjurano A, Mosha J, Kavishe RA,
1047 Mwakalinga SB, Minja DTR, Lusingu JP, *et al* (2017) The severity of *Plasmodium falci-*
1048 *parum* infection is associated with transcript levels of *var* genes encoding EPCR-binding
1049 PfEMP1. *Infect Immun*: IAI.00841-16
1050 Nag S, Dalgaard MD, Kofoed PE, Ursing J, Crespo M, Andersen LOB, Aarestrup FM, Lund
1051 O & Alifrangis M (2017) High throughput resistance profiling of Plasmodium falciparum
1052 infections based on custom dual indexing and Illumina next generation sequencing-
1053 technology. *Sci Rep* 7: 1–13
1054 Obeng-Adjei N, Larremore DB, Turner L, Ongoiba A, Li S, Doumbo S, Yazew TB, Kayentao
1055 K, Miller LH, Traore B, *et al* (2020) Longitudinal analysis of naturally acquired PfEMP1
1056 CIDR domain variant antibodies identifies associations with malaria protection. *JCI In-*
1057 *sight*
1058 Obiero JM, Campo JJ, Scholzen A, Randall A, Bijker EM, Roestenberg M, Hermsen CC,
1059 Teng A, Jain A, Davies DH, *et al* (2019) Antibody Biomarkers Associated with Sterile
1060 Protection Induced by Controlled Human Malaria Infection under Chloroquine Prophy-
1061 laxis. *mSphere* 4
1062 Otto TD (2019) ThomasDOtto/varDB: First release of varDB.
1063 Otto TD, Assefa SA, Böhme U, Sanders MJ, Kwiatkowski D, Berriman M, Newbold C &

1064 Newbold C (2019) Evolutionary analysis of the most polymorphic gene family in falcipa-
1065 rum malaria. *Wellcome Open Res* 4: 193
1066 Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, Bull PC, Pearson RD, Abdi A,
1067 Dimonte S, *et al* (2018a) Long read assemblies of geographically dispersed Plasmo-
1068 dium falciparum isolates reveal highly structured subtelomeres. *Wellcome open Res* 3:
1069 52
1070 Otto TD, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, Oyola SO, Okouga AP,
1071 Boundenga L, Willaume E, *et al* (2018b) Genomes of all known members of a Plasmo-
1072 dium subgenus reveal paths to virulent human malaria. *Nat Microbiol* 3: 687–697
1073 Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C (2017) Salmon provides fast and bias-
1074 aware quantification of transcript expression. *Nat Methods* 14: 417–419
1075 Rask TS, Hansen DA, Theander TG, Gorm Pedersen A & Lavstsen T (2010) Plasmodium
1076 falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and
1077 Conquer. *PLoS Comput Biol* 6: e1000933
1078 Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H & Vilo J (2019) G:Profiler: A
1079 web server for functional enrichment analysis and conversions of gene lists (2019 up-
1080 date). *Nucleic Acids Res* 47: W191–W198
1081 Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, Sarthou JL & Mercereau-
1082 Puijalon O (1996) Extensive genetic diversity of Plasmodium falciparum isolates col-
1083 lected from patients with severe malaria in Dakar, Senegal. *Trans R Soc Trop Med Hyg*
1084 90: 704–711
1085 Rowe JA, Claessens A, Corrigan RA & Arman M (2009) Adhesion of Plasmodium falcipa-
1086 rum-infected erythrocytes to human cells: Molecular mechanisms and therapeutic impli-
1087 cations. *Expert Rev Mol Med* 11: e16 doi:10.1016/0925-4005(95)85135-6
1088 Rowe JA, Obiero J, Marsh K & Raza A (2002) Short report: Positive correlation between ro-
1089 setting and parasitemia in Plasmodium falciparum clinical isolates. *Am J Trop Med Hyg*
1090 66: 458–460
1091 Salanti A, Dahlbäck M, Turner L, Nielsen MA, Barfod L, Magistrado P, Jensen ATR, Lavst-
1092 sen T, Ofori MF, Marsh K, *et al* (2004) Evidence for the involvement of VAR2CSA in
1093 pregnancy-associated malaria. *J Exp Med* 200: 1197–1203
1094 Saul A (1999) The role of variant surface antigens on malaria-infected red blood cells. *Para-*
1095 *sitol Today* 15: 455–7
1096 Schwartz E, Sadetzki S, Murad H & Raveh D (2001) Age as a risk factor for severe Plasmo-
1097 dium falciparum malaria in nonimmune patients. *Clin Infect Dis* 33: 1774–1777
1098 Shabani E, Hanisch B, Opoka RO, Lavstsen T & John CC (2017) Plasmodium falciparum
1099 EPCR-binding PfEMP1 expression increases with malaria disease severity and is ele-
1100 vated in retinopathy negative cerebral malaria. *BMC Med* 15: 183

