From genomic to LC-MS/MS evidence

Claire Kamaliddin, David Rombaut, Emilie Guillochon, Jade Royo, Sem Ezinmegnon, Gino Agbota, Stephanie Huguet, Sayeh Guemouri, Celine Peirera, Romain Coppée, et al.

To cite this version:

Claire Kamaliddin, David Rombaut, Emilie Guillochon, Jade Royo, Sem Ezinmegnon, et al.. From genomic to LC-MS/MS evidence: Analysis of PfEMP1 in Benin malaria cases. PLoS ONE, 2019, 14 (6), pp.e0218012. 10.1371/journal.pone.0218012 . hal-02618395

HAL Id: hal-02618395
https://hal.inrae.fr/hal-02618395

Submitted on 25 May 2020

RESEARCH ARTICLE

From genomic to LC-MS/MS evidence: Analysis of PfEMP1 in Benin malaria cases

Claire Kamaliddin[1], **David Rombaut**[2☯], **Emilie Guillochon**[3☯], **Jade Royo**[4], **Sem Ezinmegnon**[1,5], **Gino Agbota**[1,5], **Stéphanie Huguet**[6,7], **Sayeh Guemouri**[1], **Céline Peirera**[1], **Romain Coppée**[1], **Cédric Broussard**[2], **Jules M. Alao**[8], **Agnès Aubouy**[4], **François Guillonneau**[2], **Philippe Deloron**[1], **Gwladys I. Bertin**[1]*

1 UMR 261 – MERIT, IRD, Université de Paris, Paris, France, **2** 3p5 Proteomic Facility, Université de Paris, Paris, France, **3** Inovarion, Paris, France, **4** UMR 152 – PHARMADEV, IRD, Paul Sabatier Toulouse III University, Toulouse, France, **5** Centre pour la Recherche et l'Etude du paludisme associé à la grossesse et à l'enfance, Cotonou, Bénin, **6** Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA, Université Paris-Sud, Université d'Evry, Université Paris-Saclay, Gif sur Yvette, France, **7** Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA Université Paris-Diderot, Sorbonne Paris-Cité, Gif sur Yvette, France, **8** CHU-MEL, Pediatric department, Cotonou, Bénin

☯ These authors contributed equally to this work.
* gwladys.bertin@ird.fr

Citation: Kamaliddin C, Rombaut D, Guillochon E, Royo J, Ezinmegnon S, Agbota G, et al. (2019) From genomic to LC-MS/MS evidence: Analysis of PfEMP1 in Benin malaria cases. PLoS ONE 14(6): e0218012. https://doi.org/10.1371/journal.pone.0218012

Editor: Takafumi Tsuboi, Ehime Daigaku, JAPAN

Received: October 24, 2018

Accepted: May 23, 2019

Published: June 28, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files. Additionally, the WGS sequences are deposited in the European Nucleotide Archive (ENA). Accession numbers are uploaded as supplemental materials.

Funding: This study was funded by Merieux Research Grant to GIB (http://www.institut-merieux.com/fr/accueil/); Laboratoire d'Excellence GR-Ex, Paris, France, reference ANR-11-LABX-0051 - ANR-11-IDEX-0005-02 to PD and FG; NeuroCM project, that is funded by ANR-17-CE17-

Abstract

Background

PfEMP1 is the major protein from parasitic origin involved in the pathophysiology of severe malaria, and PfEMP1 domain subtypes are associated with the infection outcome. In addition, PfEMP1 variability is endless and current publicly available protein repositories do not reflect the high diversity of the sequences of PfEMP1 proteins. The identification of PfEMP1 protein sequences expressed with samples remains challenging. The aim of our study is to identify the different PfEMP1 proteins variants expressed within patient samples, and therefore identify PfEMP1 proteins domains expressed by patients presenting uncomplicated malaria or severe malaria in malaria endemic setting in Cotonou, Benin.

Methods

We performed a multi-omic approach to decipher PfEMP1 expression at the patient's level in different clinical settings. Using a combination of whole genome sequencing approach and RNA sequencing, we were able to identify new PfEMP1 sequences and created a new custom protein database. This database was used for protein identification in mass spectrometry analysis.

