Direct *ex vivo* gene expression profiling of malaria patients reveals common virulence gene expression in first-time infected patients and severe cases

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

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

corresponding author

Running title: In host expression of *var* genes

Abstract

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

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

Introduction

Despite considerable efforts during recent years to combat malaria, 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 2018, 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. The virulence of *P. falciparum* is linked to the infected erythrocytes binding to endothelial cell surface molecules expressed on blood vessel walls. This phenomenon, known as sequestration, prevents the passage of infected erythrocytes through the spleen, which would otherwise remove the infected erythrocytes from the circulation and kill the parasite (Saul, 1999). The membrane proteins mediating sequestration are exposed to the host's immune system and through evolution *P. falciparum* parasites have acquired several multi-copy gene families coding for variant surface antigens (VSAs) allowing immune escape through extensive sequence polymorphisms. Endothelial sequestration is mediated by the *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1) family, which members have different binding capacities for host vascular tissue receptors such as CD36, EPCR,

ICAM-1, PECAM1, receptor for complement component C1q (gC1qR) and CSA (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 a EPCR-binding CIDRα1 domain or a CIDRβ/γ/δ domain of unknown function. The group A includes two conserved subfamilies: the *var1* subfamily, 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 domains. 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 other, DBLβ12, have 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 and typically DBLβ12 domains.

Group A *var* gene expression has been associated with severe forms of malaria, whereas mild malaria may be associated with group C expression (Avril *et al*, 2012; Claessens *et al*, 2012; Lavstsen *et al*, 2012; Falk *et al*, 2009; Warimwe *et al*, 2009; Kyriacou *et al*, 2006; Rottmann *et al*, 2006; Jensen *et al*, 2004; Kirchgatter & Del Portillo, 2002). Conflicting results were reported on group B expression during severe disease. However, this is likely explained by the B/A-type *Pf*EMP1 sharing the EPCR receptor-binding phenotype with A-type *Pf*EMP1. Indeed, consensus from a range of gene expression studies is that severe malaria is associated with expression of *Pf*EMP1 with EPCR-binding CIDR α1 domains (Jespersen *et al*, 2016; Kessler *et al*, 2017; Storm *et al*, 2019; Shabani *et al*, 2017; Mkumbaye *et al*, 2017; Bernabeu *et al*, 2016; Magallón-Tejada *et al*, 2016). No other domain has been consistently associated with severity of disease, but elevated expression of dual EPCR and ICAM-1-binding *Pf*EMP1 (Lennartz *et al*, 2017) and the group A associated DC5 and DC6 have been associated with severe disease. (Bernabeu *et al*, 2016; Magallón-Tejada *et al*, 2016; Berger *et al*, 2013; Avril *et al*, 2013, 2012; Claessens *et al*, 2012; Lavstsen *et al*, 2012).

In this study, we applied genome-wide expression profiling using RNA-seq to study gene expression, in particular *var* gene expression, in parasites from hospitalized travelers returning from malaria endemic countries with certified *P. falciparum* infections. Individuals were clustered into i) first-time infected and ii) pre-exposed individuals on the basis of serological data or into iii) severe and iv) non-severe cases according to medical reports. Our multi-dimensional analysis reveals a clear association of domain cassettes with EPCR-binding properties with a naïve immune status and severe malaria, whereas CD36-binding *Pf*EMP1 proteins and the conserved *var1*-3D7 variant were expressed at higher levels in pre-exposed patients and non-severe cases. Interestingly, circulating parasites from severe cases tended to be younger than parasites from non-severe cases, indicating that the EPCR-binding phenotype confers more efficient sequestration of infected erythrocytes.

Results

Cohort characterization

This study is based on a cohort of 32 malaria patients hospitalized in Hamburg, Germany. Parasitemia, recorded clinical symptoms and patient sub-grouping are summarized in Table 1. For ten patients, the present malaria episode was their first recorded *P. falciparum* infection. Nine individuals had previously experienced malaria episodes according to the medical reports, whereas malaria exposure was unknown for 13 patients. In order to determine if patients already had an immune response to *P. falciparum* antigens, indicative of previous exposure to malaria, plasma samples were analyzed by a Luminex-based assay displaying the antigens AMA1, MSP1 and CSP (Table S1). Immune responses to AMA1 and MSP1 are known to be long-lasting and seroconversion to AMA1 is assumed to occur after only a single or very few

140 infections (Drakeley *et al*, 2005). Principal component analysis of the Luminex data resulted in
141 separation of the patients into two discrete groups corresponding to first-time infected adults
142 ('naïve cluster') and malaria pre-exposed individuals ('pre-exposed cluster') (Figure 1A). The
143 13 patients with unknown malaria exposure status could be grouped into either the naïve or
144 pre-exposed groups defined by the PCA of the antigen reactivity. The only outlier in the clus-
145 tering was a 19-year-old patient (#21) from Sudan, who reported several malaria episodes
146 during childhood, but clustered with the malaria-naïve patients.
147 Plasma samples were further subjected to i) a merozoite-directed antibody-dependent respir-
148 atory burst (mADRB) assay (Kapelski *et al*, 2014), ii) a parasitophorous vacuolar membrane-
149 enclosed merozoite structures (PEMS)-specific ELISA and iii) a protein microarray with 228
150 *P. falciparum* antigens (Borrmann, 2020). Analysis of these serological assays in relation to
151 the patient clustering confirmed the expected higher and broader antigen recognition by
152 ELISA, protein microarray, and stronger ability to induce burst of neutrophils by serum from
153 the group of malaria pre-exposed patients (Figure 1B–D, Table S1). Data from all the serolog-
154 ical assays were next used for an unsupervised random forest machine learning approach to
155 build models predictive of individual's protective status. A multidimensional scaling plot was
156 used to visualize cluster allocation (Figure 1E) confirming the classification of patient #21 as
157 being non-immune (Figure 1E). The patient #26, positioned at the borderline to pre-exposed
158 patients, was grouped into the naïve cluster in accordance with the Luminex data and the
159 medical report showing that this patient from Jamaica was infected during his first trip to Africa.
160 The calculated variable importance highlighted the relevance of the mADRB assay, the ELISA
161 and the Luminex to allocate patients into cluster (Figure 1F).
162 Using protein microarrays, the antibody response against described antigens was analyzed in
163 detail. As expected, pre-exposed individuals showed significantly elevated IgG antibody re-
164 sponses against a broad range of parasite antigens, especially typical parasite blood stage
165 markers, including MSP1, MSP2, MSP4, MSP10, EBA175, ring exported protein 1 (REX1),
166 and AMA1 (Figure 1D, upper panel). Markers for a recent infection, MSP1, MSP4, GLURP and
167 Early transcribed membrane protein 5 (ETRAMP5) (Van Den Hoogen *et al*, 2019), were sig-
168 nificantly elevated in the pre-exposed individuals in comparison to the defined first-time in-
169 fected group. In addition, further members of the ETRAMP family, including ETRAMP10,
170 ETRAMP14, ETRAMP10.2 and ETRAMP4, and also antibodies against pre-erythrocytic anti-
171 gens, such as CSP, STARP and LSA3, were highly elevated. Similar effects were detectable
172 for IgM antibodies; previous exposure to the malaria parasite led to higher antibody levels
173 (Figure 1D, lower panel).
174 Based on medical reports eight patients from the malaria-naïve group were considered as
175 more severe cases, having a substantially higher parasitemia and an impaired function of di-
176 verse organs especially the brain (Table 1). Patient #1 was included into the severe group due

177 to increasing parasitemia during hospitalization and circulating schizonts indicative of a very
178 high sequestering parasite biomass associated with severity (Bernabeu *et al*, 2016). The re-
179 maining 24 cases were summarized into the non-severe malaria group. Comparing the IgG
180 antibody response of severe and non-severe cases within the previously malaria-naïve group,
181 elevated antibody levels were found in the severe subgroup. The highest fold change was
182 observed for antibodies directed against intracellular proteins, such as DnaJ protein, GTPase-
183 activating protein or heat shock protein 70 (Figure S1). Interestingly, IgM antibodies against
184 ETRAMP5 were detectable in the severely infected individuals, suggesting they were infected
185 for a prolonged period of time compared to the mild malaria population (Helb *et al*, 2015; Van
186 Den Hoogen *et al*, 2019).
187
188 ***RNA-seq transcriptomics***
189 Parasites were isolated from the venous blood of all patients for subsequent transcriptional
190 profiling. Transcriptome libraries were sequenced for all 32 patient samples (NCBI BioProject
191 ID: PRJNA679547). The number of trimmed reads ranged between 29,142,684 and
192 82,000,248 (median: 41,383,289) within the individual libraries derived from patients. The pro-
193 portion of total reads specific for *P. falciparum* were 87.7% (median; IQR: 76.7–91.3) for the
194 30 samples included in the *de novo* assembly (Table S2). Variation in parasite ages in the
195 different patient samples was analyzed with a mixture model in accordance to Tonkin-Hill *et*
196 *al.* (Tonkin-Hill *et al*, 2018) using published data from López-Barragán *et al.* as a reference
197 (López-Barragán *et al*, 2011). Parasites from first-time infected and pre-exposed patients re-
198 vealed no obvious difference in the proportion of the different parasite stages (Figure 2A) or in
199 median age (Figure 2B, Table 2). However, a small bias towards younger parasites in the
200 severe cases was observed with a median age of 8.2 hpi (IQR: 8.0–9.8) in comparison to
201 9.8 hpi in the non-severe cases (IQR: 8.2–11.4) (Figure 2C, D, Table 2). None of the samples
202 revealed high proportions of late trophozoites (all <3%), schizonts (0%) or gametocytes (all
203 <6%) (Figure 2A, C).
204
205 ***Genome-wide analysis of differential gene expression***
206 Global gene expression analysis according to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) identi-
207 fied 420 genes to be higher and 236 to be lower expressed ($p \leq 0.05$) in first-time infected
208 patients, together corresponding to 11.3% of all *P. falciparum* genes (Table S3). Similar, 362
209 genes were significantly higher and 219 lower expressed in severe cases. A gene set enrich-
210 ment analysis (GSEA) of the differentially expressed genes using Gene Ontology (GO) terms
211 and KEGG pathway annotations showed that the KEGG pathway 03410 'base excision repair'
212 facilitating the maintenance of genome integrity by repairing small bases lesions in the DNA
213 was expressed at significantly higher levels in first-time infected patient samples (Figure 3A,

214 Figure S2). In total, six out of 15 *P. falciparum* genes included in this KEGG pathway were
215 found to be statistically significant enriched upon first-time infection, including the putative en-
216 donuclease III (PF3D7_0614800) from the short-patch pathway and the putative A-/G-specific
217 adenine glycosylase (PF3D7_1129500), the putative apurinic/apyrimidinic endonuclease
218 Apn1 (PF3D7_1332600), the proliferating cell nuclear antigens 1 (PF3D7_1361900), the cat-
219 alytic (PF3D7_1017000) and small (PF3D7_0308000) subunits from the DNA polymerase
220 delta from the long-patch pathway (Figure 3B). Additionally, a significantly lower expression
221 level for genes associated with several GO terms involved in antigenic variation and host cell
222 remodeling was found in first-time infected patients (Figure 3A, Table S3) and severe cases
223 (Table S4). However, as variant surface antigens like *var*, *rif*, *stevor*, *surf* and *pfmc-2tm* are
224 largely clone-specific, analysis of reads from the clinical isolates mapping to homologous re-
225 gions in 3D7 genes would be distorted and flawed. Therefore, we analyzed *var* gene expres-
226 sion by first *de novo* assembling *var* genes from the RNA-seq data and subsequently analyzing
227 the expression of specific *var* gene subsets or according to the domains encoded. In addition,
228 we manually screened differentially expressed genes known to be involved in *var* gene regu-
229 lation or correct display of *Pf*EMP1 at the host cell surface (Table S3, S4).
230
231 ***Differential* var *gene expression***
232 To correlate individual *var* genes with a naïve immune status or disease severity, differential
233 *var* gene expression was analyzed according to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018). First,
234 *var* gene assembly was performed by assembling each sample separately, which reduces the
235 risk for generating false chimeric genes and results in longer contigs compared to a combined
236 all sample assembly approach. In total, 6,441 contigs with over 500 bp-length were generated
237 with an N50 of 2,302 bp and a maximum length of 10,412 bp (Data S1). One or more DBL or
238 CIDR domains could be annotated to 5,488 of the contigs, whereas the remaining contigs
239 could only be annotated by these domains smaller building blocks, the so-called homology
240 blocks defined by Rask *et al.* (Rask *et al*, 2010) (Table S5, S6).
241
242 *Differential* var *gene expression by specific transcript sequence*
243 A median of 200 contigs (IQR: 137–279) with >500 bp was assembled per sample. Around
244 half (47.2%: 2,592) of the domain-annotated contigs shared >98% nucleotide identity over
245 least 80% of the length of a known *var* gene found in the database varDB (Otto, 2019) (Table
246 S6). The Salmon RNA-seq quantification pipeline (Patro *et al*, 2017), which identifies equiva-
247 lence classes allowing reads to contribute to the expression estimates of multiple transcripts,
248 was used to estimate expression levels for each transcript. As a result, it accounts for the
249 redundancy present in our whole set of *var* gene contigs from all separate sample-specific

250 assemblies. We compared this approach with Corset (Davidson & Oshlack, 2014) which pre-
251 viously has been used to investigate differential expression of *var* genes in severe malaria and
252 found it gave similar results (see methods) (Tonkin-Hill *et al*, 2018). However, due to the high
253 diversity in *var* genes, both of these approaches are only able to identify significant associa-
254 tions between transcripts and phenotypes when there is sufficient similarity within the associ-
255 ated sequences.
256 By comparing transcripts expressed in first-time infected patients with those from parasites
257 isolated from pre-exposed patients using the Salmon approach, transcript levels were higher
258 for twelve and lower for three contigs in malaria-naïve hosts (Table 3, Figure 4A, B, Table S8).
259 Assembled transcripts of the conserved *var1*, *var2csa* and *var3* gene families were also found
260 at higher frequencies in first-time infected patients. Notably, the *var1-IT* variant was expressed
261 at higher levels in parasites from first-time infected patients, whereas the *var1-3D7* variant was
262 expressed at higher levels in parasites from pre-exposed and non-severe patients (Figure 4,
263 Table 3, 4). This was confirmed by mapping normalized reads from all patients to both *var1*
264 variants as well *as var2csa* (Figure S3). Several *var* fragments from B- or C-type *var* genes
265 were associated with a naïve immune status and three transcripts from A, DC8 and B-type *var*
266 genes as well as *var2csa* were linked to severe malaria patients (Figure 4, Table 3, 4, Table
267 S8, S9).
268
269 *Differential* var *gene expression on domain level:*
270 To assess differential expression of specific *var* domains, read counts corresponding to do-
271 mains with the same classification was calculated to provide domain level read counts. This
272 showed that different EPCR-binding CIDRα1 domain variants and other domains found in DCs
273 with CIDRα1 domains were expressed at significantly higher levels in first-time infected pa-
274 tients (Table 5, Figure 5A, B, Table S8). Thus, besides domains from DC8 (DBLα2, CIDRα1.1,
275 DBLβ12) and DC13 (DBLα1.7, CIDRα1.4), the CIDRα1.7 and DBLα1.2 (DC15) were increased
276 upon infection of malaria-naïve individuals. The DBLα1.2 domain was in all of the 32 gene
277 assemblies flanked by an EPCR-binding CIDRα1 domain, 56% of these were a CIDRα1.5
278 domain (Table S5, S6). In addition, parasites from first-time infected patients showed signifi-
279 cantly more transcripts encoding the CIDRδ1 domain of DC16 (DBLα1.5/6-CIDRδ1/2) (Table
280 5, Figure 5), and the DBLβ6 domain. The DBLβ6 is associated with A-type *var* genes and often
281 found adjacent to DC15 and DC16 (Otto *et al*, 2019), supporting the association of both domain
282 cassettes with a naïve immune status. In general, all domains associated with the same do-
283 main cassettes showed the same trend even if some domains did not reach statistical signifi-
284 cance set at $p < 0.05$ (Table S8).
285 Domains found expressed at lower levels in malaria-naïve included group B and C N-terminal
286 head structure domains NTSB, DBLα0.13/22/23 and CD36-binding CIDR domains