1101 Silver JD, Ritchie ME & Smyth GK (2009) Microarray background correction: Maximum likeli-
1102 hood estimation for the normal-exponential convolution. *Biostatistics* 10: 352–363
1103 Smyth GK (2005) limma: Linear Models for Microarray Data. In *Bioinformatics and Computa-*
1104 *tional Biology Solutions Using R and Bioconductor* pp 397–420. Springer-Verlag
1105 Storm J, Jespersen JS, Seydel KB, Szestak T, Mbewe M, Chisala N V, Phula P, Wang CW,
1106 Taylor TE, Moxon CA, *et al* (2019) Cerebral malaria is associated with differential cy-
1107 toadherence to brain endothelial cells. *EMBO Mol Med* 11
1108 Subudhi AK, Boopathi PA, Pandey I, Kohli R, Karwa R, Middha S, Acharya J, Kochar SK,
1109 Kochar DK & Das A (2015) Plasmodium falciparum complicated malaria: Modulation
1110 and connectivity between exportome and variant surface antigen gene families. *Mol*
1111 *Biochem Parasitol* 201: 31–46
1112 Taghavian O, Jain A, Joyner CJ, Ketchum S, Nakajima R, Jasinskas A, Liang L, Fong R,
1113 King C, Greenhouse B, *et al* (2018) Antibody Profiling by Proteome Microarray with Mul-
1114 tiplex Isotype Detection Reveals Overlap between Human and Aotus nancymaae Con-
1115 trolled Malaria Infections. *Proteomics* 18
1116 Tenenbaum D (2020) KEGGREST: Client-side REST access to KEGG. R package version
1117 1.30.0. [PREPRINT]
1118 Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, Marfurt J,
1119 Cobbold SA, Rambhatla JS, McConville MJ, *et al* (2018) The Plasmodium falciparum
1120 transcriptome in severe malaria reveals altered expression of genes involved in impor-
1121 tant processes including surface antigen–encoding var genes. *PLoS Biol* 16: e2004328
1122 Turewicz M, Ahrens M, May C, Marcus K & Eisenacher M (2016) PAA: An R/bioconductor
1123 package for biomarker discovery with protein microarrays. *Bioinformatics* 32: 1577–
1124 1579
1125 Turner GDH, Morrison H, Jones M, Davis TME, Looareesuwan S, Buley ID, Gatter KC, New-
1126 bold CI, Pukritayakamee S, Nagachinta B, *et al* (1994) An immunohistochemical study
1127 of the pathology of fatal malaria: Evidence for widespread endothelial activation and a
1128 potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J*
1129 *Pathol* 145: 1057–1069
1130 Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JE V., Avril M, Brazier AJ, Freeth J,
1131 Jespersen JS, Nielsen MA, *et al* (2013) Severe malaria is associated with parasite bind-
1132 ing to endothelial protein C receptor. *Nature* 498: 502–505
1133 Turner L, Lavstsen T, Mmbando BP, Wang CW, Magistrado PA, Vestergaard LS, Ishengoma
1134 DS, Minja DTR, Lusingu JP & Theander TG (2015) IgG antibodies to endothelial protein
1135 C receptor-binding cysteine-rich interdomain region domains of Plasmodium falciparum
1136 erythrocyte membrane protein 1 are acquired early in life in individuals exposed to ma-
1137 laria. *Infect Immun* 83: 3096–3103

Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, Bouzek H, Moon W, Babak T, Fried M, *et al* (2011) NSR-seq transcriptional profiling enables identification of a gene signature of Plasmodium falciparum parasites infecting children. *J Clin Invest* 121: 1119–1129

Warimwe GM, Fegan G, Musyoki JN, Newton CRJC, Opiyo M, Githinji G, Andisi C, Menza F, Kitsao B, Marsh K, *et al* (2012) Prognostic indicators of life-threatening malaria are associated with distinct parasite variant antigen profiles. *Sci Transl Med* 4

Warimwe GM, Keane TM, Fegan G, Musyoki JN, Newton CRJC, Pain A, Berriman M, Marsh K & Bull PC (2009) Plasmodium falciparum var gene expression is modified by host immunity. *Proc Natl Acad Sci U S A* 106: 21801–21806

WHO (2019) WHO World Malaria Report 2019

Wickham H (2016) Ggplot2□: elegant graphics for data analysis Springer

Winter G, Chen Q, Flick K, Kremsner P, Fernandez V & Wahlgren M (2003) The 3D7var5.2 (varCOMMON) type var gene family is commonly expressed in non-placental Plasmodium falciparum malaria. *Mol Biochem Parasitol* 127: 179–191

World Health Organization (WHO) (2014) Severe malaria. *Trop Med Int Health* 19 Suppl 1: 7–131

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, *et al* (2014) SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30: 1660–1666

Zhao S, Guo Y SY (2020) KEGGprofile: An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway. R package version 1.32.0. [PREPRINT]

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**). Venn chart showing the patient subgroups used for differential expression analysis (**G**). 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. Patient #21 is shown as filled circle in grey with a cross, patient #26 is represented by an open circle with cross. ELISA: Enzyme-linked immunosorbent assay, IQR: interquartile range, PCA: Principal component analysis

Figure 2: Overview of the methodology and differential core gene expression analysis. Summary diagram of the approaches taken to analyze the differential expression of core and *var* genes. In principle, all samples were analyzed by sequencing of the RNA using next generation sequencing (NGS) and by sequencing of expressed sequences tags (EST) from the DBLα-domain (**A**). 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 (**B**). 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 $p<0.01$ (**C**).