Results

The differential expression analysis of RNAsequencing data shows an increased expression of the *var* domains transcripts DBLα1.7, DBLα1.1, DBLα2 and DBLβ12 in samples from patients suffering from Cerebral Malaria compared to Uncomplicated Malaria. Our approach allowed us to attribute PfEMP1 sequences to each sample and identify new peptides associated to PfEMP1 proteins in mass spectrometry.

0001, to PD and GIB; LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS) to the POPS platform (SH); PhD Scholarship from French Minister of Research through MTCI Doctoral School (ED 563, Paris Descartes University) awarded to CK. The European Regional Development Fund (FEDER) and Cancéropôle Île-de-France provided a grant to the 3P5 group (DR, CB, FG) to purchase the Orbitrap. Funder Merieux Research was involved in the decision to publish. All other funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Inovarion (Paris, France https://www.inovarion.com/) provided support in the form of salary to EG, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: EG is a salaried employee of Inovarion (Paris, France https://www.inovarion.com/). The 3P5 group (DR, CB, FG) group received funding from Cancéropôle Île-de-France (https://www.canceropole-idf.fr/) to purchase the Orbitrap. There are no patents, products in development or marketed products associated with this research to declare. This does not alter our adherence to PLOS ONE policies on sharing data and materials. All other authors declare that no competing interests exist.

Conclusion

We highlighted the diversity of the PfEMP1 sequences from field sample compared to reference sequences repositories and confirmed the validity of our approach. These findings should contribute to further vaccine development strategies based on PfEMP1 proteins.

Introduction

Through its asexual development in human erythrocytes, *Plasmodium falciparum* grows and reshapes its host cell. Parasite proteins exported at the host cell surface mediate infected erythrocyte's adhesion to the host's endothelium that leads to hypoxia, occlusion and endothelial activation. In cerebral malaria (CM) pathophysiology, the sequestration of infected erythrocytes (iE) in the brain capillaries is believed to trigger coma and brain swelling [1].

Among the proteins exported at the erythrocyte's surface, the *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) protein family is involved in cytoadhesion [2]. PfEMP1 proteins are encoded by the multigenic *var* gene family [3–5], consisting in ~ 60 copies per parasite genome [6]. The diversity among *var* sequences is almost endless [7,8] thus participating to the infected erythrocyte ability to evade the immune system. PfEMP1 proteins are high molecular weight transmembrane proteins (200–350 kDa), and are composed of an intra-erythrocytic segment, which is conserved, and a highly variable extracellular segment [9]. The extra-erythrocytic segment is composed of 4 to 9 alternated Duffy Binding Like (DBL) or Cystein Inter Domain Rich (CIDR) domains. The nature and the arrangement of these domains determine the binding phenotype of the iE [9,10]. More specifically, the transcripts coding for the domains cassettes DC8 (DBLα2-CIDRα1.1-DBLβ12) and DC13 (DBLα1.7-CIDRα1.4- DBLβ1/3) are preferentially expressed in severe malaria isolates [11,12].

Among the PfEMP1 receptors in human endothelium, the most common is the broadly expressed in human cell CD36, but PfEMP1 binding to CD36 is not related to any specific form of malaria [13]. Two human host receptors for PfEMP1 binding in the context of severe malaria have been identified: the InterCellular Adhesion Molecule-1 receptor (ICAM-1) [14] and the Endothelial Protein C Receptor (EPCR) [15], both expressed in brain endothelial cells [16], and co-localized with the sequestered iEs in severe malaria [14]. The binding domain for ICAM-1 receptor is located in the C-terminal third of the DBLβ3 [17], and the residues involved in PfEMP1 binding to ICAM-1 are highly variable with a limited binding pattern [18,19]. The role of EPCR in PfEMP1 binding has been more recently shown [15] and is still an important research problematic [20]. EPCR binding is mediated by highly variable but structurally conserved CIDRα1 PfEMP1 domains (more precisely CIDRα1.1 and CIDRα1.4–1.8) [21,22]. Importantly, the level of PfEMP1 transcript associated with EPCR binding is higher in samples from patients suffering from severe malaria and increases with the severity of the disease [20,21]. A dual binding with EPCR and ICAM-1 has been suggested, since not all CM isolates present an increase in binding-EPCR PfEMP1 coding transcript [23]. The expression of DBL involved in ICAM-1 binding is associated with dual ICAM-1 and EPCR binding [19].