287 (CIDRα2.8/9,6) as well as the C-terminal CIDRγ11 domain and domains of the *var1-3D7* vari-
288 ant (DBLα1.4, DBLγ15, DBLε5) (Table 5, Figure 5A, B).
289 When comparing severe to the non-severe cases, domains of DC8 (DBLα2, CIDRα1.1,
290 DBLβ12), DC15 (DBLα1.2) and DC16 (DBLα1.6) were found associated with severe disease
291 (Table 6, Figure 5C, D, Table S9). Domain types expressed at significantly higher levels in
292 non-severe cases included N-terminal head structure domains, DBLα0.23, CIDRα2.4/9 from
293 group B and C *Pf*EMP1, DC16 (DBLα1.5) and the CIDRα1.3 domain from the *var1-3D7* allele
294 (Table 6, Figure 5C, D).
295 As DBL and CIDR sub-classes are poorly defined (Otto *et al*, 2019) and different domain sub-
296 classes confer the same binding phenotype (Higgins & Carrington, 2014), the domains of the
297 N-terminal head structure were grouped according to their binding phenotype and the normal-
298 ized read counts (TPM) were summarized for each patient (Figure 5E, F). This showed clear
299 differences for domains associated with EPCR- or CD36-binding *Pf*EMP1. As expected, do-
300 mains associated with EPCR-binding as well as the CIDRγ3 domain were expressed at higher
301 levels in naïve and more severe cases, whereas domains associated with the CD36-binding
302 phenotype were higher expressed in pre-exposed and non-severe patients.
303
304 *Differential* var *gene expression on homology block level:*
305 *Pf*EMP1 domains have been described as composed of 628 homology blocks (Rask *et al*,
306 2010). Homology block expression levels were obtained by aggregating read counts for each
307 block after first identifying all occurrences of the block within the assembled contigs. Tran-
308 scripts encoding blocks number 255, 584 and 614, all typically located within DBLβ domains
309 of DC8 and CIDRα1-containing type A *Pf*EMP1 (Table 7, Figure 6A, B, Table S8), number
310 557, located in the inter-domain region between DBLβ and a DBLγ domains (no *Pf*EMP1 type
311 association) and block number 155 found in NTSA, were found associated with a naïve im-
312 mune status. Conversely, transcripts encoding block 88 from DBLα0 domains and 269 from
313 ATSB were found at lower levels in malaria-naïve patients indicating that B- and C-type genes
314 are more frequently expressed in pre-exposed individuals (Table 7, Figure 6A, B). No homol-
315 ogy blocks were associated with severe cases, but two blocks, 591 and 559, found in group B
316 *Pf*EMP1 were found to be lower expressed in severe malaria cases (Table 8, Figure 6C, D,
317 Table S9).
318
319 ***Var expression profiling by DBLα-tag sequencing***
320 To supplement the RNA-seq analysis with an orthogonal analysis, we conducted deep se-
321 quencing of RT-PCR-amplified DBLα expression-tags from 30 of the patient samples
322 (Lavstsen *et al*, 2012). Between 851 to 3,368 reads with a median of 1,666 over all samples
323 were analyzed. Identical DBLα-tag sequences were clustered to generate relative expression

324 levels of each unique *var* gene tag. Overall, the relative expression levels were similar for
325 sequences found in both the RNA-seq and the DBLα-tag approach with a mean log2(DBLα-
326 PCR/RNA-seq) ratio of 0.4 (CI of 95%: -2.5–3.3) determined by Bland-Altman plotting (Figure
327 S4). Around 82.6% (median) of all unique DBLα-tag sequences detected with >10 reads
328 (92.9% of all DBLα-tag reads) were found in the RNA-seq approach; and 81.8% of the upper
329 75th percentile of RNA-seq contigs (spanning across the DBLα-tag region) were found by the
330 DBLα-tag approach.
331 Using the Varia tool (Mackenzie *et al*, 2020) the domain composition of the *var* genes from
332 which the DBLα-tag sequences originated were predicted. The tool searches the varDB ver-
333 sion V1.1 (Otto, 2019) containing >200K annotated *var* genes for genes with near identical
334 DBLα-tag sequences and returns the consensus domain annotation among the hit sequences,
335 if such exists. A partial domain annotation was made for ~85% of the DBLα-tag sequences
336 (Table S10). In line with the RNA-seq data, this analysis showed that DBLα1 and DBLα2 se-
337 quences were enriched in first-time infected and severe malaria patients. Conversely, a signif-
338 icant higher proportion of DBLα0 sequences was found in pre-exposed individuals and mild
339 cases (Figure 7A, B). No difference was observed in the number of reads or unique DBLα-tags
340 detected between patient groups, although a trend towards more DBLα-tag clusters was ob-
341 served in first-time infected patients and severe cases (Figure 7A, B). A prediction of the NTS
342 and CIDR domains flanking the DBLα domain showed a significantly higher proportion of NTSA
343 in severe cases as well as EPCR-binding CIDRα1 domains in first-time infected and severe
344 cases. Expression of *var* genes encoding NTSB and CIDRα2-6 domains were significantly
345 associated with in pre-exposed and non-severe cases (Figure 7A, B). Subsequent analysis of
346 *var* expression in relation to other domains, showed *var* transcripts with DBLβ, γ and ζ or
347 CIDRγ domains were more frequently expressed in first-time infected and severe malaria pa-
348 tients whereas those encoding DBLδ and CIDRβ were less frequent (Figure 7C, D). Assessing
349 expression in relation to domain subtype, CIDRα1.1/5, DBLβ12, DBLγ2/12 and DBLα2 were
350 found associated with severe malaria, while CIDRα3.1/4, DBLα0.12/16, and DBLδ1 associated
351 with non-severe cases (Figure 7D).
352 Overall, these data corroborate the main observations from the RNA-seq analysis, confirming
353 the association of EPCR-binding *Pf*EMP1 variants with development of severe malaria symp-
354 toms and CD36-binding *Pf*EMP1 variants with establishment of less severe infections in semi-
355 immune individuals. A summary of the differential *var* gene expression data on the multi-, sin-
356 gle- and subdomain level from the RNA-seq and DBLα-tag approaches is shown in Figure 8.
357
358 ***Correlation of* var *gene expression with antibody levels against head structure CIDR***
359 ***domains***

360 A detailed analysis of the antibody repertoire of the patients against head structure CIDR do-
361 mains of *Pf*EMP1 was carried out using a panel of 19 different EPCR-binding CIDRα1 do-
362 mains, 12 CD36-binding CIDRα2–6 domains, three CIDRδ1 domains as well as a single
363 CIDRγ3 domain (Obeng-Adjei *et al*, 2020; Bachmann *et al*, 2019). Additionally, the minimal
364 binding region of VAR2CSA was included. In general, plasma samples from malaria-naïve as
365 well as severe cases showed lower MFI values for all antigens tested in comparison to samples
366 from pre-exposed or non-severe cases with significant differences for CIDRα2–6, CIDRδ1 and
367 CIDRγ3, but not for EPCR-binding CIDRα1 domains (Figure 9A, B).
368 The data was also analyzed using the average MFI reactivity (plus two standard deviations) of
369 an Danish control cohort as a cut off for seropositivity to calculate the coverage of antigen
370 recognition (Cham *et al*, 2010). In this analysis, almost half of the tested antigens were recog-
371 nized by pre-exposed (median: 46.3%) and non-severe patients (median: 45.1%), but only 1/4
372 of the antigens were recognized by first-time infected patients (median: 24.4%) and 1/20 by
373 severely ill patients (median: 4.9%). *Pf*EMP1 antigens recognized by over 60% of the pre-
374 exposed and/or non-severe patient sera were i) four CIDRα1 domains capable of EPCR-
375 binding (CIDRα1.5, CIDRα1.6, CIDRα1.7 and the DC8 domain CIDRα1.8), ii) two CD36-
376 binding CIDRα domains (CIDRα2.10, CIDRα3.1) and iii) two CIDR domains with unknown
377 binding phenotype (CIDRδ1 and CIDRγ3) (Figure 9C, Table S1).
378
379 **Discussion**
380
381 Analysis of blood samples from travelers returning to Germany and hospitalized with *P. falci-*
382 *parum* malaria for the first time, allows studies of parasites' development and gene expression
383 unaffected by host immune responses elicited in previous *Plasmodia* infections. Here we an-
384 alyzed gene expression in 32 such patients using direct *ex vivo* RNA-seq to exclude transcrip-
385 tional adaptation through *in vitro* cultivation. Most of the returning travelers studied were in-
386 fected with a single or very few parasite genotypes, which likely simplified the analysis of *var*
387 gene expression due to the restricted genomic repertoire. We were able to distinguish retro-
388 spectively between first-time infected and pre-exposed patients by medical history and as-
389 sessing the presence of antibodies to *P. falciparum* antigens. Eight out of 15 first-time infected
390 patients were classified as severe malaria cases due to impaired organ function according to
391 their medical reports. Despite the relatively low number of patient samples, our RNA-seq ap-
392 proach confirmed previously reported associations between transcripts encoding type A and
393 B EPCR-binding *Pf*EMP1 and infections in naïve hosts and disease severity (Table S6,
394 S7)(Duffy *et al*, 2019; Tonkin-Hill *et al*, 2018; Kessler *et al*, 2017; Bernabeu *et al*, 2016;
395 Jespersen *et al*, 2016; Lavstsen *et al*, 2012).

396 Overall, there was a high degree of consensus between observations made on *var* expression
397 analyzed at different levels of *Pf*EMP1 domain annotation. Stratifying *var* gene expression ac-
398 cording to different main and subtype of DBL and CIDR domains, showed only A- and DC8-
399 type *Pf*EMP1 domains, and predominantly those linked to EPCR-binding *Pf*EMP1, to be asso-
400 ciated with first-time infections. Conversely, domains typical for CD36-binding *Pf*EMP1 pro-
401 teins was found at higher levels in malaria-experienced patients. Specifically, expression of
402 *Pf*EMP1 domains included in DC8, DC13 and DC15 as well as all EPCR-binding CIDRα1 do-
403 mains were associated with first time infections, whereas DBLα0 and CD36-binding CIDRα2-
404 6 domains were linked to pre-exposed individuals. These differences were in large due to the
405 differential expression between the first-time infected patients with more severe symptoms and
406 patients with non-severe malaria. Here, domains of DC8 and DC15 as well as all DBLα1/2 and
407 CIDRα1 domains were associated with severe symptoms, while NTSB, DBLα0, CIDRα2-6 do-
408 mains including specific subsets of CIDRα2 were linked to non-severe symptoms. These con-
409 clusions were closely mirrored in the DBLα tag analysis, and was further corroborated by the
410 differential RNA-seq expression stratified according to the smaller homology blocks, which
411 identified mainly homology blocks of DBLβ1, 3, 5 and 12 DBLβ domains to be associated with
412 first-time infected patients. These DBLβ domains are parts of DCs associated with EPCR-
413 binding, so it is hard to distinguish between co-occurring domains and clear associations.
414 In addition, three other group A *Pf*EMP1-associated domains, CIDRγ3, CIDRδ from the DC16
415 and DBLβ9 from DC5, were found associated with first-time infected patients. DC5 could have
416 been detected due to its presence C-terminally to some EPCR-binding *Pf*EMP1. However, the
417 CIDRδ domain of DC16 (DBLα1.5/6-CIDRδ1/2) constitute a different subset of A-type *Pf*EMP1,
418 which together with CIDRβ2 and CIDRγ3-containing group A *Pf*EMP1 (found in DC11) may be
419 associated with rosetting (Carlson *et al*, 1990; Ghumra *et al*, 2012). Direct evidence that any
420 of these CIDR domains have intrinsic rosetting properties is lacking (Rowe *et al*, 2002). Rather,
421 their association with rosetting may be related to their tandem expression with DBLα1 at the
422 N-terminal head (Ghumra *et al*, 2012). The DC16 group A signature was not associated with
423 severe disease outcome in previous DBLα-tag studies or qPCR studies by Lavstsen *et al.*
424 (Lavstsen *et al*, 2012) and Bernabeu *et al.* (Bernabeu *et al*, 2016), but DBLα1.5/6 and CIDRδ
425 of DC16 were enriched in cerebral malaria cases with retinopathy in the study of Shabani *et*
426 *al.* (Shabani *et al*, 2017) and Kessler *et al.* (Kessler *et al*, 2017) using the same qPCR primer
427 set. Also, association of DC11 with severe malaria in Indonesia was found using the same
428 RNA-seq approach as used here (Tonkin-Hill *et al*, 2018). Rosetting is thought to enhance
429 microvascular obstruction but its role in severe malaria pathogenesis remains unclear
430 (McQuaid & Rowe, 2020). Together with previous observations, our data suggest that pediatric
431 cerebral malaria infections are dominated by the expansion of parasites expressing EPCR-
432 binding domains accompanied by parasites expressing other group A *Pf*EMP1, possibly acting

433 as rosetting variants. Also, this aligns with pre-exposed patients recognized several CIDRα1
434 domains more frequently than CIDRα2-6 domains, similar to what has been observed malaria
435 endemic populations, indicating that IgG against EPCR-binding CIDR domains were acquired
436 before IgG to other CIDR domains (Obeng-Adjei *et al*, 2020; Cham *et al*, 2009, 2010; Turner
437 *et al*, 2015).
438 Data from immunologically naïve malaria patients are limited, restricting our comparison mainly
439 to studies of severe malaria in malaria endemic regions. However, a recent study explored *var*
440 gene expression during infancy in Kenyan children and found that the waning of maternal an-
441 tibodies correlated with increasing transcription of DC8, DC13 and A-type *var* genes. After the
442 first year of life the amount of these transcripts decreased with age and acquired immunity
443 (Kivisi *et al*, 2019). A high expression of A-type *var* genes in naïve malaria cases imported to
444 France and an association of DC4, 8 and 13 with disease severity has also been reported
445 (Argy *et al*, 2017). Both qPCR studies are in agreement with the RNA-seq data from our cohort
446 of immunologically malaria-naïve adults, although we here in addition found a higher expres-
447 sion of DC15 (EPCR binding) and DC16 (putative rosetting variants) *var* genes in malaria-
448 naïve patients.
449 To the best of our knowledge, this study is the first description of expression differences be-
450 tween the two *var1* variants, 3D7 and IT. The *var1-IT* variant was found enriched in parasites
451 from first-time infected patients, whereas several transcripts of the *var1*-3D7 variant were in-
452 creased in pre-exposed and non-severely ill patients. Expression of the *var1* gene was previ-
453 ously observed to be elevated in malaria cases imported to France with an uncomplicated
454 disease phenotype (Argy *et al*, 2017). In general, the *var1* subfamily is ubiquitously transcribed
455 (Winter *et al*, 2003; Duffy *et al*, 2006), atypically late in the cell cycle after transcription of *var*
456 genes encoding the adhesion phenotype (Kyes *et al*, 2003; Duffy *et al*, 2002) and is annotated
457 as a pseudogene in 3D7 due to a premature stop codon. Similarly, numerous isolates display
458 frame-shift mutations often in exon 2 in the full gene sequences (Rask *et al*, 2010). However,
459 none of these studies addressed differences in the two *var1* variants that were recently identi-
460 fied by comparing *var* gene sequences from 714 *P. falciparum* genomes (Otto *et al*, 2019),
461 and to date it is still unclear if both variants fulfill the same function or have the same charac-
462 teristics previously described. Overall, the *var1* gene – and the first 3.2 kb of the 3D7 variant
463 in particular – seems to be under high evolutionary pressure (Otto *et al*, 2019) and both vari-
464 ants can be traced back before the split of *P. reichenowi* from *P. praefalciparum* and *P. falci-*
465 *parum* (Otto *et al*, 2018b). Our data indicate that the two variants, VAR1-3D7 and VAR1-IT,
466 may have different roles during disease, however, this remains to be determined in future
467 studies.
468 A recent study showed that parasites isolated from symptomatic infections were on average
469 younger than blood-circulating parasites from asymptomatic infections presumably due to a

470 more efficient sequestration of parasites in the symptomatic cases (Andrade *et al*, 2020).
471 Based on this and our findings we hypothesize that parasites circulating in severely ill patients
472 are younger due to the expression of EPCR and ICAM-1-binding *Pf*EMP1 variants, whereas
473 parasites in non-severe patients may circulate longer due to their expression of *Pf*EMP1 bind-
474 ing to CD36. Although affinity of the CIDR domains to each of these receptors is similar with
475 median K_d values of 12 nM for CD36 (Hsieh *et al*, 2016) and 16.5 nM for EPCR (Lau *et al*,
476 2015), recent papers describe a rolling binding phenotype of infected red blood cells over
477 CD36 and a static binding for EPCR and ICAM-1 under flow conditions (Lubiana *et al*, 2020;
478 Bernabeu *et al*, 2019; Dasanna *et al*, 2017; Herricks *et al*, 2013). Due to their biconcave shape
479 trophozoites seem to roll less stable by flipping over CD36-expressing cells, whereas schizonts
480 roll over longer distances at different shear stresses applied. This might explain why young
481 trophozoites, but not older trophozoites or schizont stages, are found in blood samples from
482 non-severe cases (Tonkin-Hill *et al*, 2018).
483 The parasite population in first-time infected individuals may have broader binding potential
484 after liver release as there is no pre-existing immunity to clear previously experienced *Pf*EMP1
485 variants. During the blood stage infection EPCR- and ICAM-1-binding and rosetting-mediating
486 variants may confer a parasite growth advantage in malaria-naïve hosts, and increase the risk
487 for severe malaria, and selection towards these variants may already have occurred in our
488 severe malaria patients. This assumption is supported by: (i) the observation that the severe
489 cases within the malaria-naïve patient group appeared to be have been infected for a longer
490 period (ii) the difference in parasite age between severe and non-severe malaria cases calcu-
491 lated by matching RNA-seq data to a reference data set (López-Barragán *et al*, 2011); (iii) the
492 correlation with a higher *var1* expression in parasite from non-severe and pre-exposed pa-
493 tients, which expression is not suppressed in trophozoites (Kyes *et al*, 2003); (iv) the expres-
494 sion of EPCR- & ICAM-1-binding variants in parasites from severe and first-time infected pa-
495 tients is significantly increased, conversely transcripts of CD36-binding variants are found
496 more frequently in parasites from non-severe and pre-exposed patients. For the parasite's
497 survival and transmission, it may be highly beneficial to have more less virulent *Pf*EMP1 vari-
498 ants able to bind CD36. This interaction may not, or is less likely to, lead to obstruction of blood
499 flow, inflammation and organ failure at least of the brain, where CD36 is nearly absent.
500
501 **Material and methods**
502
503 ***Ethics statement***
504 The study was conducted according to the principles of the Declaration of Helsinki in its 6th
505 revision as well as International Conference on Harmonization–Good Clinical Practice (ICH-
506 GCP) guidelines. All patients, aged 19 to 70 years, provided written informed consent for this