Figure 3: 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 all *var* gene groups with their associated binding phenotypes and typical *Pf*EMP1 domain compositions. The N-terminal head structure confers mutually exclusive receptor binding phenotypes: EPCR (beige: CIDRα1.1/4-8), CD36 (turquoise CIDRα2-6),

CSA (yellow: VAR2CSA) and yet unknown phenotypes (brown: CIDRβ/γ/δ, dark 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 conferred by CIDRβ/γ/δ domains. 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 after the first CIDR domain responsible for ICAM-1- (DBLβ1/3/5) or gC1qr-binding (DBLβ12) (**A**). Transcripts, domains and homology blocks according to Rask *et al.* (Rask *et al*, 2010) as well as domain predictions from the DBLα-tag approach 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, PAM: pregnancy-associated malaria, TM: transmembrane domain

Figure 4: Expression differences between parasites from first-time infected and pre-exposed patients at the level of *var* gene transcripts, domains and homology blocks determined by NGS. 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 (**A, B)** as well as *Pf*EMP1 domain subfamilies (**C–F**) and homology blocks (**G, H**) from Rask *et al.* (Rask *et al*, 2010) with an adjusted p-value of <0.05. Data are displayed as heat maps showing expression levels either in log transformed normalized Salmon read counts (**A**) or in log transcript per million (TPM) (**C, G**) for each individual sample. Box plots show median log transformed normalized Salmon read counts (**B**) or TPM (**D, F, H**) and interquartile range (IQR) for each group of samples. Individual domains from inter-strain conserved tandem arrangements of domains, so called domain cassettes (DC), found significantly higher expressed in samples from first-time infected (blue arrow) and pre-exposed patients (grey arrow) are indicated in bold (**E**). 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 (**F**). Normalized Salmon read counts for all assembled transcripts and TPM for *Pf*EMP1 domains and homology blocks are available in Table S8.

Figure 5: Expression differences between parasites from severe and non-severe cases at the level of *var* gene transcripts, domains and homology blocks determined by NGS. 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 *var* gene contigs (A, B) as well as *Pf*EMP1 domain subfamilies (C–F) and homology blocks from Rask *et al.* (Rask *et al*, 2010) with an adjusted p-value of <0.05 in severe (red) and non-severe patient samples (grey) (**A**, **B**). Data are displayed as heat maps showing expression levels either in log transformed normalized Salmon read counts (**A**) or in log transcript per million (TPM) (**C, G**) for each individual sample. Box plots show median log transformed normalized Salmon read counts (**B**) or TPM (**D, F, H**) and interquartile range (IQR) for each group of samples. Individual domains from inter-strain conserved tandem arrangements of domains, so called domain cassettes (DC), found significantly higher expressed in severe (red arrow) and non-severe cases (grey arrow) are indicated in bold (**E**). 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 (**F**). Normalized Salmon read counts for all assembled transcripts and TPM for *Pf*EMP1 domains and homology blocks are available in Table S8.

Figure 6: 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.

Figure 7: 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 un-

known 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**).

Table 1: Patient groups data.

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: 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

Supplement figure 3: Summary diagram of the approaches taken to analyze the RNA-seq data. Diagram created in Lucidchart (www.lucidchart.com).

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: Differential expression of the *var1* variants 3D7 and IT 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 variants 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 6: 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 7: Quantile regression analysis of Varia outputs

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 (**A**) and severe patients (**B**).

Supplement figure 8: 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. Seroprevalence of head-structure CIDR domains determined by applying a cut off from Danish controls (mean + 2 STD) to the Luminex data.

Supplement table 2: Characteristics and classification of adult malaria patients. Parasitemia, signs of organ failure and sub-grouping of each individual patient.

Supplement table 3: 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 4: Differentially expressed genes excluding *var* genes (all gene analysis) between first-time infected and pre-exposed patient samples.

Supplement table 5: Differentially expressed genes excluding *var* genes (all gene analysis) between severe and non-severe patient samples.

Supplement table 6: 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 7: 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 calcu-