Most field studies looking for *P. falciparum* binding phenotypes are based on molecular biology analysis and have shown that transcript coding for specific PfEMP1 domains expression level is associated with disease outcome [11,21,24,25]. However, this strategy is currently limited to the already identified PfEMP1 domains and does not give proficiency of the expressed proteins. Recently, several strategies have been implemented to investigate the variability of full-length *var* genes using whole genome sequencing [6,26,27] or dedicated long range sequencing with a hybrid PCR approach [23]. In addition, Tonkin Hill et *al* performed *de novo* assembly of *var* genes issued from RNA sequencing (RNAseq) and identified transcripts up-regulated in severe malaria [28]. These recent publications provide insight towards *var* genes variability in the studied areas. However, the identification of PfEMP1 proteins by mass spectrometry approach (LC-MS/MS) remains infrequent in publications.

To complement this deficiency, we aimed to conduct a mass-spectrometry-based proteomics analysis of *P. falciparum* field isolates proteome. LC-MS/MS is a powerful and sensitive tool for protein identification, however, its application for PfEMP1 identification remains challenging because PfEMP1 has highly variable sequences, yet database repository is usually simplified by eliminating redundancy. That is the reason why they do not reflect the natural sequence diversity that may occur in such a context.

To identify PfEMP1 associated with *P. falciparum* clinical outcome in endemic settings, we used a "proteogenomic" approach. Specific PfEMP1 sequences from each isolate were reconstructed *de novo* using whole genome sequencing (WGS) data to identify the expressed transcripts and enrich the protein database (Fig 1). We analysed the whole proteome of samples from patients presenting CM, Severe Anemia (SA) or Uncomplicated Malaria (UM), and attribute PfEMP1 sequences within these samples. Corresponding samples were analysed in RNAseq for PfEMP1 expression analysis, in relation to proteomic results.

We performed RNAseq successfully on 7 field samples (3 CM, 2 SA and 2 UM) and managed to identify the PfEMP1 protein sequence associated with 4 CM samples, 9 SA and 9 UM samples from Benin, West Africa, using LC-MS/MS. We confirmed the expression of several PfEMP1 within a single field isolates and provided the first identification at the patient's level of PfEMP1 expressed by the parasite in the context of acute *P. falciparum* infection (Fig 1).

Material and methods

Ethic statement

Ethical clearance was obtained from the Institutional Ethics Committee of the faculty of health science at the Abomey-Calavi University in Benin (clearance n°90, 06/06/2016). Before inclusion, written informed consent was obtained from children's guardians. Patients were treated in accordance to the national malaria program policy. The methods were carried out in accordance with the relevant guidelines and regulations.

Sample collection

Patients under age of five, presenting *P. falciparum* acute infection, were included in the Lagune Mother and Child Hospital in Cotonou (severe malaria), Benin and Saint-Joseph Hospital, in Sô-Ava, Benin (UM) in rainy season (May–August) 2016. Severe malaria patients were classified as following: CM was defined as associated with a coma (Blantyre score ≤ 2) and the absence of meningitis detected by CSF count and culture and SA was defined with Hb $<$ 5g/dL, measured using Hemocue device (Radiometer). UM was defined as a *P. falciparum* infection with fever, in the absence of any other complication. Five mL of peripheral whole blood were collected on EDTA. Parasite density was evaluated with Giemsa-stained thick blood smear. Only pure *P. falciparum* infections were retained for the study. Samples were depleted from white blood cells using a gradient-based separation technique Ficoll (GE Healthcare Life Science).