507 study, which was approved by the Ethical Review Board of the Medical Association of Ham-
508 burg (reference number PV3828).
509
510 ***Blood sampling and processing***
511 Blood samples from *P. falciparum* malaria patients collected either at the diagnostic unit of the
512 Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany, or at the Medical Clinic
513 and Polyclinic of University Clinic Hamburg-Eppendorf, Hamburg, Germany, were used in this
514 study. EDTA blood samples (1–30 mL) were obtained from the patients. The plasma was sep-
515 arated by centrifugation and immediately stored at -20°C. Erythrocytes were isolated by Ficoll
516 gradient centrifugation followed by filtration through Plasmodipur filters (EuroProxima) to clear
517 the remaining granulocytes. An aliquot of erythrocytes (about 50–100 µl) was separated and
518 further processed for gDNA purification. At least 400 µl of purified erythrocytes were rapidly
519 lysed in 5 volumes pre-warmed TRIzol (ThermoFisher Scientific) and stored at -80°C until fur-
520 ther processing (*ex vivo* samples, n = 32).
521
522 ***Serological assays***
523 *Luminex assay*
524 The Luminex assay was conducted as previously described using the same plex of antigens
525 tested (Bachmann *et al*, 2019). In brief, plasma samples from patients were screened for indi-
526 vidual recognition of 19 different CIDRα1, 12 CIDRα2–6, three CIDRδ1 domains and a single
527 CIDRγ3 domain as well as of the controls AMA1, MSP1, CSP, VAR2CSA (VAR2), tetanus
528 toxin (TetTox) and BSA. The data are shown as mean fluorescence intensities (MFI) allowing
529 comparison between different plasma samples, but not between different antigens. Alterna-
530 tively, the breadth of antibody recognition (%) was calculated using MFI values from Danish
531 controls plus two standard deviations (SD) as cut off.
532
533 *Merozoite-triggered antibody-dependent respiratory burst (mADRB)*
534 The assay to determine the mADRB activity of the patients was set up as described before
535 (Kapelski *et al*, 2014). Polymorphonuclear neutrophil granulocytes (PMNs) from one healthy
536 volunteer were isolated by a combination of dextran-sedimentation and Ficoll-gradient centrif-
537 ugation. Meanwhile, 1.25 x 10^6 merozoites were incubated with 50 µl of 1:5 diluted plasma
538 (decomplemented) for 2 h. The opsonized merozoites were pelleted (20 min, 1500 g), re-sus-
539 pended in 25 µl HBSS and then transferred to a previously blocked well of an opaque 96 well
540 high-binding plate (Greiner Bio-One). Chemiluminescence was detected in HBSS using 83.3
541 µM luminol and 1.5 x 10^5 PMNs at 37°C for 1 h to characterize the PMN response, with read-
542 ings taken at 2 min intervals using a multiplate reader (CLARIOstar, BMG Labtech). PMNs
543 were added in the dark, immediately before readings were initiated.

544

545 *ELISA*

546 Antibody reactivity against parasitophorous vacuolar membrane-enclosed merozoite struc-
547 tures (PEMS) was estimated by ELISA. PEMS were isolated as described before (Llewellyn *et*
548 *al*, 2015). For the ELISA, 0.625 x 10^5 PEMS were coated on the ELISA plates in PBS. Plates
549 were blocked using 1% Casein (Thermo Scientific #37528) and incubated for 2 h at 37°C. After
550 washing using PBS/0.1% Tween, plasma samples were added at two-fold dilutions of 1:200 to
551 1:12800 in PBS/0.1% Casein. The samples were incubated for 2 h at room temperature (RT).
552 IgG was quantified using HRP-conjugated goat anti-human IgG at a dilution of 1:20,000 and
553 incubated for 1 h. For the color reaction, 50 µl of TMB substrate was used and stopped by
554 adding 1 M HCl after 20 min. Absorbance was quantified at 450 nm using a multiplate reader
555 (CLARIOstar, BMG Labtech).

556

557 *Protein microarray*

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

568 Study serum samples as well as the European control serum were diluted 1:50 in 0.05X Super
569 G Blocking Buffer (Grace Bio-Labs, Inc.) containing 10% *E. coli* lysate (GenScript, Piscataway,
570 NJ) and incubated for 30 minutes on a shaker at RT. Meanwhile, microarray slides were rehy-
571 drated using 0.05X Super G Blocking buffer at RT. Rehydration buffer was subsequently re-
572 moved and samples added onto the slides. Arrays were incubated overnight at 4°C on a shaker
573 (180 rpm). Serum samples were removed the following day and microarrays were washed
574 using 1X TBST buffer (Grace Bio-Labs, Inc.). Secondary antibodies were then applied at a
575 dilution of 1:200 and incubated for two hours at RT on the shaker, followed by another washing
576 step and a one-hour incubation in a 1:250 dilution of Qdot™585 Streptavidin Conjugate. After
577 a final washing step, slides were dried by centrifugation at 500 g for 10 minutes. Slide images
578 were taken using the ArrayCAM® Imaging System (Grace Bio-Labs) and the ArrayCAM 400-
579 S Microarray Imager Software.

580 Microarray data were analyzed in R statistical software package version 3.6.2. All images were
581 manually checked for any noise signal. Each antigen spot signal was corrected for local back-
582 ground reactivity by applying a normal-exponential convolution model (McGee & Chen, 2006)
583 using the RMA-75 algorithm for parameter estimation (available in the LIMMA package
584 v3.28.14) (Silver *et al*, 2009). Data was log2-transformed and further normalized by subtraction
585 of the median signal intensity of mock expression spots on the particular array to correct for
586 background activity of antibodies binding to *E. coli* lysate. After log2 transformation data ap-
587 proached normal distribution. Differential antibody levels (protein array signal) in the different
588 patient groups were determined by Welch-corrected Student's t-test. Antigens with $p<0.05$ and
589 a fold change >2 of mean signal intensities were defined as differentially recognized between
590 the tested sample groups. Volcano plots were generated using the PAA package (Turewicz *et*
591 *al*, 2016) and GraphPad Prism 8. Individual antibody breadths were defined as number of
592 seropositive features with signal intensities exceeding an antigen-specific threshold set at six
593 standard deviations above the mean intensity in negative control samples.
594
595 *Unsupervised random forest model*
596 An unsupervised random forest (RF) model, a machine learning method based on multiple
597 classification and regression trees, was calculated to estimate proximity between patients.
598 Variable importance was calculated, which shows the decrease in prediction accuracy if values
599 of a variable are permuted randomly. The *k*-medoids clustering method was applied on the
600 proximity matrix to group patients according to their serological profile. Input data for random
601 forest were Luminex measurements for MSP1, AMA1 and CSP reduced by principal compo-
602 nent analysis (PCA; first principal component selected), mADRB, ELISA, and antibody breadth
603 of IgG and IgM determined by protein microarray were used to fit the RF model. Multidimen-
604 sional scaling was used to display patient cluster. All analyses were done with R (4.02) using
605 the packages randomForest (4.6-14) to run RF models and cluster (2.1.0) for k-medoids clus-
606 tering.
607
608 ***DNA purification and MSP1 genotyping***
609 Genomic DNA was isolated using the QIAamp DNA Mini Kit (Qiagen) according to the manu-
610 facturer's protocol. To assess the number of *P. falciparum* genotypes present in the patient
611 isolates, MSP1 genotyping was carried out as described elsewhere (Robert *et al*, 1996).
612
613 ***RNA extraction, RNA-seq library preparation, and sequencing***
614 TRIzol samples were thawed, mixed rigorously with 0.2 volumes of cold chloroform and incu-
615 bated for 3 min at room temperature. After centrifugation for 30 min at 4°C and maximum

616 speed, the supernatant was carefully transferred to a new tube and mixed with an equal vol-
617 ume of 70% ethanol. Afterwards the manufacturer's instruction from the RNeasy MinElute Kit
618 (Qiagen) were followed with DNase digestion (DNase I, Qiagen) for 30 min on column. Elution
619 of the RNA was carried out in 14 µl. Human globin mRNA was depleted from all samples
620 except from samples #1 and #2 using the GLOBINclear kit (ThermoFisher Scientific). The
621 quality of the RNA was assessed using the Agilent 6000 Pico kit with the Bioanalyzer 2100
622 (Agilent) (Figure S5), the RNA quantity using the Qubit RNA HA assay kit and a Qubit 3.0
623 fluorometer (ThermoFisher Scientific). Upon arrival at BGI Genomics Co. (Hong Kong), the
624 RNA quality of each sample was double-checked before sequencing. The median RIN value
625 over all *ex vivo* samples was 6.75 (IQR: 5.93–7.40) (Figure S5), although this measurement
626 has only limited significance for samples containing RNA of two species. Customized library
627 construction in accordance to Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) including amplification
628 with KAPA polymerase and HiSeq 2500 100 bp paired-end sequencing was also performed
629 by BGI Genomics Co. (Hong Kong).
630
631 ***RNA-seq read mapping and data analysis***
632 *Differential expression of core genes*
633 Differential gene expression analysis of *P. falciparum* core genes was done in accordance to
634 Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018) by applying the script as given in the Github repository
635 (https://github.com/gtonkinhill/falciparum_transcriptome_manuscript/tree/master/all_gene_an
636 alysis). In brief, subread-align v1.4.6 (Liao *et al*, 2013) were used to align the reads to the
637 *H. sapiens* and *P. falciparum* reference genomes. Read counts for each gene were obtained
638 with FeatureCounts v1.20.2 (Liao *et al*, 2014). To account for parasite life cycle, each sample
639 is considered as a composition of six parasite life cycle stages excluding the ookinete stage
640 (López-Barragán *et al*, 2011). Unwanted variations were determined with the 'RUV' (Remove
641 Unwanted Variation) algorithm implemented in the R package ruv v0.9.6 (Gagnon-Bartsch &
642 Speed, 2012) adjusting for systematic errors of unknown origin by using the genes with the
643 1009 lowest p-values as controls as described in (Vignali *et al*, 2011). The gene counts and
644 estimated ring-stage factor, and factors of unwanted variation were then used as input for the
645 Limma /Voom (Law *et al*, 2014; Smyth, 2005) differential analysis pipeline.
646
647 *Functional enrichment analysis of differentially expressed core genes*
648 Genes that were identified as significantly differentially expressed (defined as -1<logFC>1,
649 p<0.05) during prior differential gene expression analysis were used for functional enrichment
650 analysis using the R package gprofiler2 (Kolberg *et al*, 2020). Enrichment analysis was per-
651 formed on multiple input lists containing genes expressed significantly higher (logFC > 1, P <
652 0.05) and lower (logFC < - 1, P < 0.05) between different patient cohorts. All *var* genes were

653 excluded from the enrichment analysis. For custom visualization of results, gene set data
654 sources available for *P. falciparum* were downloaded from gprofiler (Raudvere *et al*, 2019).
655 Pathway data available in the KEGG database (https://www.kegg.jp/kegg/) was accessed via
656 the KEGG API using KEGGREST (Tenenbaum, 2020) to supplement gprofiler data sources
657 and build a custom data source in Gene Matrix Transposed file format (*.gmt) for subsequent
658 visualization. Functional enrichment results were then output to a Generic Enrichment Map
659 (GEM) for visualization using the Cytoscape EnrichmentMap app (Merico *et al*, 2010) and
660 RCy3 (Gustavsen *et al*, 2019). Bar plots of differential gene expression values for genes of
661 selected KEGG pathways were generated using ggplot2 (Wickham, 2016) and enriched KEGG
662 pathways were visualized using KEGGprofile (Zhao S, Guo Y, 2020).
663
664 Var *gene assembly*
665 Samples from patient #1 and #2 not subjected to globin-mRNA depletion due to their low RNA
666 content after multiple rounds of DNase treatment showed low percentages of *P. falciparum*
667 specific reads (12.4% and 15.68%) (Table S2). Consequently, less than one million *P. falcipa-*
668 *rum* reads were obtained for each of these samples and they were omitted from assembly due
669 to low coverage, but included in the differential gene expression analysis.
670 *Var* genes were assembled using the pipeline described in Tonkin-Hill *et al.* (Tonkin-Hill *et al*,
671 2018). Briefly, non-*var* reads were first filtered out by removing reads that aligned to *H. sapi-*
672 *ens*, *P. vivax* or non-*var P. falciparum*. Assembly of the remaining reads was then performed
673 using a pipeline combining SOAPdenovo-Trans and Cap3 (Xie *et al*, 2014; Huang & Madan,
674 1999; Liao *et al*, 2013). Finally, contaminants were removed from the resulting contigs and
675 they were then translated into the correct reading frame. Reads were mapped to the contigs
676 using BWA-MEM (Li, 2013) and RPKM values were calculated for each *var* transcript to com-
677 pare individual transcript levels in each patient. Although transcripts might be differentially cov-
678 ered by RNA-seq due to their variable GC content, this seems not to be an issue between *var*
679 genes (Tonkin-Hill *et al*, 2018).
680
681 Var *transcript differential expression*
682 Expression for the assembled *var* genes was quantified using Salmon v0.14.1 (Patro *et al*,
683 2017) for 531 transcripts with five read counts in at least 3 patient isolates. Both the naïve and
684 pre-exposed groups as well as the severe and non-severe groups were compared. The com-
685 bined set of all *de novo* assembled transcripts was used as a reference in addition to the coding
686 regions of the 3D7 reference genome. The Salmon algorithm identifies equivalence sets be-
687 tween transcripts allowing a single read to support the expression of multiple transcripts. This
688 enables it to account for the redundancy present in our dataset. To confirm the suitability of
689 this approach we also ran the Corset algorithm as used in Tonkin-Hill *et al.*, (Tonkin-Hill *et al*,

690 2018; Davidson & Oshlack, 2014). Unlike Salmon which attempts to quantify the expression
691 of transcripts themselves, Corset copes with the redundancy present in *de novo* transcriptome
692 assemblies by clustering similar transcripts together using both the sequence identify of the
693 transcripts as well as multi-mapping read alignments. Of the transcripts identified using the
694 Salmon analysis 5/15 in the naïve versus pre-exposed and 4/13 in the severe versus non-
695 severe were identified in the significant clusters produced using Corset. As the two algorithms
696 take very different approaches and as Salmon is quantifying transcripts rather than the 'gene'
697 like clusters of Corset this represents a fairly reasonable level of concordance between the
698 two methods. In both the Salmon and Corset pipelines differential expression analysis of the
699 resulting *var* expression values was performed using DESeq2 v1.26 (Love *et al*, 2014). The
700 Benjamini-Hochberg method was used to control for multiple testing (Benjamini & Hochberg,
701 1995).
702 To check differential expression of the conserved *var* gene variants *var1*-3D7, *var1*-IT and
703 *var2csa* raw reads were mapped with BWA-MEM (AS score >110) to the reference genes from
704 the 3D7 and the IT strains. The mapped raw read counts (bam files) were normalized with the
705 number of 3D7 mappable reads in each isolate using bamCoverage by introducing a scaling
706 factor to generate bigwig files displayed in Artemis (Carver *et al*, 2012).
707
708 Var *domain and segment differential expression*
709 Differential expression analysis at the domain and segment level was performed using a similar
710 approach to that described previously (Tonkin-Hill *et al*, 2018). Initially, the domain families
711 and homology blocks defined in Rask *et al.* were annotated to the assembled transcripts using
712 HMMER v3.1b2 (Rask *et al*, 2010; Eddy, 2011). Domains and segments previously identified
713 to be significantly associated with severe disease in Tonkin-Hill *et al.,* 2018 were also anno-
714 tated by single pairwise comparison in the assembled transcripts using USEARCH v11.0.667
715 (Tonkin-Hill *et al*, 2018; Edgar, 2010). Overall, 336 contigs (5.22% of all contigs >500 bp) pos-
716 sess partial domains in an unusual order, e.g., an NTS in an internal region or a tandem ar-
717 rangement of two DBLα or CIDRα domains. This might be caused by *de novo* assembly errors,
718 which is challenging from transcriptome data. Therefore, in both cases the domain or segment
719 with the most significant alignment was taken as the best annotation for each region of the
720 assembled *var* transcripts (E-value cutoff of $1e^{-}8$). The expression at each of these annotations
721 was then quantified using featureCounts v1.6.4 before the counts were aggregated to give a
722 total for each domain and segment family in each sample. Finally, similar to the transcript level
723 analysis, DESeq2 was used to test for differences in expression levels of both domain and
724 segment families in the naïve versus pre-exposed groups as well as the severe versus non-
725 severe groups. Again, more than five read counts in at least three patient isolates were re-
726 quired for inclusion into differential expression analysis.