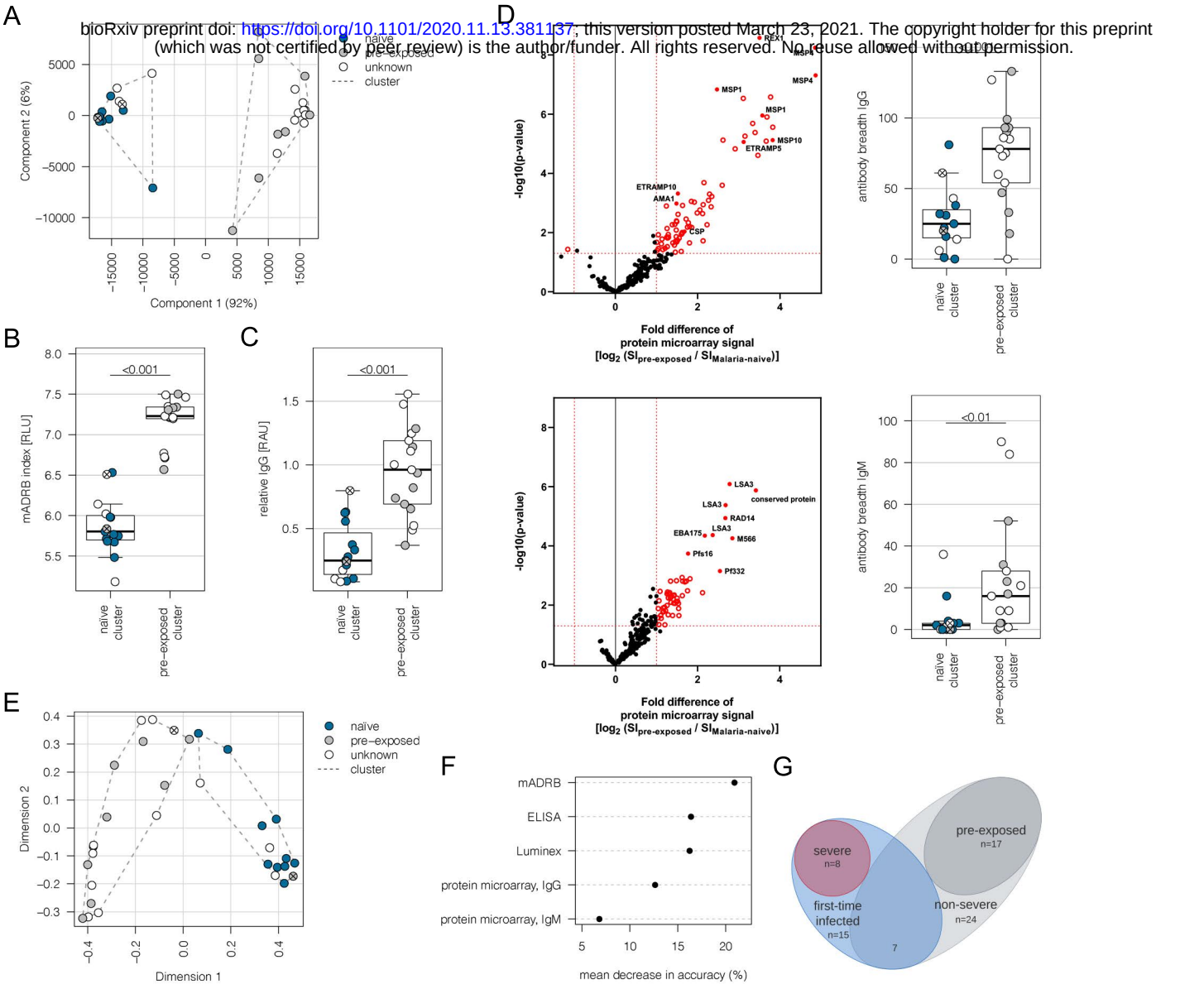
lated as number of mapped reads and normalized by the number of mapped reads against all transcript in each isolate, respectively. Therefore, RPKM expression values are only valid to compare within a single sample since RNA-seq reads were mapped only to the contigs of the respective patient isolate using 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 8: 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.

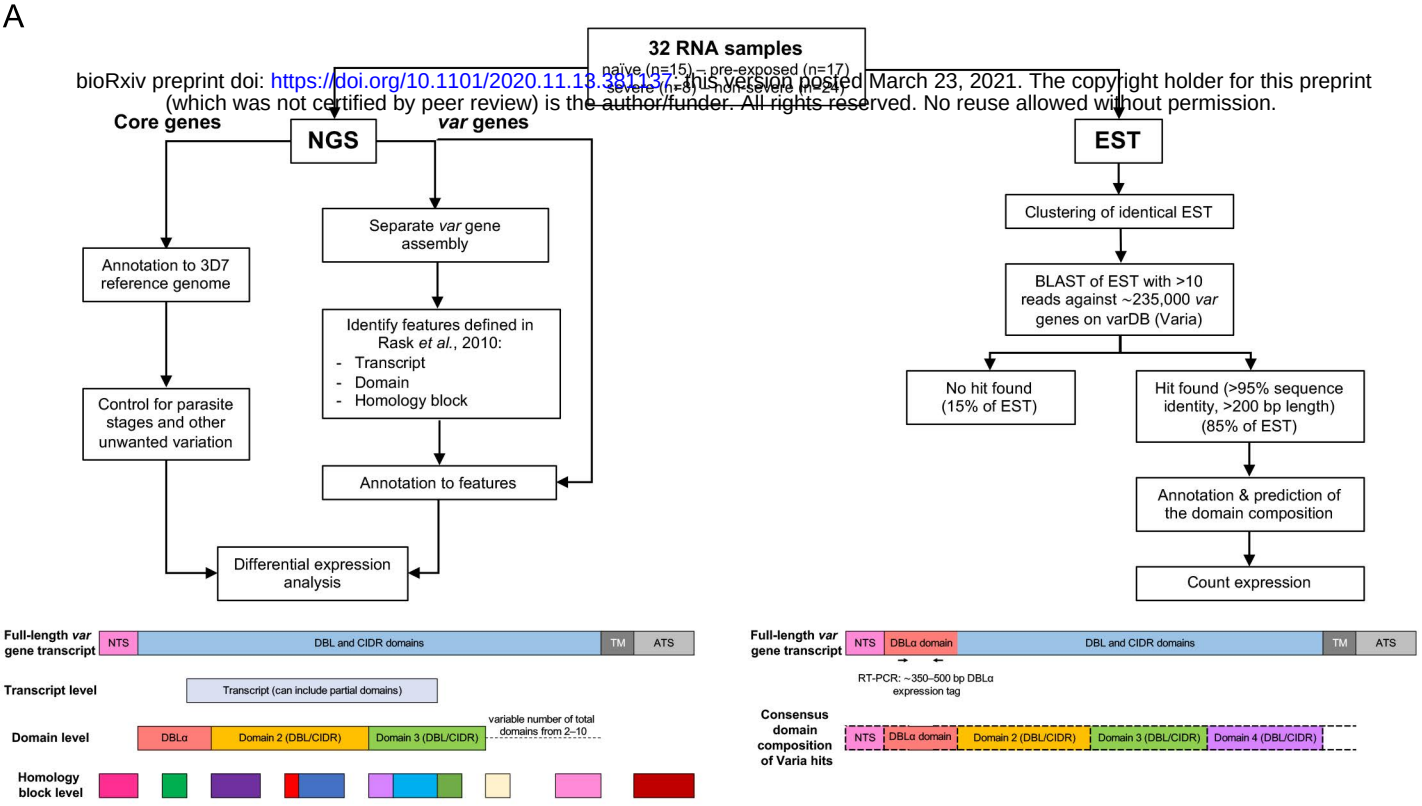
Supplement table 9: Differently expressed *var* transcripts, domains and homology blocks between first-time infected and pre-exposed patient samples.