Whole genome sequencing

Fifty μL of erythrocyte's pellet was extracted using DNEasy Blood kit (Qiagen). WGS was performed by the Malaria Genomic Epidemiology Network (MalariaGEN) at the Welcome Trust

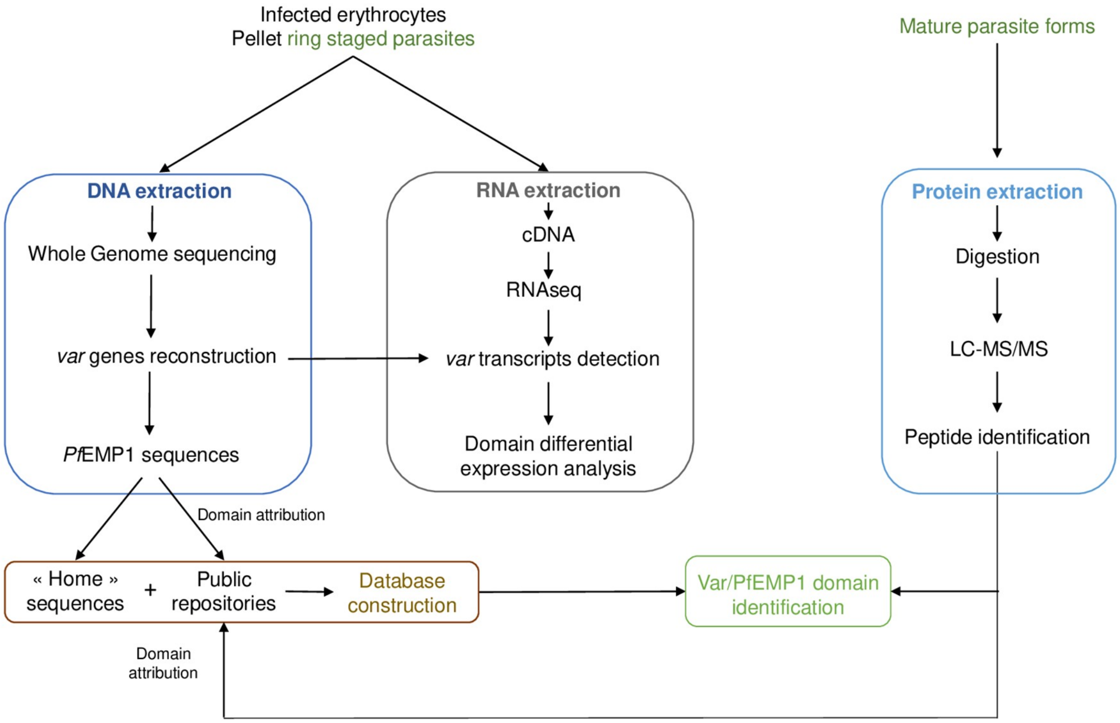

Fig 1. Experimental design—Proteogenomic approach on field samples for PfEMP1 identification. Whole blood sample from patients are collected. DNA and RNA are extracted from parasite's ring forms. For LC-MS/MS analysis, parasites are matured, and the corresponding proteins are extracted and analysed using the mass spectrometer. Whole genome sequencing data provides the var repertoire from each isolate and allows the assessment of RNA expression in each sample. In addition, WGS data were used to enrich the protein database for protein identification with LC-MS/MS data.

https://doi.org/10.1371/journal.pone.0218012.g001

Sanger Institute (Hinxton, UK). Reconstructed *var* genes were kindly provided by Thomas Otto, Matt Berriman and Chris Newbold from the Welcome Trust Sanger Institute and translated into putative PfEMP1 protein sequences for protein identification. The raw reads from whole genome sequencing are available on the ENA server under the accession number listed in S1 Table.