727

728 ***DBLα-tag sequencing***

729 For DBLα-tag PCR the forward primer varF_dg2 (5'-tcgtcggcagcgtcagatgtgtataagaga-
730 cagGCAMGMAGTTTYGCNGATATWGG-3') and the reverse primer brlong2 (5'-
731 gtctcgtgggctcggagatgtgtataagagacagTCTTCDSYCCATTCVTCRAACCA-3') were used result-
732 ing in an amplicon size of 350-500 bp (median 422 bp) plus the 67 bp overhang (small type).
733 Template cDNA (1 µl) was mixed with 5x KAPA HiFi buffer, 0.3 µM of each dNTP, 2 µM of
734 each primer and 0.5 U KAPA HiFi Hotstart Polymerase in a final reaction volume of 25 µl.
735 Reaction mixtures were incubated at 95°C for 2 min and then subjected to 35 cycles of 98°C
736 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
737 first 5 cycles cooling from denaturation temperature is performed to 65°C at a maximal ramp
738 of 3°C per second, then cooled to 54°C with a 0.5°C per second ramp. Heating from annealing
739 temperature to elongation temperature was performed with 1°C per second, all other steps
740 with a ramp of 3°C per second. Agarose gel images taken afterwards showed clean amplicons.
741 The DBLα-tag primers contain an overhang, which was used to conduct a second indexing
742 PCR reaction using sample-specific indexing primers as described in Nag *et al.* (Nag *et al*,
743 2017). The overhang sequence also serves as annealing site for Illumina sequencing primers
744 and indexing primers include individual 8-base combinations and adapter sequences that will
745 allow the final PCR product to bind in MiSeq Illumina sequencing flow cells. Indexing PCR
746 reactions were performed with a final primer concentration of 0.065 µM and 1 µl of first PCR
747 amplicon in a final volume of 20 µl; and by following steps: Heat activation at 95 °C, 15 min, 20
748 cycles of 95 °C for 20 s, 60 °C for 1 min and 72 °C for 1 min, and one final elongation step at
749 72 °C for 10 min. Indexing PCR amplicons were pooled (4 µl of each) and purified using AM-
750 Pure XP beads (Beckman Coulter, California, United States) according to manufacturer's pro-
751 tocol, using 200 µl pooled PCR product and 0.6 x PCR-pool volume of beads, to eliminate
752 primer dimers. The purified PCR pool were analyzed on agarose gels and Agilent 2100 Bio-
753 analyser to verify elimination of primer dimers, and correct amplicon sizes. Concentration of
754 purified PCR pools was measured by Nanodrop2000 (Thermo Fisher Scientific, Waltham, MA,
755 USA) and an aliquot adjusted to 4 nM concentration was pooled with other unrelated DNA
756 material and added to an Illumina MiSeq instrument for paired end 300 bp reads using a MiSeq
757 v3 flow cell.

758

759 ***DBLα-tag sequence analysis***

760 The paired-end DBLα-tag sequences were identified and partitioned into correct sample origin
761 based on unique index sequences. Each indexed raw sequence-pair were then processed
762 through the *Galaxy* webtool (usegalaxy.eu). Read quality checks was first performed with
763 *FastQC* to ensure a good NGS run (sufficient base quality, read length, duplication etc.). Next,

764 the sequences were trimmed by the Trimmomatic application, with a four base sliding window
765 approach and a *Phred* quality score above 20 to ensure high sequence quality output. The
766 trimmed sequences were then paired and converted, following analysis using the Varia tool for
767 quantification and prediction of the domain composition of the full-length *var* sequences from
768 which the DBLα-tag originated (Mackenzie *et al*, 2020). In brief, Varia clusters DBLα-tags with
769 99% sequence identity using *Vsearch* program (v2.14.2), and each unique tag is used to
770 search a database consisting of roughly 235,000 annotated *var* genes for near identical *var*
771 sequences (95% identity over 200 nucleotides). The domain composition of all "hit" sequences
772 is checked for conflicting annotations and the most likely domain composition is retuned. The
773 tool validation indicated prediction of correct domain compositions for around 85% of randomly
774 selected *var* tags, with higher hit rate and accuracy of the N-terminal domains. An average of
775 2,223.70 reads per patient sample was obtained and clusters consisting of less than 10 reads
776 were excluded from the analysis. The raw Varia output file is given in Table S10. The propor-
777 tion of transcripts encoding a given *Pf*EMP1 domain type or subtype was calculated for each
778 patient. These expression levels were used to first test the hypothesis that N-terminal domain
779 types associated with EPCR are found more frequently in first-time infections or upon severity
780 of disease, while those associated with non-EPCR binding were associated with pre-exposed
781 or mild cases. Secondly, quantile regression was used to calculate median differences (with
782 95%-confidence intervals) in expression levels for all main domain classes and subtypes be-
783 tween severity and exposure groups. All analyses were done with R (4.02) using the package
784 quantreg (5.73) for quantile regression.
785 For the comparison of both approaches, DBLα-tag sequencing and RNA-seq, only RNA-seq
786 contigs spanning the whole DBLα-tag region were considered. All conserved variants, the sub-
787 families *var1*, *var2csa* and *var3*, detected by RNA-seq were omitted form analysis since they
788 were not properly amplified by the DBLα-tag primers. To scan for the occurrence of DBLα-tag
789 sequences within the contigs assembled from the RNA-seq data we applied BLAST (basic
790 local alignment search tool) v2.9.0 software (Altschul *et al*, 1990). Therefore, we created a
791 BLAST database from the RNA-seq assemblies and screened for the occurrence of those
792 DBLα-tag sequence with more than 97% percent sequence identity using the "megablast op-
793 tion".
794 Calculation of the proportion of RNA-seq data covered by DBLα-tag was done with the upper
795 75th percentile based on total RPKM values determined for each patient. Vice versa, only
796 DBLα-tag clusters with more than 10 reads were considered and percent coverage of reads
797 and clusters calculated for each individual patient.
798 For all samples the agreement between the two molecular methods DBLα-tag sequencing and
799 RNA-seq was analyzed with a Bland-Altman plot, each individually and summarized. The ratio

between %-transformed measurements are plotted on the y-axis and the mean of the respective DBLα-tag and RNA-seq results are plotted on the x-axis. The bias and the 95% limits of agreement were calculated using GraphPad Prism 8.4.2.

Acknowledgments

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

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

Bachmann A, Bruske E, Krumkamp R, Turner L, Wichers JS, Petter M, Held J, Duffy MF, Sim BKL, Hoffman SL, *et al* (2019) Controlled human malaria infection with Plasmodium

837 falciparum demonstrates impact of naturally acquired immunity on virulence gene
838 expression. *PLOS Pathog* 15: e1007906
839 Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and
840 Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 57: 289–300
841 Berger SS, Turner L, Wang CW, Petersen JEV, Kraft M, Lusingu JPA, Mmbando B, Marquard
842 AM, Bengtsson DBAC, Hviid L, *et al* (2013) Plasmodium falciparum Expressing Domain
843 Cassette 5 Type PfEMP1 (DC5-PfEMP1) Bind PECAM1. *PLoS One* 8: 69117
844 Bernabeu M, Danziger SA, Avril M, Vaz M, Babar PH, Brazier AJ, Herricks T, Maki JN, Pereira
845 L, Mascarenhas A, *et al* (2016) Severe adult malaria is associated with specific PfEMP1
846 adhesion types and high parasite biomass. *Proc Natl Acad Sci U S A* 113: E3270–E3279
847 Bernabeu M, Gunnarsson C, Vishnyakova M, Howard CC, Nagao RJ, Avril M, Taylor TE,
848 Seydel KB, Zheng Y & Smith JD (2019) Binding Heterogeneity of Plasmodium falciparum
849 to Engineered 3D Brain Microvessels Is Mediated by EPCR and ICAM-1. *MBio* 10:
850 e00420-19
851 Borrmann S (2020) Mapping of safe and early chemo-attenuated live Plasmodium falciparum
852 immunization identifies immune signature of vaccine efficacy. *bioRxiv*:
853 2020.09.14.296152
854 Carlson J, Helmby H, Wahlgren M, Carlson J, Helmby H, Wahlgren M, Hill A, Brewster D &
855 Greenwood BM (1990) Human cerebral malaria: association with erythrocyte rosetting
856 and lack of anti-rosetting antibodies. *Lancet* 336: 1457–1460
857 Carver T, Harris SR, Berriman M, Parkhill J & McQuillan JA (2012) Artemis: An integrated
858 platform for visualization and analysis of high-throughput sequence-based experimental
859 data. *Bioinformatics* 28: 464–469
860 Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, Lavstsen T, Hviid L,
861 Duffy PE & Theander TG (2010) Hierarchical, domain type-specific acquisition of
862 antibodies to Plasmodium falciparum erythrocyte membrane protein 1 in Tanzanian
863 children. *Infect Immun* 78: 4653–4659
864 Cham GKK, Turner L, Lusingu J, Vestergaard L, Mmbando BP, Kurtis JD, Jensen ATR, Salanti
865 A, Lavstsen T & Theander TG (2009) Sequential, Ordered Acquisition of Antibodies to
866 Plasmodium falciparum Erythrocyte Membrane Protein 1 Domains . *J Immunol* 183:
867 3356–3363
868 Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, Bull PC, Mok S, Gupta
869 AP, Wang CW, *et al* (2012) A subset of group A-like var genes encodes the malaria
870 parasite ligands for binding to human brain endothelial cells. *Proc Natl Acad Sci U S A*
871 109: E1772
872 Dasanna AK, Lansche C, Lanzer M & Schwarz US (2017) Rolling Adhesion of Schizont Stage
873 Malaria-Infected Red Blood Cells in Shear Flow. *Biophys J* 112: 1908–1919

874 Davidson NM & Oshlack A (2014) Corset: Enabling differential gene expression analysis for
875 de novo assembled transcriptomes. *Genome Biol* 15: 410
876 Dent AE, Malhotra I, Wang X, Babineau D, Yeo KT, Anderson T, Kimmel RJ, Angov E, Lanar
877 DE, Narum D, *et al* (2016) Contrasting Patterns of Serologic and Functional Antibody
878 Dynamics to Plasmodium falciparum Antigens in a Kenyan Birth Cohort. *Clin Vaccine*
879 *Immunol* 23: 104–116
880 Doolan DL, Mu Y, Unal B, Sundaresh S, Hirst S, Valdez C, Randall A, Molina D, Liang X,
881 Freilich DA, *et al* (2008) Profiling humoral immune responses to P. falciparum infection
882 with protein microarrays. *Proteomics* 8: 4680–4694
883 Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, Malima R,
884 Lusingu J, Manjurano A, Nkya WMM, *et al* (2005) Estimating medium- and long-term
885 trends in malaria transmission by using serological markers of malaria exposure. *Proc*
886 *Natl Acad Sci U S A* 102: 5108–5113
887 Duffy F, Bernabeu M, Babar PH, Kessler A, Wang CW, Vaz M, Chery L, Mandala WL,
888 Rogerson SJ, Taylor TE, *et al* (2019) Meta-analysis of Plasmodium falciparum var
889 Signatures Contributing to Severe Malaria in African Children and Indian Adults. *MBio* 10:
890 e00217-19
891 Duffy MF, Brown G V., Basuki W, Krejany EO, Noviyanti R, Cowman AF & Reeder JC (2002)
892 Transcription of multiple var genes by individual, trophozoite-stage Plasmodium
893 falciparum cells expressing a chondroitin sulphate A binding phenotype. *Mol Microbiol* 43:
894 1285–1293
895 Duffy MF, Caragounis A, Noviyanti R, Kyriacou HM, Choong EK, Boysen K, Healer J, Rowe
896 JA, Molyneux ME, Brown G V., *et al* (2006) Transcribed var genes associated with
897 placental malaria in Malawian women. *Infect Immun* 74: 4875–4883
898 Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7: 1002195
899 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
900 26: 2460–2461
901 Falk N, Kaestli M, Qi W, Ott M, Baea K, Cortés A & Beck HP (2009) Analysis of plasmodium
902 falciparum var genes expressed in children from Papua New Guinea. *J Infect Dis* 200:
903 347–356
904 Felgner PL, Roestenberg M, Liang L, Hung C, Jain A, Pablo J, Nakajima-Sasaki R, Molina D,
905 Teelen K, Hermsen CC, *et al* (2013) Pre-erythrocytic antibody profiles induced by
906 controlled human malaria infections in healthy volunteers under chloroquine prophylaxis.
907 *Sci Rep* 3
908 Gagnon-Bartsch JA & Speed TP (2012) Using control genes to correct for unwanted variation
909 in microarray data. *Biostatistics* 13: 539–552
910 Ghumra A, Semblat J-P, Ataide R, Kifude C, Adams Y, Claessens A, Anong DN, Bull PC,

911 Fennell C, Arman M, *et al* (2012) Induction of Strain-Transcending Antibodies Against
912 Group A PfEMP1 Surface Antigens from Virulent Malaria Parasites. *PLoS Pathog* 8:
913 e1002665
914 Gustavsen JA, Pai S, Isserlin R, Demchak B & Pico AR (2019) RCy3: Network biology using
915 Cytoscape from within R. *F1000Research* 8: 1774
916 Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, Mayanja-Kizza H,
917 Ssewanyana I, Kamya MR, Beeson JG, *et al* (2015) Novel serologic biomarkers provide
918 accurate estimates of recent Plasmodium falciparum exposure for individuals and
919 communities. *Proc Natl Acad Sci U S A* 112: E4438–E4447
920 Herricks T, Avril M, Janes J, Smith JD & Rathod PK (2013) Clonal variants of Plasmodium
921 falciparum exhibit a narrow range of rolling velocities to host receptor CD36 under
922 dynamic flow conditions. *Eukaryot Cell* 12: 1490–1498
923 Higgins MK & Carrington M (2014) Sequence variation and structural conservation allows
924 development of novel function and immune evasion in parasite surface protein families.
925 *Protein Sci* 23: 354–365
926 Van Den Hoogen LL, Walk J, Oulton T, Reuling IJ, Reiling L, Beeson JG, Coppel RL, Singh
927 SK, Draper SJ, Bousema T, *et al* (2019) Antibody responses to antigenic targets of recent
928 exposure are associated with low-density parasitemia in controlled human plasmodium
929 falciparuminfections. *Front Microbiol* 10
930 Hsieh FL, Turner L, Bolla JR, Robinson C V., Lavstsen T & Higgins MK (2016) The structural
931 basis for CD36 binding by the malaria parasite. *Nat Commun* 7
932 Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9: 868–
933 877
934 Janes JH, Wang CP, Levin-Edens E, Vigan-Womas I, Guillotte M, Melcher M, Mercereau-
935 Puijalon O & Smith JD (2011) Investigating the host binding signature on the Plasmodium
936 falciparum PfEMP1 protein family. *PLoS Pathog* 7
937 Jensen ATR, Magistrado P, Sharp S, Joergensen L, Lavstsen T, Chiucchiuini A, Salanti A,
938 Vestergaard LS, Lusingu JP, Hermsen R, *et al* (2004) Plasmodium falciparum Associated
939 with Severe Childhood Malaria Preferentially Expresses PfEMP1 Encoded by Group A
940 var Genes. *J Exp Med* 199: 1179–1190
941 Jespersen JS, Wang CW, Mkumbaye SI, Minja DT, Petersen B, Turner L, Petersen JE,
942 Lusingu JP, Theander TG & Lavstsen T (2016) *Plasmodium falciparum var* genes
943 expressed in children with severe malaria encode <scp>CIDR</scp> α1 domains. *EMBO*
944 *Mol Med* 8: 839–850
945 Kapelski S, Klockenbring T, Fischer R, Barth S & Fendel R (2014) Assessment of the
946 neutrophilic antibody-dependent respiratory burst (ADRB) response to Plasmodium
947 falciparum . *J Leukoc Biol* 96: 1131–1142

948 Kessler A, Dankwa S, Bernabeu M, Harawa V, Danziger SA, Duffy F, Kampondeni SD,
949 Potchen MJ, Dambrauskas N, Vigdorovich V, *et al* (2017) Linking EPCR-Binding PfEMP1
950 to Brain Swelling in Pediatric Cerebral Malaria. *Cell Host Microbe* 22: 601-614.e5
951 Kirchgatter K & Del Portillo HA (2002) Association of severe noncerebral Plasmodium
952 falciparum malaria in Brazil with expressed PfEMP1 DBL1α sequences lacking cysteine
953 residues. *Mol Med* 8: 16–23
954 Kivisi CA, Muthui M, Hunt M, Fegan G, Otto TD, Githinji G, Warimwe GM, Rance R, Marsh K,
955 Bull PC, *et al* (2019) Exploring Plasmodium falciparum Var Gene Expression to Assess
956 Host Selection Pressure on Parasites During Infancy. *Front Immunol* 10: 2328
957 Kolberg L, Raudvere U, Kuzmin I, Vilo J & Peterson H (2020) gprofiler2 -- an R package for
958 gene list functional enrichment analysis and namespace conversion toolset g:Profiler.
959 *F1000Research* 9: 709
960 Kraemer SM & Smith JD (2003) Evidence for the importance of genetic structuring to the
961 structural and functional specialization of the Plasmodium falciparum var gene family. *Mol*
962 *Microbiol* 50: 1527–1538
963 Kyes SA, Christodoulou Z, Raza A, Horrocks P, Pinches R, Rowe JA & Newbold CI (2003) A
964 well-conserved Plasmodium falciparum var gene shows an unusual stage-specific
965 transcript pattern. *Mol Microbiol* 48: 1339–1348
966 Kyes SA, Kraemer SM & Smith JD (2007) Antigenic variation in Plasmodium falciparum: Gene
967 organization and regulation of the var multigene family. *Eukaryot Cell* 6: 1511–1520
968 doi:10.1128/EC.00173-07 [PREPRINT]
969 Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, Koné AK, Doumbo OK,
970 Plowe C V. & Rowe JA (2006) Differential var gene transcription in Plasmodium
971 falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia.
972 *Mol Biochem Parasitol* 150: 211–218
973 Lau CKY, Turner L, Jespersen JS, Lowe ED, Petersen B, Wang CW, Petersen JEV, Lusingu
974 J, Theander TG, Lavstsen T, *et al* (2015) Structural conservation despite huge sequence
975 diversity allows EPCR binding by the pfemp1 family implicated in severe childhood
976 malaria. *Cell Host Microbe* 17: 118–129
977 Lavstsen T, Salanti A, Jensen ATR, Arnot DE & Theander TG (2003) Sub-grouping of
978 Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-
979 coding regions. *Malar J* 2: 1–14
980 Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, Wang CW, Berger SS,
981 Baraka V, Marquard AM, *et al* (2012) Plasmodium falciparum erythrocyte membrane
982 protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc*
983 *Natl Acad Sci U S A* 109: E1791–E1800
984 Law CW, Chen Y, Shi W & Smyth GK (2014) Voom: Precision weights unlock linear model