Supplement table 10: Differently expressed *var* transcripts, domains and homology blocks between severe and non-severe patient samples.

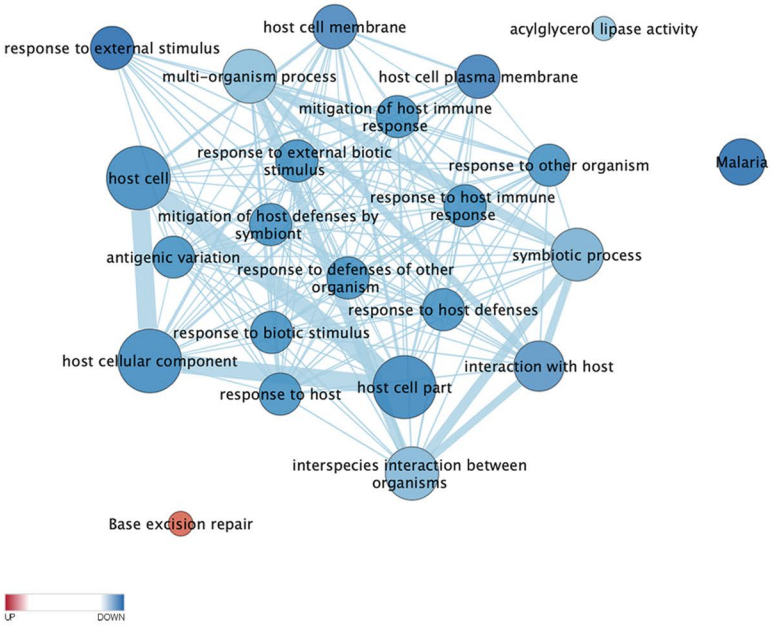
Supplement table 11: Data from DBLα-tag sequencing.