Transcriptome studies of ring staged parasites

Ring staged parasites were preserved in 5 volumes of pre-warmed (37˚C) TriZol (Life Technology), vortexed then immediately frozen at -80˚C until further utilization. RNA were extracted as described [29], then digested with DNAse I (Qiagen) and purified using RNEasy MinElute Cleanup kit column (Qiagen). Only RNA presenting a RNA Integrity Number (RIN) > 7 evaluated with PicoChip Agilent 2100TM Bioanalyzeur (Agilent) were retained for downstream analysis [30]. RNAseq libraries were performed using TruSeq Stranded mRNA protocol (Illumina, California, U.S.A.). RNAseq samples have been sequenced in paired-end (PE) with a sizing of 260 base pairs and a read length of 150 bases. Fifty four samples by lane of Illumina NextSeq500 (IPS2 POPS platform) were generated using individual barcoded adapters. Approximately 5 million of PE reads by sample were obtained. The raw reads (fastq) were trimmed with Trimmomatic [31] tool for Phred Quality Score Qscore >20, read length >30 bases, and ribosome sequences were removed using sortMeRNA [32]. RNAseq paired-end reads were mapped to the human reference genome Hg38 (UCSC Genome Browser). Unattributed reads were mapped to the *P. falciparum* 3D7 strain reference genome (PlasmoDB release 41), the reference *var* genes removed and replaced by the *var* genes of each sample issued from its own whole genome using HISAT2 (v2.1.0) [33].

Raw counts for each *var* transcript were obtained using HTSeq-count (0.11.1) [34]. Transcript abundance was evaluated using RPKM values. We considered a transcript as present if the RPKM value was > 1. To assess the potential expression differences according to the sample group (patients' clinical presentation—severe or uncomplicated malaria), we performed a selective read count on each *var* domain subtype from the cognate isolate *var* transcripts. The differential expression analysis was performed on the obtained counts using DESeq2 R package [35].

Proteome analysis of *P. falciparum* late trophozoites using LC-MS/MS

Blood samples were matured *in vitro* for 18 to 32 hours in RPMI medium supplemented with human serum and Albumax (Gibco) and preserved after MACS (Myltenyi Biotech) enrichment as described [36].

Whole cell infected erythrocyte lysates were solubilized and digested in solution using trypsin (Promega, sequencing Grade). Briefly, 50 μg of proteins from whole cell lysates were diluted to 25 μl in solubilization buffer (1% sodium desoxycholate, 100 mM Tris/HCl, pH 8.5, 10mM TCEP, 40 mM chloroacetamide), heated for 5 min at 95°C and sonicated three times for 30 s. Once at room temperature, extracts were diluted with 25 μl Tris-ACN buffer (50 mM Tris/HCl pH 8.5, 10% ACN). Collected peptides were fractionated in 5 fractions per sample by strong cationic exchange (SCX) StageTips [37].

LC-MS/MS analysis was performed on a Dionex U3000 RSLC nano-LC-system coupled to an Orbitrap-fusion mass spectrometer (Thermo Fisher Scientific) as described [38]. Peptides from each SCX fraction were solubilized in 0.1% trifluoracetic acid (TFA) containing 10% acetonitrile (ACN) and were separated on a C18 reverse-phase resin (75-μm inner diameter and 15-cm length) with a 3-hr gradient. The mass spectrometer acquired data throughout the elution process and operated in a data-dependent scheme.

For protein identification using LC-MS/MS, we created a custom database containing both the human proteome (to identify peptides issued from the erythrocyte) and *P. falciparum* proteome. In order to perform PfEMP1 protein identification, we concatenated *P. falciparum* proteins sequences from PlasmoDB (v35), Uniprot and NCBI. In addition, we implemented our own PfEMP1 sequences, obtained after *in silico* translation from *var* genes reconstruction. Duplicate sequences were removed.

The LC-MS/MS data were analyzed using MaxQuant version 1.5.2.8 [36] as described [39]. The database used was our homemade database and the list of contaminant sequences from Maxquant. For analysis, LFQ results from MaxQuant were imported into the Perseus software (version 1.5.1.6). Reverse and contaminant proteins were excluded. Only proteins from *P. falciparum* were selected for further analysis. We then focused on the membrane associated and putative proteins from *P. falciparum*.