985 analysis tools for RNA-seq read counts. *Genome Biol* 15
986 Lennartz F, Adams Y, Bengtsson A, Olsen RW, Turner L, Ndam NT, Ecklu-Mensah G,
987 Moussiliou A, Ofori MF, Gamain B, *et al* (2017) Structure-Guided Identification of a Family
988 of Dual Receptor-Binding PfEMP1 that Is Associated with Cerebral Malaria. *Cell Host*
989 *Microbe* 21: 403–414
990 Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
991 Liao Y, Smyth GK & Shi W (2013) The Subread aligner: Fast, accurate and scalable read
992 mapping by seed-and-vote. *Nucleic Acids Res* 41
993 Liao Y, Smyth GK & Shi W (2014) FeatureCounts: An efficient general purpose program for
994 assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930
995 Llewellyn D, Miura K, Fay MP, Williams AR, Murungi LM, Shi J, Hodgson SH, Douglas AD,
996 Osier FH, Fairhurst RM, *et al* (2015) Standardization of the antibody-dependent
997 respiratory burst assay with human neutrophils and Plasmodium falciparum malaria. *Sci*
998 *Rep* 5
999 López-Barragán MJ, Lemieux J, Quiñones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-
1000 Mury C, Zhao K & Su X (2011) Directional gene expression and antisense transcripts in
1001 sexual and asexual stages of Plasmodium falciparum. *BMC Genomics* 12: 587
1002 Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for
1003 RNA-seq data with DESeq2. *Genome Biol* 15: 550
1004 Lubiana P, Bouws P, Roth LK, Dörpinghaus M, Rehn T, Brehmer J, Wichers JS, Bachmann
1005 A, Höhn K, Roeder T, *et al* (2020) Adhesion between P. falciparum infected erythrocytes
1006 and human endothelial receptors follows alternative binding dynamics under flow and
1007 febrile conditions. *Sci Rep* 10: 4548
1008 Mackenzie G, Jensen RW, Lavstsen T & Otto TD (2020) Varia: Prediction, analysis and
1009 visualisation of variable genes. *bioRxiv*: 2020.12.15.422815 [PREPRINT]
1010 Magallón-Tejada A, Machevo S, Cisteró P, Lavstsen T, Aide P, Rubio M, Jiménez A, Turner
1011 L, Valmaseda A, Gupta H, *et al* (2016) Cytoadhesion to gC1qR through Plasmodium
1012 falciparum Erythrocyte Membrane Protein 1 in Severe Malaria. *PLOS Pathog* 12:
1013 e1006011
1014 McGee M & Chen Z (2006) Parameter estimation for the exponential-normal convolution model
1015 for background correction of affymetrix GeneChip data. *Stat Appl Genet Mol Biol* 5
1016 McQuaid F & Rowe JA (2020) Rosetting revisited: A critical look at the evidence for host
1017 erythrocyte receptors in Plasmodium falciparum rosetting. *Parasitology* 147: 1–11
1018 doi:10.1017/S0031182019001288 [PREPRINT]
1019 Merico D, Isserlin R, Stueker O, Emili A & Bader GD (2010) Enrichment Map: A Network-
1020 Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* 5:
1021 e13984

1022 Mkumbaye SI, Wang CW, Lyimo E, Jespersen JS, Manjurano A, Mosha J, Kavishe RA,
1023 Mwakalinga SB, Minja DTR, Lusingu JP, *et al* (2017) The severity of *Plasmodium*
1024 *falciparum* infection is associated with transcript levels of *var* genes encoding EPCR-
1025 binding PfEMP1. *Infect Immun*: IAI.00841-16
1026 Nag S, Dalgaard MD, Kofoed PE, Ursing J, Crespo M, Andersen LOB, Aarestrup FM, Lund O
1027 & Alifrangis M (2017) High throughput resistance profiling of Plasmodium falciparum
1028 infections based on custom dual indexing and Illumina next generation sequencing-
1029 technology. *Sci Rep* 7: 1–13
1030 Obeng-Adjei N, Larremore DB, Turner L, Ongoiba A, Li S, Doumbo S, Yazew TB, Kayentao
1031 K, Miller LH, Traore B, *et al* (2020) Longitudinal analysis of naturally acquired PfEMP1
1032 CIDR domain variant antibodies identifies associations with malaria protection. *JCI Insight*
1033 Obiero JM, Campo JJ, Scholzen A, Randall A, Bijker EM, Roestenberg M, Hermsen CC, Teng
1034 A, Jain A, Davies DH, *et al* (2019) Antibody Biomarkers Associated with Sterile Protection
1035 Induced by Controlled Human Malaria Infection under Chloroquine Prophylaxis. *mSphere*
1036 4
1037 Otto TD (2019) ThomasDOtto/varDB: First release of varDB.
1038 Otto TD, Assefa SA, Böhme U, Sanders MJ, Kwiatkowski D, Berriman M, Newbold C &
1039 Newbold C (2019) Evolutionary analysis of the most polymorphic gene family in
1040 falciparum malaria. *Wellcome Open Res* 4: 193
1041 Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, Bull PC, Pearson RD, Abdi A,
1042 Dimonte S, *et al* (2018a) Long read assemblies of geographically dispersed Plasmodium
1043 falciparum isolates reveal highly structured subtelomeres. *Wellcome open Res* 3: 52
1044 Otto TD, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, Oyola SO, Okouga AP,
1045 Boundenga L, Willaume E, *et al* (2018b) Genomes of all known members of a
1046 Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol* 3: 687–697
1047 Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C (2017) Salmon provides fast and bias-
1048 aware quantification of transcript expression. *Nat Methods* 14: 417–419
1049 Rask TS, Hansen DA, Theander TG, Gorm Pedersen A & Lavstsen T (2010) Plasmodium
1050 falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and
1051 Conquer. *PLoS Comput Biol* 6: e1000933
1052 Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H & Vilo J (2019) G:Profiler: A
1053 web server for functional enrichment analysis and conversions of gene lists (2019
1054 update). *Nucleic Acids Res* 47: W191–W198
1055 Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, Sarthou JL & Mercereau-
1056 Puijalon O (1996) Extensive genetic diversity of Plasmodium falciparum isolates collected
1057 from patients with severe malaria in Dakar, Senegal. *Trans R Soc Trop Med Hyg* 90: 704–
1058 711

1059 Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen ATR, Müller D, Theander T & Beck
1060 HP (2006) Differential expression of var gene groups is associated with morbidity caused
1061 by Plasmodium falciparum infection in Tanzanian children. *Infect Immun* 74: 3904–3911
1062 Rowe JA, Claessens A, Corrigan RA & Arman M (2009) Adhesion of Plasmodium falciparum-
1063 infected erythrocytes to human cells: Molecular mechanisms and therapeutic
1064 implications. *Expert Rev Mol Med* 11: e16 doi:10.1016/0925-4005(95)85135-6
1065 [PREPRINT]
1066 Rowe JA, Obiero J, Marsh K & Raza A (2002) Short report: Positive correlation between
1067 rosetting and parasitemia in Plasmodium falciparum clinical isolates. *Am J Trop Med Hyg*
1068 66: 458–460
1069 Salanti A, Dahlbäck M, Turner L, Nielsen MA, Barfod L, Magistrado P, Jensen ATR, Lavstsen
1070 T, Ofori MF, Marsh K, *et al* (2004) Evidence for the involvement of VAR2CSA in
1071 pregnancy-associated malaria. *J Exp Med* 200: 1197–1203
1072 Saul A (1999) The role of variant surface antigens on malaria-infected red blood cells. *Parasitol*
1073 *Today* 15: 455–7
1074 Shabani E, Hanisch B, Opoka RO, Lavstsen T & John CC (2017) Plasmodium falciparum
1075 EPCR-binding PfEMP1 expression increases with malaria disease severity and is
1076 elevated in retinopathy negative cerebral malaria. *BMC Med* 15: 183
1077 Silver JD, Ritchie ME & Smyth GK (2009) Microarray background correction: Maximum
1078 likelihood estimation for the normal-exponential convolution. *Biostatistics* 10: 352–363
1079 Smyth GK (2005) limma: Linear Models for Microarray Data. In *Bioinformatics and*
1080 *Computational Biology Solutions Using R and Bioconductor* pp 397–420. Springer-Verlag
1081 Storm J, Jespersen JS, Seydel KB, Szestak T, Mbewe M, Chisala N V, Phula P, Wang CW,
1082 Taylor TE, Moxon CA, *et al* (2019) Cerebral malaria is associated with differential
1083 cytoadherence to brain endothelial cells. *EMBO Mol Med* 11
1084 Taghavian O, Jain A, Joyner CJ, Ketchum S, Nakajima R, Jasinskas A, Liang L, Fong R, King
1085 C, Greenhouse B, *et al* (2018) Antibody Profiling by Proteome Microarray with Multiplex
1086 Isotype Detection Reveals Overlap between Human and Aotus nancymaae Controlled
1087 Malaria Infections. *Proteomics* 18
1088 Tenenbaum D (2020) KEGGREST: Client-side REST access to KEGG. R package version
1089 1.30.0. [PREPRINT]
1090 Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, Marfurt J,
1091 Cobbold SA, Rambhatla JS, McConville MJ, *et al* (2018) The Plasmodium falciparum
1092 transcriptome in severe malaria reveals altered expression of genes involved in important
1093 processes including surface antigen–encoding var genes. *PLoS Biol* 16: e2004328
1094 Turewicz M, Ahrens M, May C, Marcus K & Eisenacher M (2016) PAA: An R/bioconductor
1095 package for biomarker discovery with protein microarrays. *Bioinformatics* 32: 1577–1579

1096 Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JE V., Avril M, Brazier AJ, Freeth J,
1097 Jespersen JS, Nielsen MA, *et al* (2013) Severe malaria is associated with parasite binding
1098 to endothelial protein C receptor. *Nature* 498: 502–505
1099 Turner L, Lavstsen T, Mmbando BP, Wang CW, Magistrado PA, Vestergaard LS, Ishengoma
1100 DS, Minja DTR, Lusingu JP & Theander TG (2015) IgG antibodies to endothelial protein
1101 C receptor-binding cysteine-rich interdomain region domains of Plasmodium falciparum
1102 erythrocyte membrane protein 1 are acquired early in life in individuals exposed to
1103 malaria. *Infect Immun* 83: 3096–3103
1104 Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, Bouzek H, Moon W, Babak
1105 T, Fried M, *et al* (2011) NSR-seq transcriptional profiling enables identification of a gene
1106 signature of Plasmodium falciparum parasites infecting children. *J Clin Invest* 121: 1119–
1107 1129
1108 Warimwe GM, Keane TM, Fegan G, Musyoki JN, Newton CRJC, Pain A, Berriman M, Marsh
1109 K & Bull PC (2009) Plasmodium falciparum var gene expression is modified by host
1110 immunity. *Proc Natl Acad Sci U S A* 106: 21801–21806
1111 WHO (2020) WHO World Malaria Report 2020
1112 Wickham H (2016) Ggplot2 : elegant graphics for data analysis Springer
1113 Winter G, Chen Q, Flick K, Kremsner P, Fernandez V & Wahlgren M (2003) The 3D7var5.2
1114 (varCOMMON) type var gene family is commonly expressed in non-placental
1115 Plasmodium falciparum malaria. *Mol Biochem Parasitol* 127: 179–191
1116 Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, *et al* (2014)
1117 SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads.
1118 *Bioinformatics* 30: 1660–1666
1119 Zhao S, Guo Y SY (2020) KEGGprofile: An annotation and visualization package for multi-
1120 types and multi-groups expression data in KEGG pathway. R package version 1.32.0.
1121 [PREPRINT]
1122
1123 **Figure legends**
1124
1125 **Figure 1: Subgrouping of patients into first-time infected and pre-exposed individuals**
1126 **based on antibody levels against *P. falciparum*.** In order to further characterize the patient
1127 cohort plasma samples (n=32) were subjected to Luminex analysis with the *P. falciparum* an-
1128 tigens AMA1, MSP1, CSP known to induce a strong antibody response in humans. With ex-
1129 ception of patient #21 unsupervised clustering of the PCA-reduced data clearly discriminates
1130 between first-time infected (naïve) and pre-exposed patients with higher antibody levels
1131 against tested *P. falciparum* antigens and also assigns plasma samples from patients with
1132 unknown immune status into naïve and pre-exposed clusters (**A**). Classification of patient #21

1133 into the naïve subgroup was confirmed using different serological assays assessing antibody
1134 levels against *P. falciparum* on different levels: a merozoite-directed antibody-dependent res-
1135 piratory burst (mADRB) assay (Kapelski *et al*, 2014) (**B**), a PEMS-specific ELISA (**C**) and a
1136 262-feature protein microarray covering 228 well-known *P. falciparum* antigens detecting re-
1137 activity with individual antigens and the antibody breadth of IgG (upper panel) and IgM (lower
1138 panel) (**D**). The boxes represent medians with IQR; the whiskers depict minimum and maxi-
1139 mum values (range) with outliers located outside the whiskers. Serological assays revealed
1140 significant differences between patient groups (Mann Whitney U test). Reactivity of patient
1141 plasma IgG and IgM with individual antigens in the protein microarray is presented as volcano
1142 plot highlighting the significant hits in red. Box plots represent antibody breadths by summa-
1143 rizing the number of recognized antigens out of 262 features tested. Data from all assays were
1144 used for an unsupervised random forest approach (**E**). The variable importance plot of the
1145 random forest model shows the decrease in prediction accuracy if values of a variable are
1146 permuted randomly. The decrease in accuracy was determined for each serological assay
1147 indicating that the mADRB, ELISA and Luminex assays are most relevant in the prediction of
1148 patient clusters (**F**). Patients with known immune status based on medical reports were marked
1149 in all plots with filled circles in blue (naïve) and grey (pre-exposed), samples from patients with
1150 unknown immune status are shown as open circles. ELISA: Enzyme-linked immunosorbent
1151 assay, IQR: interquartile range, PCA: Principal component analysis
1152
1153 **Figure 2: Estimated stage proportions for each sample.** Patient samples consist of a com-
1154 bination of different parasite stages. To estimate the proportion of different life cycle stages in
1155 each sample a constrained linear model was fit using data from López-Barragán *et al.* (López-
1156 Barragán *et al*, 2011). The proportions of rings (8 hpi), early trophozoites (19 hpi), late tropho-
1157 zoites (30 hpi), schizonts (42 hpi) and gametocytes stages shown in the columns of the bar
1158 plots must add to 1 for each sample. Shown are the comparisons between first-time infected
1159 (naïve; blue) and pre-exposed samples (grey) (**A**) and severe (red) and non-severe cases
1160 (grey) (**B**). A bias towards the early trophozoite appears in the non-severe malaria sample
1161 group, which was confirmed by calculating the age in hours post infection (hpi) for each para-
1162 site sample. The boxes represent medians with IQR; the whiskers depict minimum and maxi-
1163 mum values (range) with outliers located outside the whiskers (**C**, **D**). IQR: interquartile range
1164
1165 **Figure 3: Differential all gene expression analysis.** Gene set enrichment analysis (GSEA)
1166 of GO terms and KEGG pathways indicate gene sets deregulated in first-time infected malaria
1167 patients. GO terms related to antigenic variation and host cell remodeling are significantly
1168 down-regulated, only the KEGG pathway 03410 'base excision repair' shows a significant up-
1169 regulation in malaria-naïve patients (**A**). Log fold changes (logFC) for the 15 *P. falciparum*

genes assigned to the KEGG pathway 03410 'base excision repair' are plotted with the six significant hits marked with * for p<0.05 and ** for p<0.01 (**B**).

Figure 4: Analysis of RNA-seq data at the level of *var* gene transcripts using the separate assembly approach. RNA-seq reads of each patient sample were matched to *de novo* assembled *var* contigs with varying length, domain and homology block composition. Shown are significant differently expressed *var* gene contigs with an adjusted p-value of <0.05 in first-time infected (blue) and pre-exposed patient samples (grey) (**A**, **B**) as well as severe (red) and non-severe cases (grey) (**C**, **D**). Data are displayed as heatmaps showing expression levels in log transformed normalized Salmon read counts for each individual sample (**A**, **C**) or as box plot with median log transformed normalized Salmon read counts and interquartile range (IQR) for each group of samples (**B**, **D**). Normalized Salmon read counts for all assembled transcripts are available in Table S4.

Figure 5: Analysis of RNA-seq data via *de novo* assembly at the level of *var* gene domains. RNA-seq reads of each patient sample were matched to *de novo* assembled *var* contigs with varying length, domain and homology block composition. Shown are significantly differently expressed *Pf*EMP1 domain subfamilies from Rask *et al.* (Rask *et al*, 2010) with an adjusted p-value of <0.05 in first-time infected (blue) and pre-exposed patient samples (grey) (**A**, **B**) as well as severe (red) and non-severe cases (grey) (**C**, **D**) using HMMER3 models. The N-terminal head structure (NTS-DBLα□-CIDRα□β/γ/δ) confers a mutually exclusive binding phenotype either to EPCR-, CD36-, CSA- or an unknown receptor. Expression values of the N-terminal domains were summarized for each patient and differences in the distribution among patient groups were tested using the Mann-Whitney U test (**E**, **F**). Data are displayed as heatmaps showing expression levels in log transcripts per million (TPM) for each individual sample (**A**, **C**) or as box plot with median log TPM and interquartile range (IQR) for each group of samples (**B**, **D, E, F**).