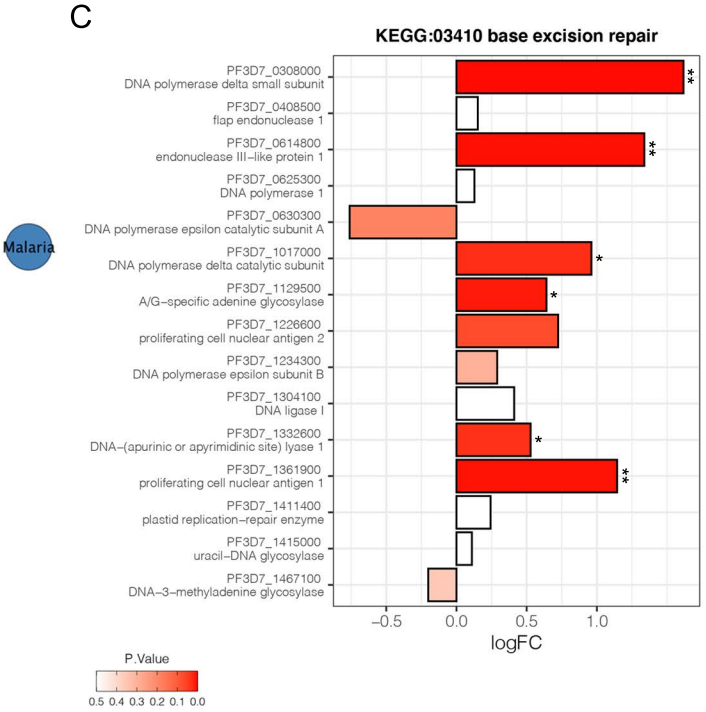
A



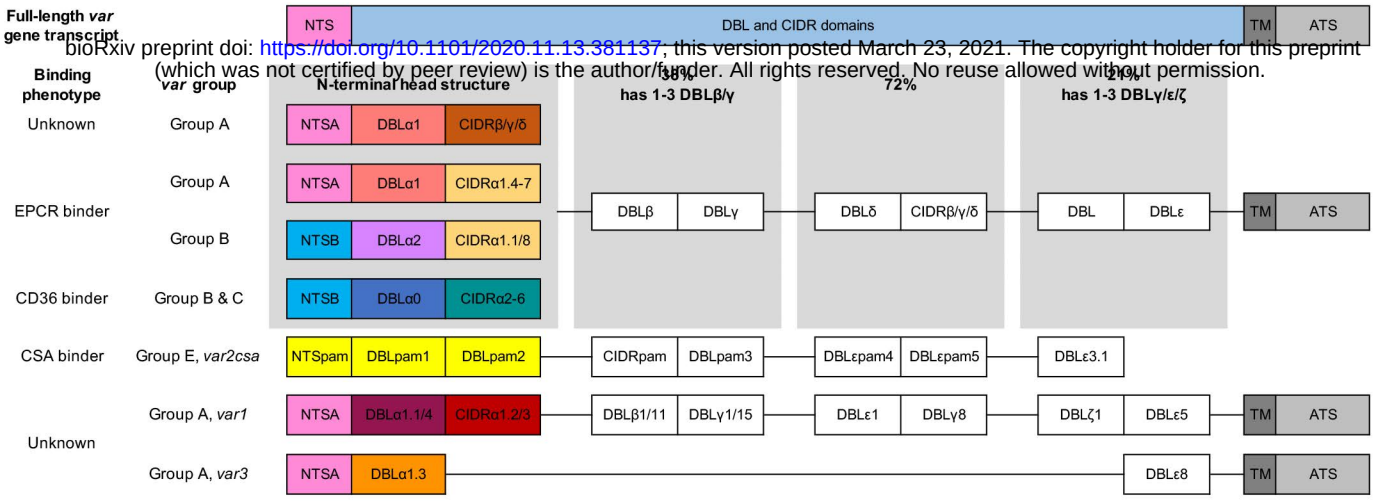
B



C

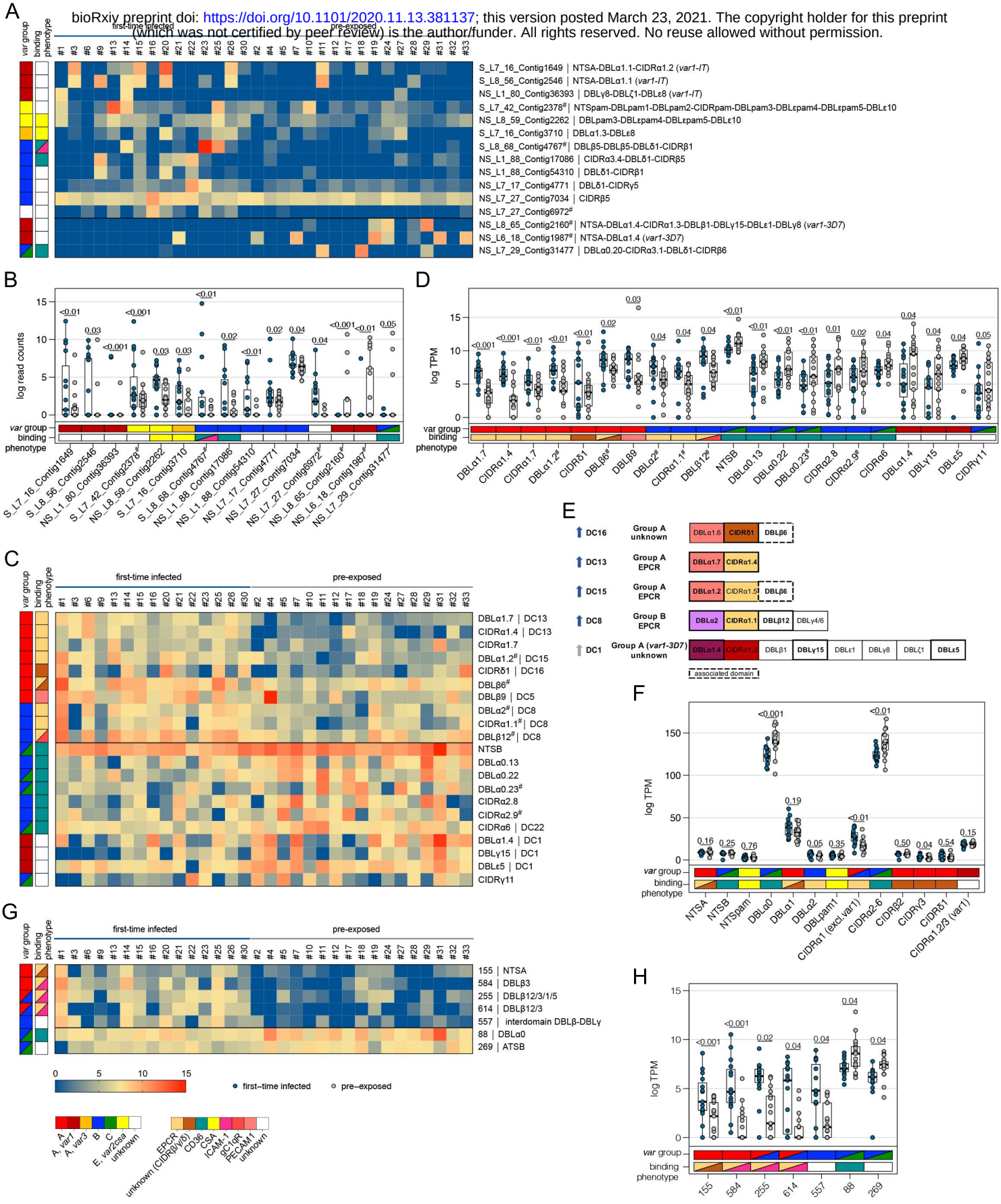


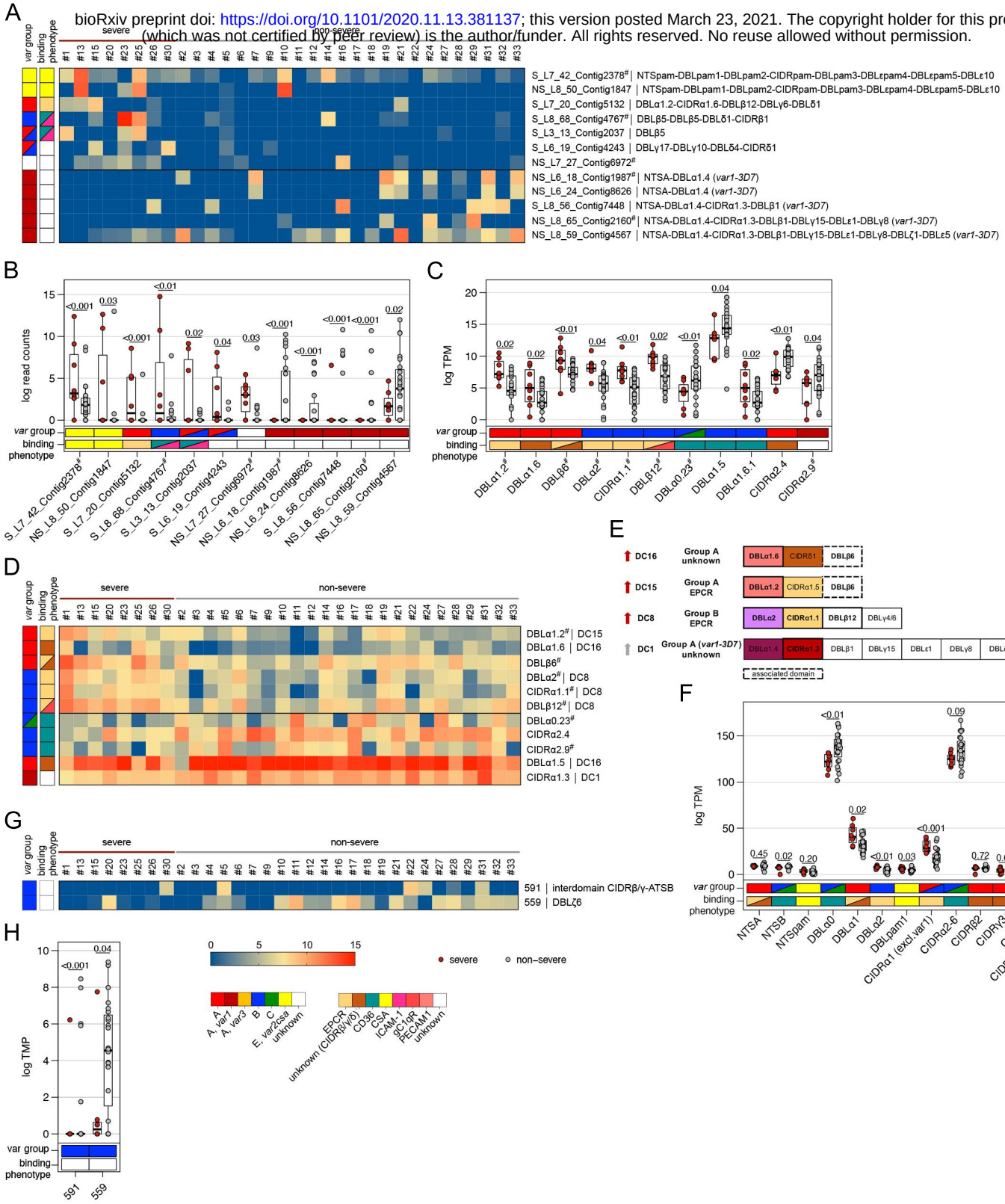
A



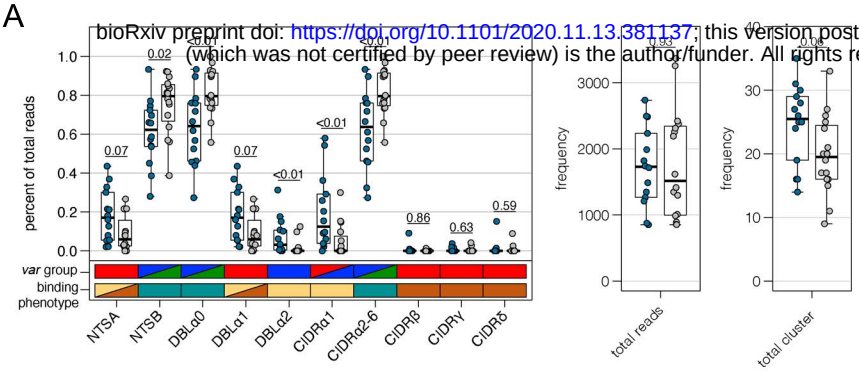
B