Figure 6: Analysis of RNA-seq data via *de novo* assembly at the level of *var* gene homology blocks. RNA-seq reads of each patient sample were matched to *de novo* assembled *var* contigs with varying length, domain and homology block composition. Shown are significantly differently expressed homology blocks from Rask *et al.* (Rask *et al*, 2010) with an adjusted p-value of <0.05 in first-time infected (blue) and pre-exposed patient samples (grey) (**A**, **B**) as well as severe (red) and non-severe cases (grey) (**C**, **D**). Data are displayed as heatmaps showing expression levels in log transcripts per million (TPM) for each individual sample (**A**, **C**) or as box plot with median log TPM and interquartile range (IQR) for each group of samples (**B**, **D**).

1207

1208 **Figure 7: Verification of RNA-seq results using DBLα-tag sequencing.** Amplified DBLα-
1209 tag sequences were blasted against the ~2,400 genomes on varDB (Otto, 2019) to obtain
1210 subclassification into DBLα0/1/2 and prediction of adjacent head-structure NTS and CIDR do-
1211 mains and their related binding phenotype. Proportion of each NTS and DBLα subclass as well
1212 as CIDR domains grouped according to binding phenotype (CIDRα1.1/4-8: EPCR-binding,
1213 CIDRα2-6: CD36-binding, CIDRβ/γ/δ unknown binding phenotype/rosetting) was calculated
1214 and shown separately on the left, number of total reads and individual sequence cluster with
1215 n ≥10 sequences are shown on the right. Differences in the distribution among first-time in-
1216 fected (blue) and pre-exposed individuals (grey) (**A**) as well as severe (red) and non-severe
1217 cases (grey) (**B**) were tested using the Mann-Whitney U test. The boxes represent medians
1218 with IQR; the whiskers depict minimum and maximum values (range) with outliers located out-
1219 side the whiskers. Quantile regression was applied to look for differences between patient
1220 groups on the level of domain main classes (left) and subdomains (right). Shown are median
1221 differences with 95%-confidence intervals of domains with values unequal 0. Domains with
1222 positive values tend to be higher expressed in naïve (**C**) and severe patients (**D**).

1223

1224 **Figure 8: Summary of *Pf*EMP1 transcripts, domains, and homology blocks that were**
1225 **found more or less frequently in malaria-naïve and severely ill patients.** A schematic
1226 presentation of typical *Pf*EMP1 domain compositions (**A**). The N-terminal head structure con-
1227 fers mutually exclusive receptor binding phenotypes: EPCR (yellow: CIDRα1.1/4-8), CD36
1228 (salmon CIDRα2-6), CSA (VAR2CSA) and yet unknown phenotypes (orange: CIDRβ/γ/δ, red:
1229 CIDRα1.2/3 from VAR1, VAR3). Group A includes the conserved subfamilies VAR1 and VAR3,
1230 EPCR binding variants and those with unknown binding phenotypes sometimes associated
1231 with rosetting. Group B *Pf*EMP1 can have EPCR-binding capacities, but most variants share
1232 a four-domain structure with group C-type variants capable of CD36-binding. Dual binder can
1233 be found within group A and B with an DBLβ domain responsible for ICAM-1- (DBLβ1/3/5) or
1234 gC1qr-binding (DBLβ12). Inter-strain conserved tandem arrangements of domains, so called
1235 domain cassettes (DC), can be found within all groups as selectively indicated.
1236 Transcripts, domains and homology blocks according to Rask *et al.* (Rask *et al*, 2010) found
1237 significant differently expressed (p-value <0.05) between patient groups of both comparisons:
1238 first-time infected (blue) versus pre-exposed (black) cases and severe (red) versus non-severe
1239 (black) cases (**B**).
1240 ATS: acidic terminal sequence, CIDR: cysteine-rich interdomain region, CSA: chondroitin sul-
1241 phate A, DBL: Duffy binding-like, DC: domain cassette, EPCR: endothelial protein C receptor,
1242 gC1qr: receptor for complement component C1q, ICAM-1: intercellular adhesion molecule 1,
1243 NTS: N-terminal segment, TM: transmembrane domain

1244

1245 **Figure 9: Correlation of *var* gene expression with antibody levels against head structure**
1246 **CIDR domains.** Patient plasma samples (n=32) were subjected to Luminex analysis with 35
1247 *Pf*EMP1 head structure CIDR domains. The panel includes EPCR-binding CIDRα1 domains
1248 (n = 19), CD36-binding CIDRα2–6 domains (n = 12) and CIDR domains with unknown binding
1249 phenotype (CIDRγ3: n = 1, CIDRδ1: n = 3) as well as the minimal binding region of VAR2CSA
1250 (VAR2). Box plots showing mean fluorescence intensities (MFI) extending from the 25th to the
1251 75th percentiles with a line at the median indicate higher reactivity of the pre-exposed (**A**) and
1252 non-severe cases (**B**) with all *Pf*EMP1 domains tested. Significant differences were observed
1253 for recognition of CIDRα2–6, CIDRδ1 and CIDRγ3; VAR2CSA recognition differed only be-
1254 tween severe and non-severe cases (Mann Whitney U test). Furthermore, the breadth of IgG
1255 recognition (%) of CIDR domains for the different patient groups was calculated and shown as
1256 a heat map (**C**).
1257
1258 **Table 1: Characteristics and classification of malaria patients.**
1259
1260 **Table 2: Patient groups data.**
1261
1262 **Table 3: *var* transcripts up- and down-regulated in first-time infected patients.**
1263
1264 **Table 4: *var* transcripts up- and down-regulated in severe cases.**
1265
1266 **Table 5: *var* domains defined by Rask *et al.* (Rask *et al*, 2010) up- and down-regulated**
1267 **in first-time infected patients.**
1268
1269 **Table 6: *var* domains defined by Rask *et al.* (Rask *et al*, 2010) up- and down-regulated**
1270 **in severe cases.**
1271
1272 **Table 7: *var* blocks defined by Rask *et al.* (Rask *et al*, 2010) up- and down-regulated in**
1273 **first-time infected patients.**
1274
1275 **Table 8: *var* blocks defined by Rask *et al.* (Rask *et al*, 2010) down-regulated in severe**
1276 **cases.**
1277
1278 **Supplement**
1279

1280 **Supplement figure 1: Early immune response in mild and severe malaria within the naïve**
1281 **patient cluster.** Antibody reactivity against individual antigens within the three subgroups 'na-
1282 ïve with mild symptoms', 'naïve with severe symptoms' and 'pre-exposed with mild symp-
1283 toms'. Sera from all volunteers were assessed on protein microarrays and data normalized to
1284 control spots containing no antigen (no DNA control spots). Median reactivity of the mild in-
1285 fected malaria-naïve, severely infected malaria-naïve as well as the mild infected with pre-
1286 exposure to malaria are represented as bar-charts. IgG data is given for all 262 *P. falciparum*
1287 proteins spotted on the microarray representing 228 unique antigens (**A**). To estimate differ-
1288 ences in immune response in mild and severe malaria within the malaria-naïve population,
1289 normalized IgG (**B**) and IgM (**C**) antibody responses were compared in the two subpopulations.
1290 Differentially recognized antigens ($p < 0.05$ and fold change >2) are depicted in red.
1291
1292 **Supplement figure 2: The base excision repair (KEGG:03410) in *P. falciparum*.**
1293 Orthologues present in *P. falciparum* are indicated by gene IDs, log fold changes (logFC) are
1294 indicated by color code (red: up-regulated, blue: down-regulated) (**A**). Summary of logFC in
1295 gene expression in first-time infected relative to pre-exposed patients and p-values for the
1296 logFC.
1297
1298 **Supplement figure 3: Differential expression of the *var1* variants 3D7 and IT and *var2csa***
1299 **between patient groups.** RNA-seq reads from each patient were normalized against the num-
1300 ber of mappable reads to the 3D7 genome and aligned to the *var1*-3D7 and *var1*-IT variants
1301 as well as *var2csa*. The resulting bigwig files were displayed in Artemis (Carver *et al*, 2012).
1302 Individual samples are colored according to the patient group: first-time infected in blue (A),
1303 severe in red (B) and the respective pre-exposed or non-severe samples in grey.
1304
1305 **Supplement figure 4: Comparison of the DBLα-tag sequencing with RNA-seq analysis.**
1306 DBLα-tag sequencing and RNA-seq data compared in Bland-Altman plots for all patients sum-
1307 marized (**A**) and for each individual patient (**B**), where the mean log expression of each gene
1308 is indicated on the X-axis and the log ratio between normalized DBLα-tag (% of reads) and
1309 RNA-seq values (% of RPKM from all contigs containing both DBLα-tag primer binding sites)
1310 on the y-axis. The mean (equal to bias) of all ratios (line) and the confidence interval (CI) of
1311 95% (dotted lines) are indicated. Data points with negative values for one of the approaches
1312 are displayed in dependence of their mean log expression on top (DBLα-tag sequence clusters
1313 not detected by RNA-seq) or bottom (RNA-seq contigs not found within DBLα-tag sequence
1314 cluster) of the graph.
1315

1316 **Supplement figure 5: RNA quality.** The Bioanalyzer automated RNA electrophoresis system
1317 was used to characterize the total RNA quality prior library synthesis. The calculated RIN value
1318 is provided, although this measurement is questionable for samples from mixed species. From
1319 the four rRNA peaks visible in all samples, the inner peaks represent *P. falciparum* 18S and
1320 28S rRNA, the outer peaks are of human origin.
1321
1322 **Data S1: Sequences of assembled *var* contigs from all patient isolates.**
1323
1324 **Supplement table 1: Data from Luminex, mADRB, ELISA and protein microarray.** Sero-
1325 prevalence of head-structure CIDR domains determined by applying a cut off from Danish
1326 controls (mean + 2 STD) to the Luminex data.
1327
1328 **Supplement table 2: Raw read counts by sample for *H. sapiens*, *P. falciparum*, *var* exon**
1329 **1 and percentage of reads that mapped either to *P. falciparum* or *var* exon 1 as well as**
1330 **the number of assembled *var* contigs >500 bp in length.**
1331
1332 **Supplement table 3: Differentially expressed genes excluding *var* genes (all gene anal-**
1333 **ysis) between first-time infected and pre-exposed patient samples.**
1334
1335 **Supplement table 4: Differentially expressed genes excluding *var* genes (all gene anal-**
1336 **ysis) between severe and non-severe patient samples.**
1337
1338 **Supplement table 5: Features of the assembled *var* fragments annotated in accordance**
1339 **with Rask *et al. .* (Rask *et al*, 2010) and Tonkin-Hill *et al.* (Tonkin-Hill *et al*, 2018).** The
1340 reading frame used for translation is given after the contig ID, the position of each annotation
1341 is provided by starting and ending amino acid followed by the p-value from the blast search
1342 against the respective database. For annotations in accordance with Tonkin-Hill *et al.* (Tonkin-
1343 Hill *et al*, 2018) either the short ID or 'NA' (not applicable) is listed at the end. Short IDs are
1344 only available for significant differently expressed domains and blocks between severe and
1345 non-severe cases (Tonkin-Hill *et al*, 2018).
1346
1347 **Supplement table 6: Summary of *var* gene fragments assembled for each patient isolate**
1348 **showing length, raw read counts, RPKM, blast hits, domain and block annotations in**
1349 **accordance with Rask *et al. .* (Rask *et al*, 2010).** The RPKM for the contigs was calculated
1350 as number of mapped reads and normalized by the number of mapped reads against all tran-
1351 script in each isolate, respectively. Therefore, RPKM expression values are only valid to com-
1352 pare within a single sample since RNA-seq reads were mapped only to the contigs of the

1353 respective patient isolate using BWA-MEM (Li, 2013). Further, the amount of blast hits with
1354 500 bp or 80% of overlap against the ~2400 samples from varDB (Otto, 2019) with an identity
1355 cutoff of 98%. Further hits of 1 kb (>98% identity) against the *var* genes from the 15 reference
1356 genomes (Otto *et al*, 2018a) are listed. The last two column show the annotations from Rask
1357 *et al.* (Rask *et al*, 2010) associated to each contig.
1358
1359 **Supplement table 7: Log transformed normalized Salmon read counts for assembled**
1360 ***var* transcripts, TPM for collapsed domains and homology blocks from each patient iso-**
1361 **late.** Normalized counts and TPM values calculated for transcripts, domains and blocks with
1362 expression in at least three patient isolates with more than five read counts.
1363
1364 **Supplement table 8: Differently expressed *var* transcripts, domains and homology**
1365 **blocks between first-time infected and pre-exposed patient samples.**
1366
1367 **Supplement table 9: Differently expressed *var* transcripts, domains and homology**
1368 **blocks between severe and non-severe patient samples.**
1369
1370 **Supplement table 10: Data from DBLα-tag sequencing.**

Figure 1

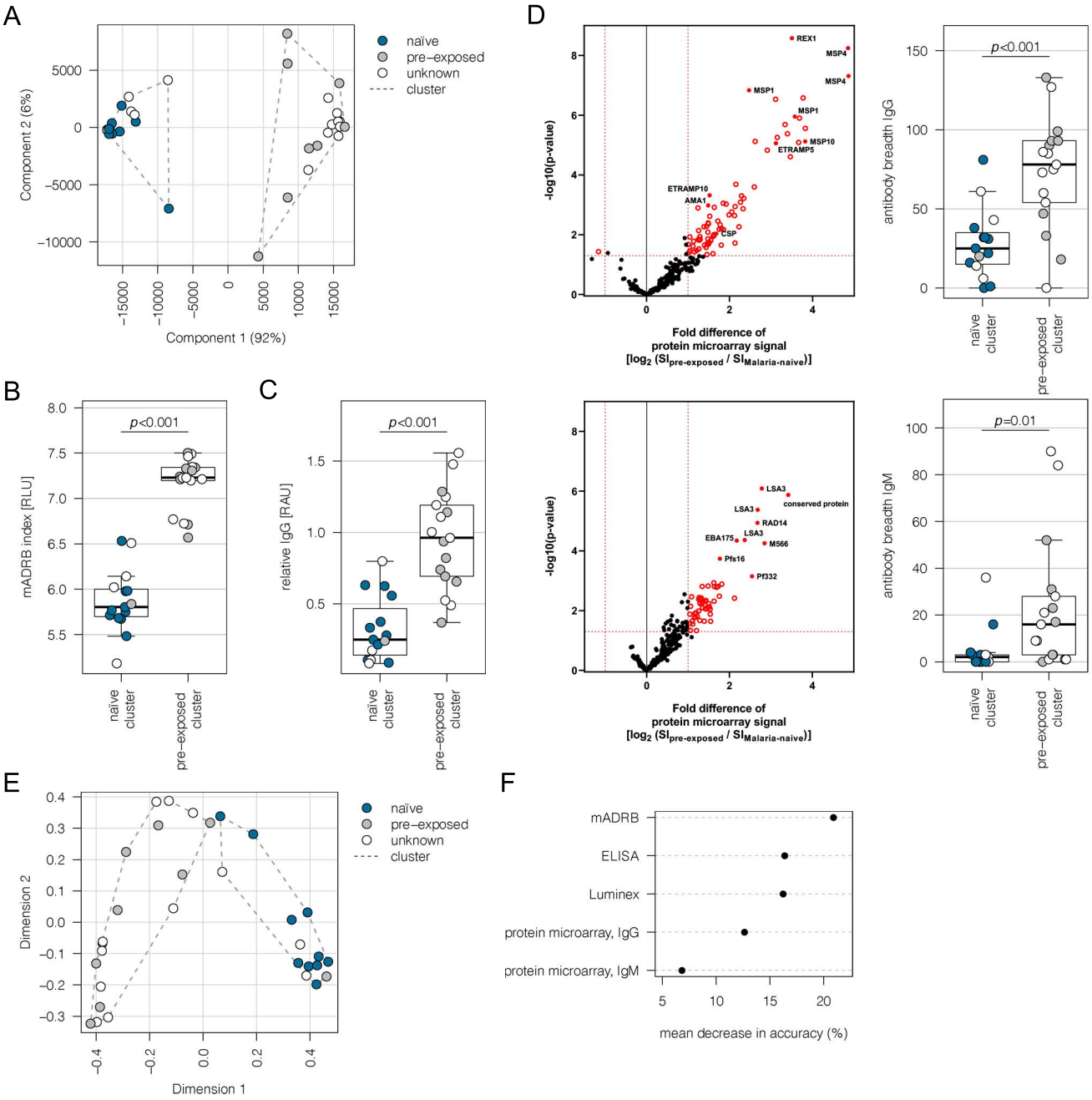

A
naïve
pre−exposed
unknown
cluster
Component 2 (6%)
Component 1 (92%)
B
mADRB index [RLU]
p<0.001
naïve cluster
pre−exposed cluster
C
relative IgG [RAU]
p<0.001
naïve cluster
pre−exposed cluster
D
-log10(p-value)
Fold difference of protein microarray signal [log2 (SIpre-exposed / SIMalaria-naive)]
REX1
MSP4
MSP4
MSP1
MSP1
MSP10
ETRAMP5
ETRAMP10
AMA1
CSP
antibody breadth IgG
p<0.001
naïve cluster
pre−exposed cluster
-log10(p-value)
Fold difference of protein microarray signal [log2 (SIpre-exposed / SIMalaria-naive)]
LSA3
conserved protein
LSA3
RAD14
EBA175
LSA3
M566
Pfs16
Pf332
antibody breadth IgM
p=0.01
naïve cluster
pre−exposed cluster
E
naïve
pre−exposed
unknown
cluster
Dimension 2
Dimension 1
F
mADRB
ELISA
Luminex
protein microarray, IgG
protein microarray, IgM
mean decrease in accuracy (%)

Figure 2

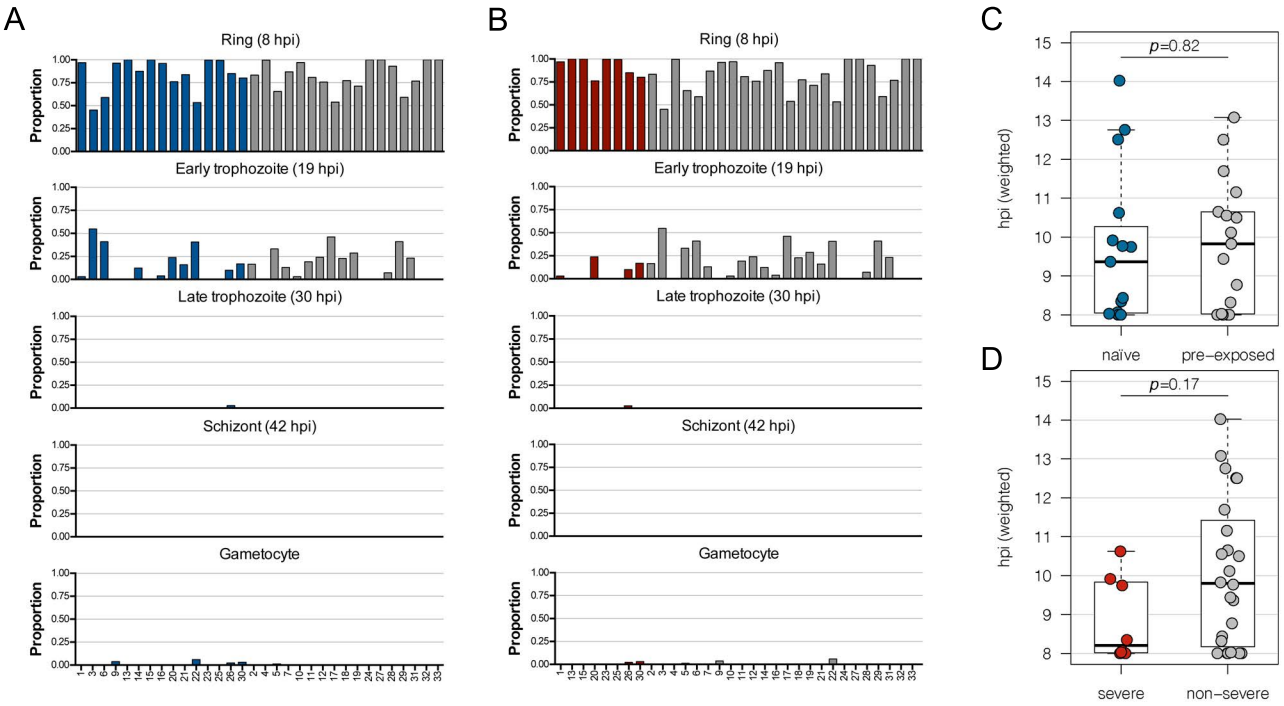
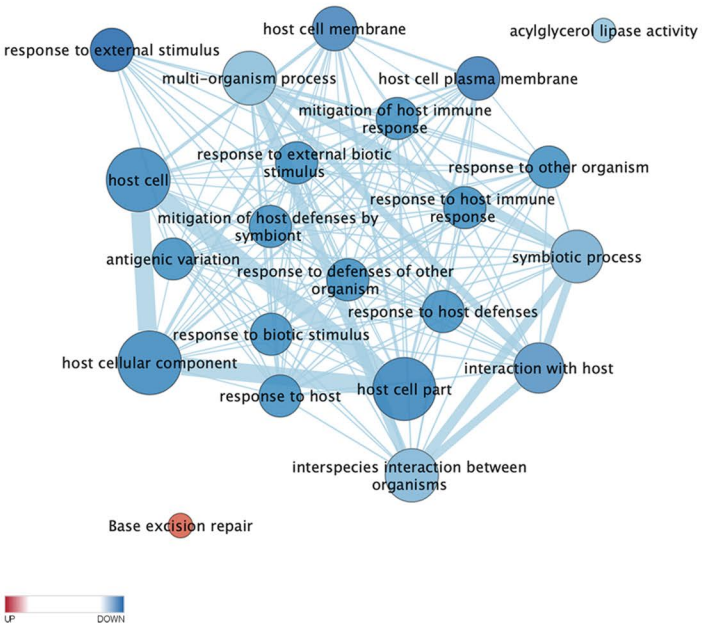

A
Ring (8 hpi)
Early trophozoite (19 hpi)
Late trophozoite (30 hpi)
Schizont (42 hpi)
Gametocyte
Proportion
B
C
p=0.82
hpi (weighted)
naïve
pre−exposed
D
p=0.17
severe
non−severe

Figure 3

A


host cell membrane
acylglycerol lipase activity
response to external stimulus
multi-organism process
host cell plasma membrane
mitigation of host immune response
response to external biotic stimulus
host cell
response to other organism
response to host immune response
mitigation of host defenses by symbiont
antigenic variation
symbiotic process
response to defenses of other organism
response to host defenses
response to biotic stimulus
host cellular component
interaction with host
response to host
host cell part
interspecies interaction between organisms
Base excision repair
UP
DOWN

B

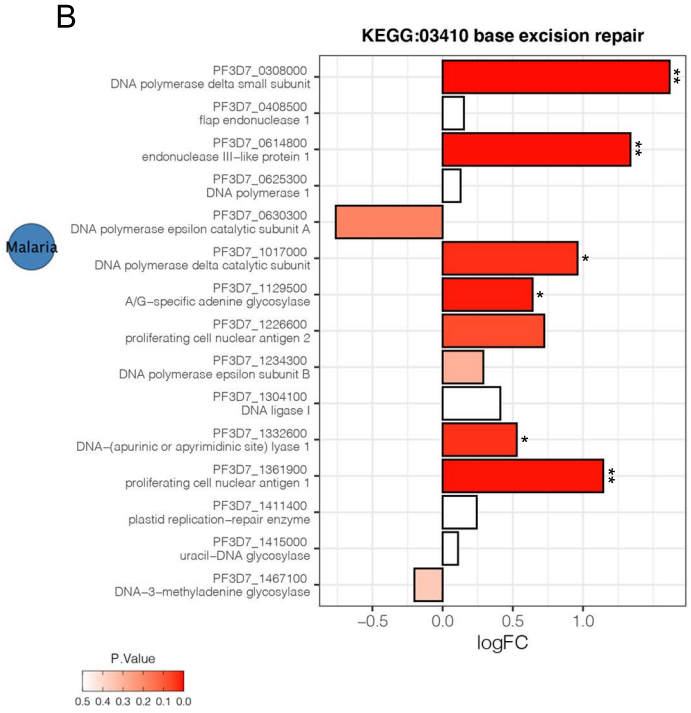

KEGG:03410 base excision repair
Malaria
PF3D7_0308000
DNA polymerase delta small subunit
PF3D7_0408500
flap endonuclease 1
PF3D7_0614800
endonuclease III-like protein 1
PF3D7_0625300
DNA polymerase 1
PF3D7_0630300
DNA polymerase epsilon catalytic subunit A
PF3D7_1017000
DNA polymerase delta catalytic subunit
PF3D7_1129500
A/G-specific adenine glycosylase
PF3D7_1226600
proliferating cell nuclear antigen 2
PF3D7_1234300
DNA polymerase epsilon subunit B
PF3D7_1304100
DNA ligase I
PF3D7_1332600
DNA-(apurinic or apyrimidinic site) lyase 1
PF3D7_1361900
proliferating cell nuclear antigen 1
PF3D7_1411400
plastid replication-repair enzyme
PF3D7_1415000
uracil-DNA glycosylase
PF3D7_1467100
DNA-3-methyladenine glycosylase
-0.5
0.0
0.5
1.0
logFC
P.Value
0.5 0.4 0.3 0.2 0.1 0.0

Figure 4

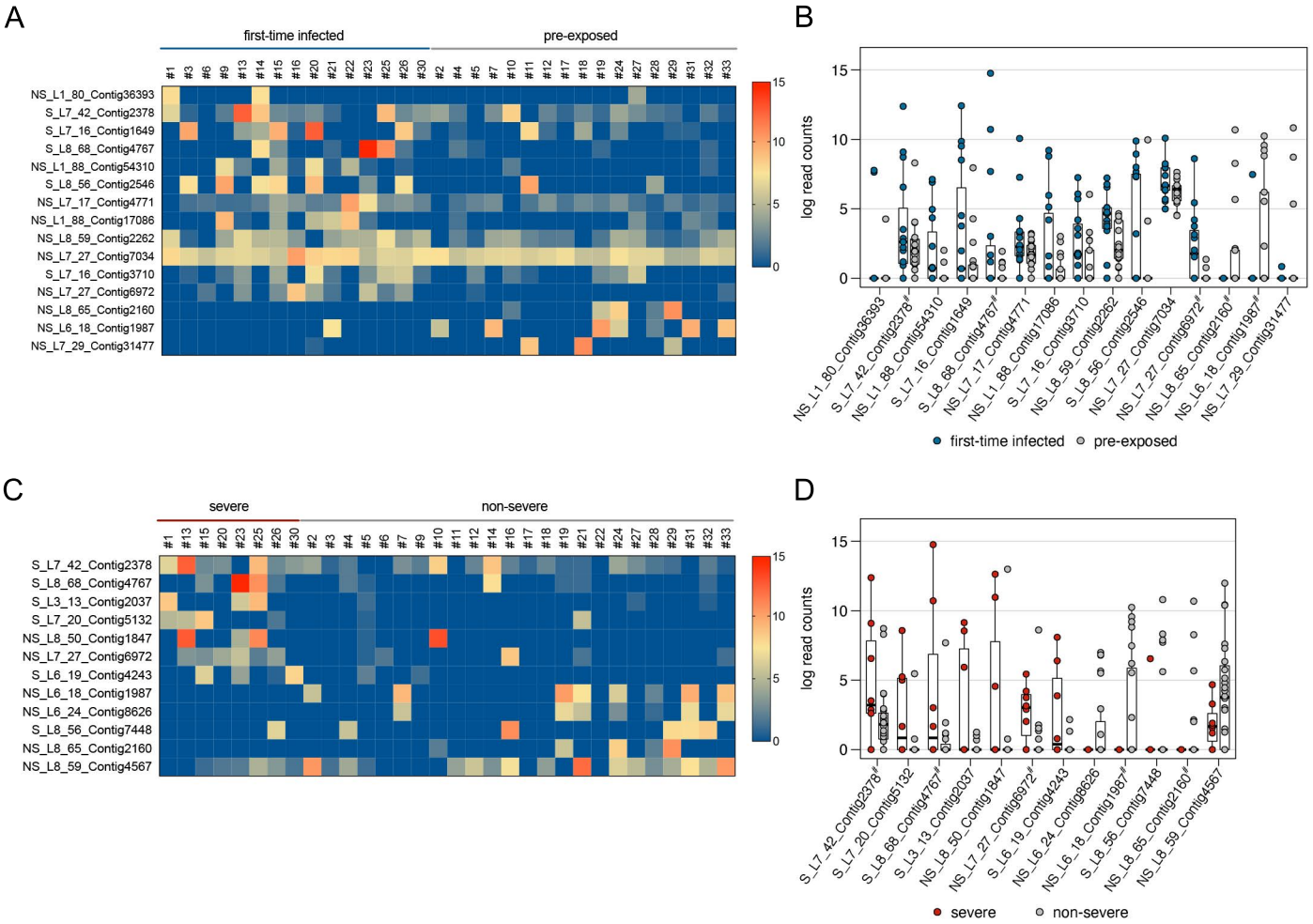

A
first-time infected
pre-exposed
NS_L1_80_Contig36393
S_L7_42_Contig2378
S_L7_16_Contig1649
S_L8_68_Contig4767
NS_L1_88_Contig54310
S_L8_56_Contig2546
NS_L7_17_Contig4771
NS_L1_88_Contig17086
NS_L8_59_Contig2262
NS_L7_27_Contig7034
S_L7_16_Contig3710
NS_L7_27_Contig6972
NS_L8_65_Contig2160
NS_L6_18_Contig1987
NS_L7_29_Contig31477
B
log read counts
first-time infected
pre-exposed
C
severe
non-severe
S_L7_42_Contig2378
S_L8_68_Contig4767
S_L3_13_Contig2037
S_L7_20_Contig5132
NS_L8_50_Contig1847
NS_L7_27_Contig6972
S_L6_19_Contig4243
NS_L6_18_Contig1987
NS_L6_24_Contig8626
S_L8_56_Contig7448
NS_L8_65_Contig2160
NS_L8_59_Contig4567
D
log read counts
severe
non-severe

Figure 5

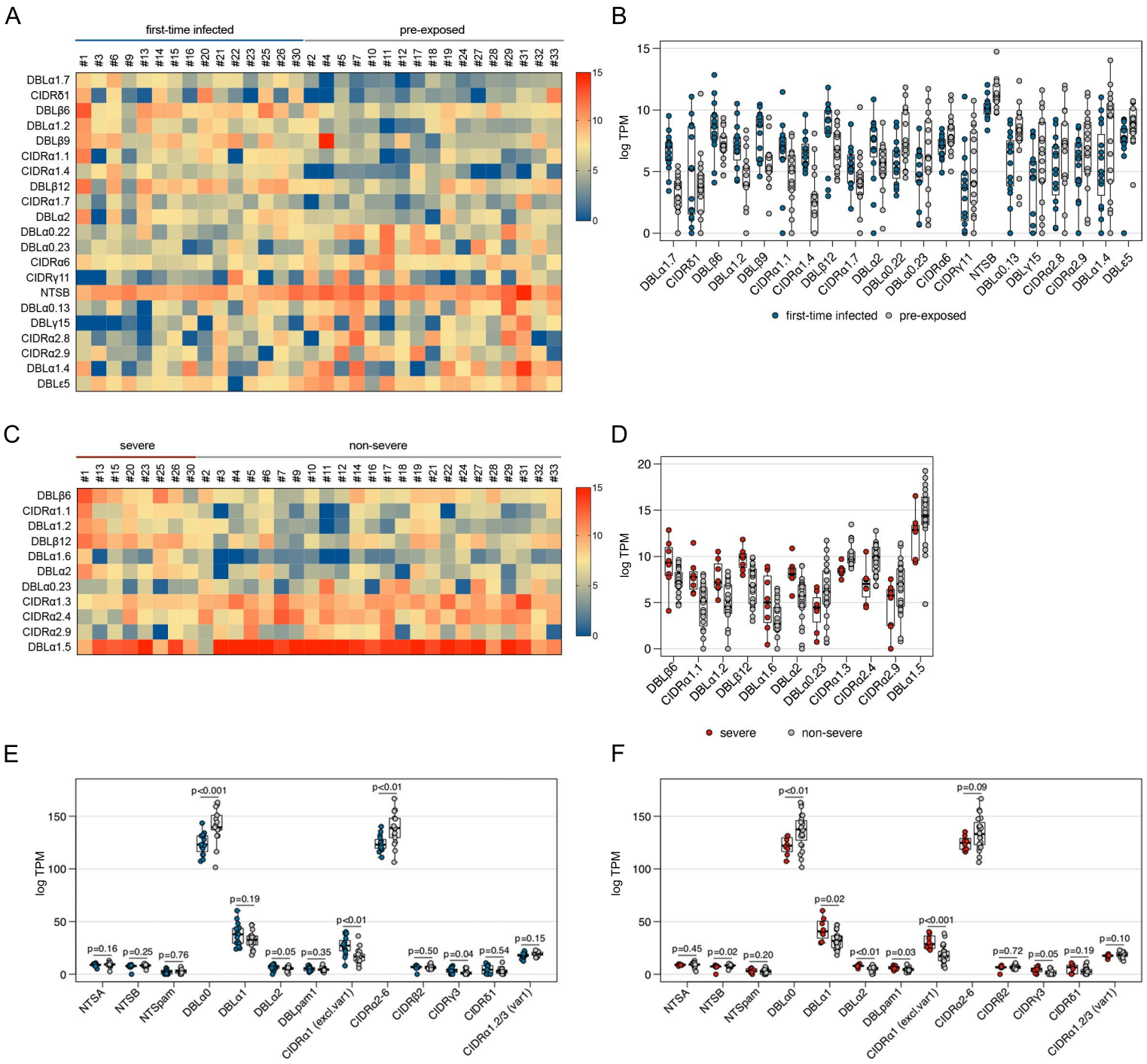

A
first-time infected
pre-exposed
DBLα1.7
CIDRδ1
DBLβ6
DBLα1.2
DBLβ9
CIDRα1.1
CIDRα1.4
DBLβ12
CIDRα1.7
DBLα2
DBLα0.22
DBLα0.23
CIDRα6
CIDRγ11
NTSB
DBLα0.13
DBLγ15
CIDRα2.8
CIDRα2.9
DBLα1.4
DBLε5
B
log TPM
first-time infected
pre-exposed
C
severe
non-severe
DBLβ6
CIDRα1.1
DBLα1.2
DBLβ12
DBLα1.6
DBLα2
DBLα0.23
CIDRα1.3
CIDRα2.4
CIDRα2.9
DBLα1.5
D
log TPM
severe
non-severe
E
log TPM
NTSA
NTSB
NTSpam
DBLα0
DBLα1
DBLα2
DBLpam1
CIDRα1 (excl.var1)
CIDRα2-6
CIDRβ2
CIDRγ3
CIDRδ1
CIDRα1.2/3 (var1)
F
log TPM