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

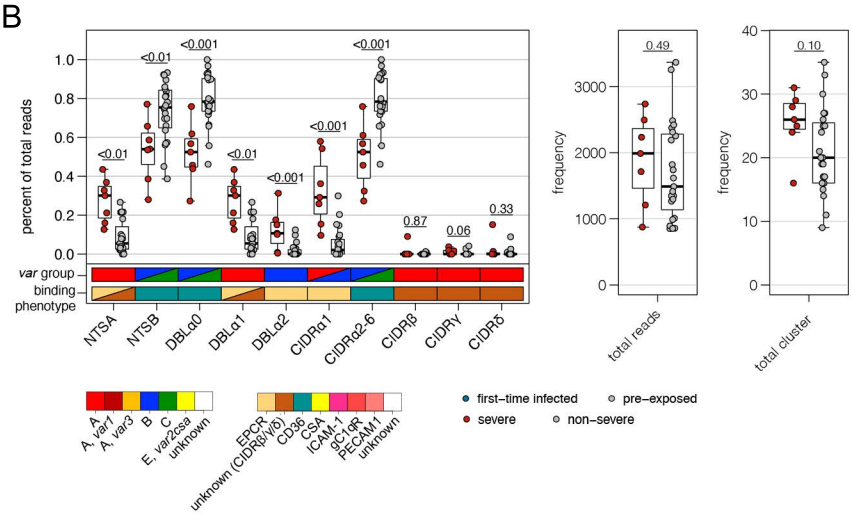




A



B



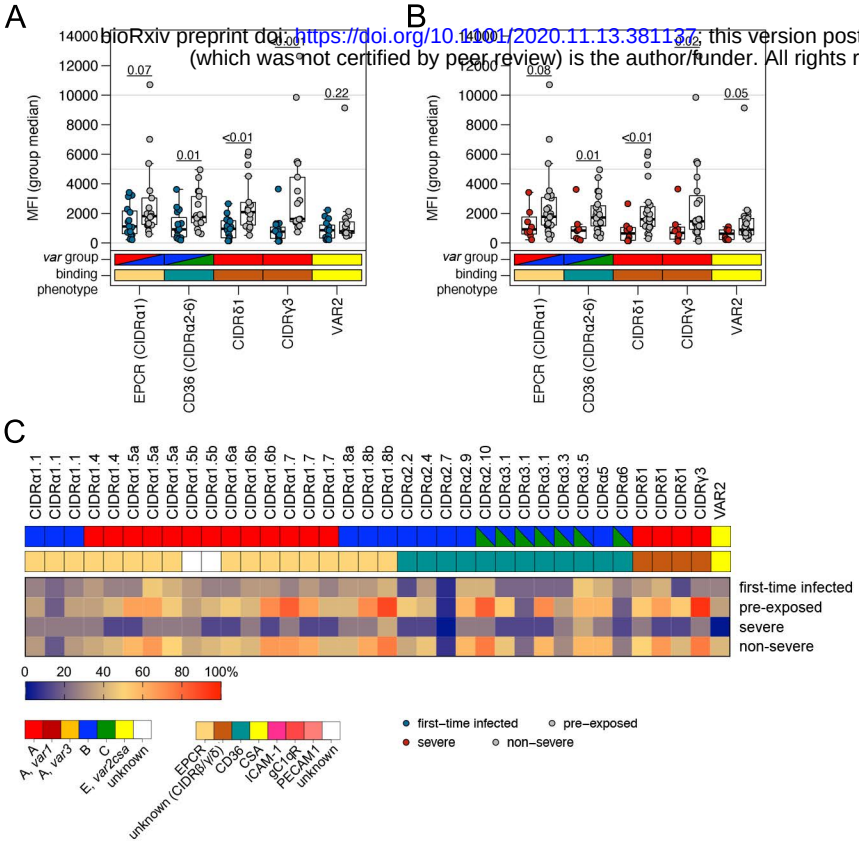


Table 1: Patient data.

	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)
Number of MSP1 genotypes [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)

Geographic origin of the parasite isolates: Ghana (n=10), Nigeria (n=6), other Sub-Saharan African countries (n=15), unknown (n=1)

* n=21