Figure 6

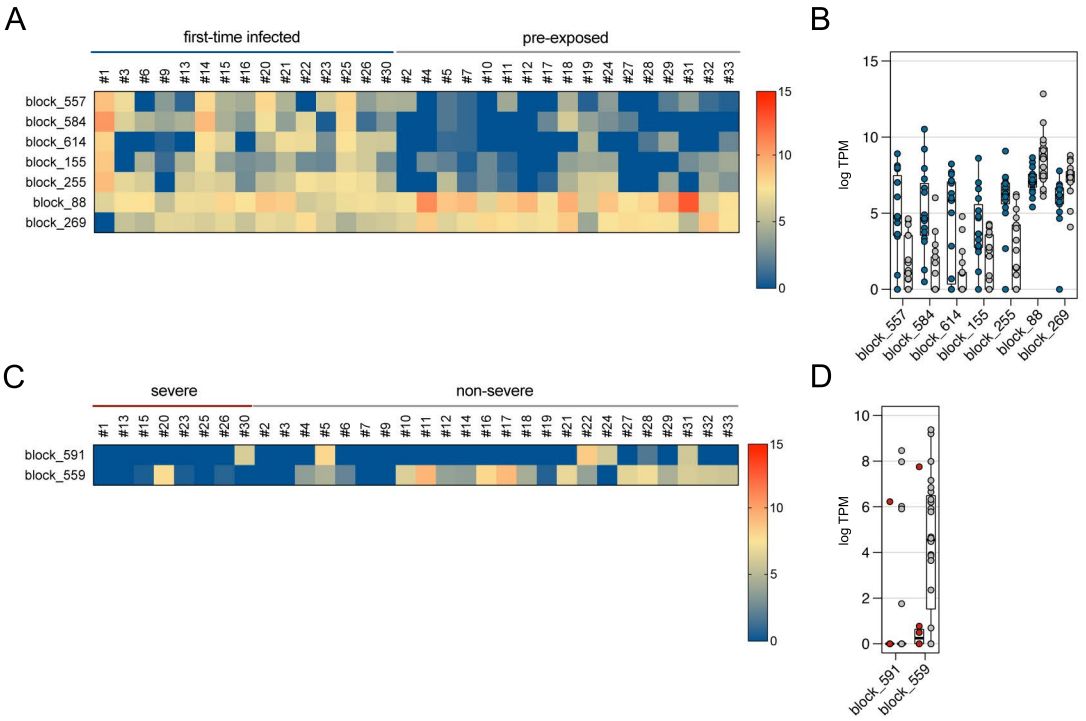
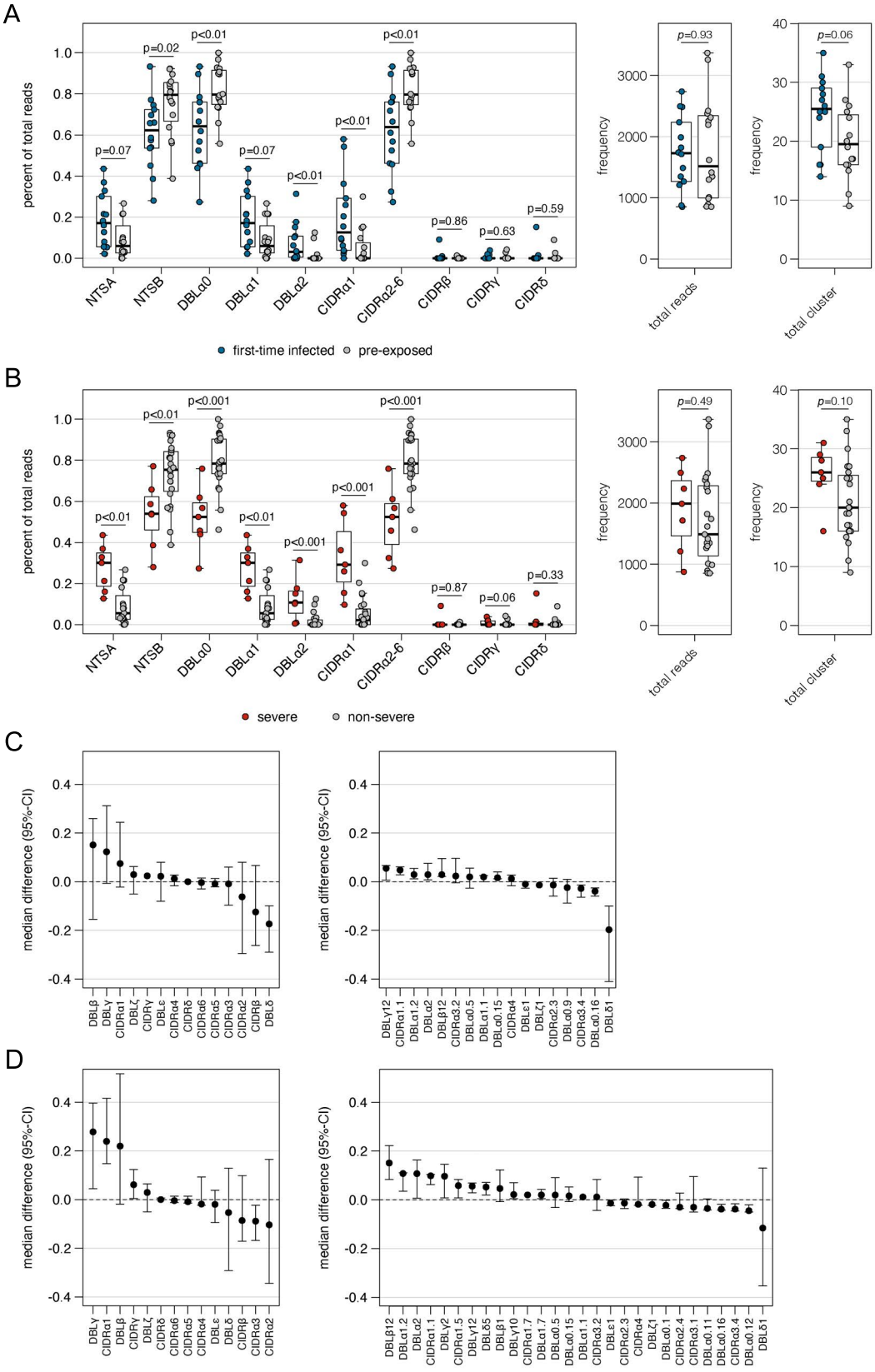
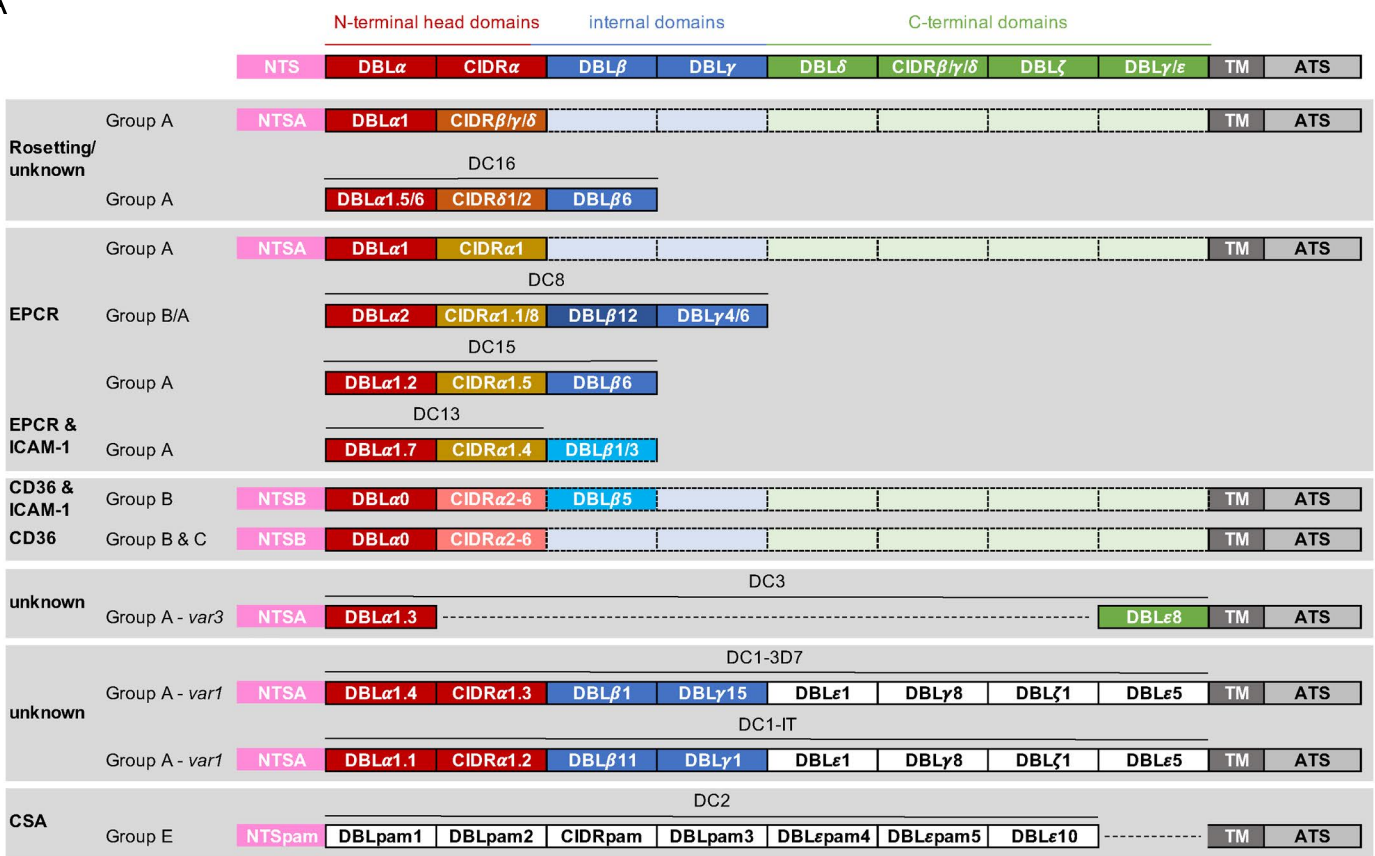

A
first-time infected
pre-exposed
#1 #3 #6 #9 #13 #14 #15 #16 #20 #21 #22 #23 #25 #26 #30 #2 #4 #5 #7 #10 #11 #12 #17 #18 #19 #24 #27 #28 #29 #31 #32 #33
block_557
block_584
block_614
block_155
block_255
block_88
block_269
15
10
5
0
B
log TPM
15
10
5
0
block_557
block_584
block_614
block_155
block_255
block_88
block_269
C
severe
non-severe
#1 #13 #15 #20 #23 #25 #26 #30 #2 #3 #4 #5 #6 #7 #9 #10 #11 #12 #14 #16 #17 #18 #19 #21 #22 #24 #27 #28 #29 #31 #32 #33
block_591
block_559
15
10
5
0
D
log TPM
10
8
6
4
2
0
block_591
block_559

Figure 7


A
percent of total reads
p=0.07
p=0.02
p<0.01
p=0.07
p<0.01
p<0.01
p<0.01
p=0.86
p=0.63
p=0.59
NTSA
NTSB
DBLα0
DBLα1
DBLα2
CIDRα1
CIDRα2-6
CIDRβ
CIDRγ
CIDRδ
first-time infected
pre-exposed
frequency
p=0.93
total reads
p=0.06
total cluster
B
percent of total reads
p<0.01
p<0.01
p<0.001
p<0.01
p<0.001
p<0.001
p<0.001
p=0.87
p=0.06
p=0.33
NTSA
NTSB
DBLα0
DBLα1
DBLα2
CIDRα1
CIDRα2-6
CIDRβ
CIDRγ
CIDRδ
severe
non-severe
frequency
p=0.49
total reads
p=0.10
total cluster
C
median difference (95%-CI)
DBLβ
DBLγ
CIDRα1
DBLζ
CIDRγ
DBLε
CIDRα4
CIDR5
CIDRα6
CIDRα5
CIDRα3
CIDRα2
CIDRβ
DBLδ
DBLγ12
CIDRα1.1
DBLα1.2
DBLα2
DBLβ12
CIDRα3.2
DBLα0.5
DBLα1.1
DBLα0.15
CIDRα4
DBLε1
DBLζ1
CIDRα2.3
DBLα0.9
CIDRα3.4
DBLα0.16
DBLδ1
D
median difference (95%-CI)
DBLγ
CIDRα1
DBLβ
CIDRγ
DBLζ
CIDR6
CIDRα6
CIDRα5
CIDRα4
DBLε
DBLδ
CIDRβ
CIDRα3
CIDRα2
DBLβ12
DBLα1.2
DBLα2
CIDRα1.1
DBLγ2
CIDRα1.5
DBLγ12
DBLδ5
DBLβ1
DBLγ10
CIDRα1.7
DBLα1.7
DBLα0.5
DBLα0.15
DBLα1.1
CIDRα3.2
DBLε1
CIDRα2.3
CIDRα4
DBLζ1
DBLα0.1
CIDRα2.4
CIDRα3.1
DBLα0.11
DBLα0.16
CIDRα3.4
DBLα0.12
DBLδ1

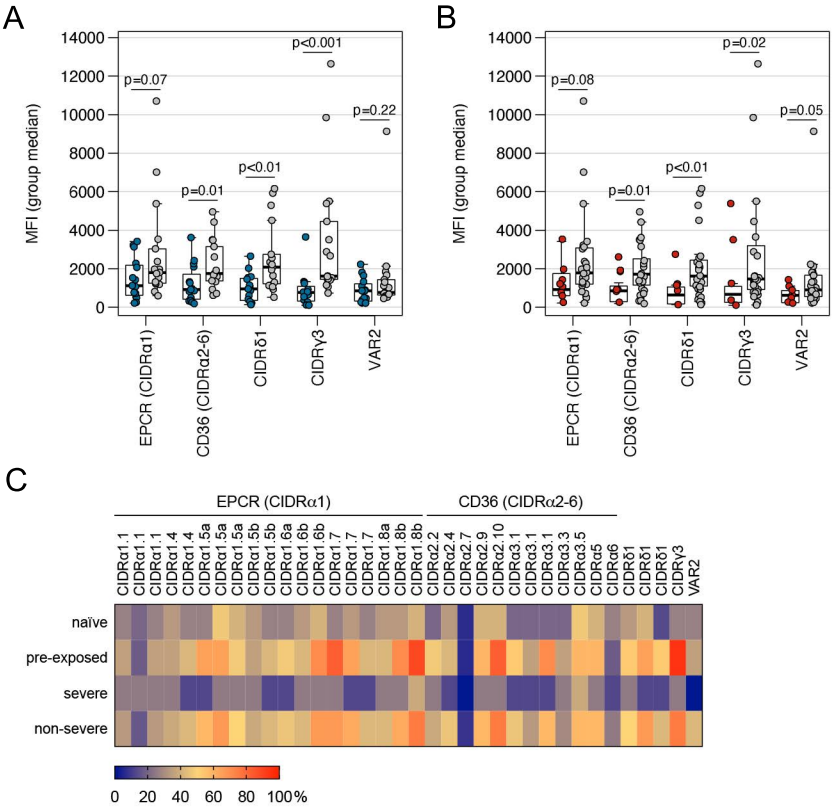
A



B

		First-time infected – Pre-exposed				Severe – Non-severe			
Binding phenotype	DC	Transcripts	Domains	Homology blocks	DBLα-tag	Transcripts	Domains	Homology blocks	DBLα-tag
Rosetting/ Unknown	DC16		CIDRδ1 DBLβ6	155 (NTSA)			DBLα1.6 DBLα1.5 DBLβ6		NTSA DBLα1
EPCR	DC8, DC15		DBLα1.2 DBLα2 CIDRα1.1 CIDRα1.7 DBLβ6 DBLβ12	155 (NTSA)	DBLα2 CIDRα1	S_L7_20_Contig5132	DBLα1.2 DBLα2 CIDRα1.1 DBLβ6 DBLβ12		NTSA DBLα1 DBLα2 CIDRα1
EPCR & ICAM-1	DC13		DBLα1.7 CIDRα1.4	155 (NTSA) 584 (DBLβ3)					
CD36 & ICAM-1		S_L8_68_Contig4767				S_L8_68_Contig4767 S_L3_13_Contig2037			
CD36		NS_L1_88_Contig17086 NS_L7_29_Contig31477	NTSB DBLα0.13 DBLα0.22 DBLα0.23 CIDRα2.8 CIDRα2.9 CIDRα6	88 (DBLα0) 69 (ATSB)	NTSB DBLα0 CIDRα2-6		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)					
Others		NS_L1_88_Contig54310 NS_L7_17_Contig4771 NS_L7_27_Contig7034	DBLβ9 CIDRγ11	557 (DBLβ-DBLγ) 614 (DBLβ12/3) 255 (DBLβ12/3/1/5)		S_L6_19_Contig4243		559 (DBLζ6)	

Figure 9


A
B
C
MFI (group median)
14000
12000
10000
8000
6000
4000
2000
0
p=0.07
p=0.01
p<0.01
p<0.001
p=0.22
p=0.08
p=0.01
p<0.01
p=0.02
p=0.05
EPCR (CIDRα1)
CD36 (CIDRα2-6)
CIDRδ1
CIDRγ3
VAR2
naïve
pre-exposed
severe
non-severe
0 20 40 60 80 100%