Analysis of *var* transcripts and PfEMP1 proteins

PfEMP1 sequences from expressed *var* transcripts and proteins identified in LC-MS/MS were aligned using the VarDom server against reference sequences for domain identification [7]. We specifically searched the pattern for ICAM-1 binding $I[V/L]x_9N[E]GG[P/A]xYx_{27}GPPx_3H$ [19] using the ProSite online interface [40]. To identify the nature of each domain within the identified sequences from RNAseq and LC-MS/MS, we aligned each DBL and CIDR domain with the VarDom database domain sequences using MAFFT tool (v7) [41]. Using the MAFFT output, we generated a phylogenic tree using PhyML online tool with default parameters [42].

Results were displayed using iTOL online tool [43]. PfEMP1 domains were attributed to all identified peptides. We considered a peptide specific of a subdomain if a peptide was corresponding to the same subdomain in at least 3 different PfEMP1 proteins.

Statistical analysis

Patient's samples information's were compared between the 3 patient's groups (UM, CM and SA) using one-way ANOVA. Bonferroni's Multiple Comparison Test was applied for individual group comparison. We considered a p value < 0.05 as significant. Qualitative data were compared with Chi Squared test using contingency table. All analyses were performed using Prism v5 (Graphpad). For the differential expression analysis, a domain subtype was considered as differentially expressed in a condition compared to another for log2 (fold-change) value > 1 and adjusted p-value < 0.1.

Results

Included samples

We included 95 patients, covering 31 SA, 18 CM and 46 UM. The average patient age was similar among all inclusion groups. Parasite density geometric mean was 8,055 p/μL for UM group, 34,191 p/μL for CM and 24,313 p/μL for SA. Parasite density was only significantly different between SA and UM samples ($p = 0.0158$ with Bonferroni's Multiple Comparison Test). Hemoglobin level was measured for 20 UM, 17 CM and 31 SA, respectively 11.28 [10.26; 12.75], 5.51 [4.10; 6.56] and 4.393 [3.90; 5.00] g/dL. Hemoglobin level was statistically different for UM samples (*vs*. CM and *vs*. SA) ($p < 0.05$). No difference in erythrocyte count ($p = 0.1274$) and temperature ($p = 0.9125$) was retrieved between SA and CM. All CM patients presented a coma (average BS 2 [2; 2]), while SA patients did not (average BS 4.6 [4; 5]) ($p < 0.0001$).

For LC-MS/MS analysis, we selected samples among those which showed successful maturation. The analysis has been performed on 4 CM, 9 SA and 9 UM samples. 25 samples qualified for RNAseq among which 7 were successfully sequenced.

Var genes transcripts identification with RNAseq

Overall 165 *var* transcripts were identified (S2 and S3 Tables) among which 134 sequences corresponded to Severe Malaria (SM) samples (52 CM and 82 SA) and 31 to UM samples. We then focused on the corresponding sequences domains combination, considering the sequences with at least one NTS domain. We found 102/134 *var* transcripts in the SM groups (30 in the CM samples and 72 in the SA samples) and 24/31 of UM associated *var* sequences.

Among these sequences, the domain combination the most representative is NTS-DBLα-CIDRα-DBLδ. This combination was identified in 17/24 (71%) UM samples and in 45/102 (44%) SM group thus 10/30 CM and 35/72 SA associated *var* transcripts.

The second domain combination is NTS-DBLα-CIDRα-DBLβ and corresponded to 6/24 (25%) of UM associated *var* transcripts and 42/102 (41%) of SM (17/30 CM and 25/72 SA) associated *var* transcripts.

Regarding the domain cassette distribution, we have identified 2 DC8 but no DC13 among *var* transcripts of CM samples, and 2 DC8 and 3 DC13 were identified in SA samples. From to UM samples, we have identified neither domain cassettes DC8 nor DC13.

The specific search of the binding pattern for ICAM-1 retrieved three identifications from CM samples, two in the SA group within the *var* transcripts sequences and no identification among the UM samples.

In addition, we performed a differential expression analysis on the *var* domains subtypes of each sample. Twelve domains subtypes were up-regulated in CM samples compared to the UM samples (Fig 2A), among which the DBLα2 and DBLβ12. These domains match to the organisation of DC8. The DBLα1.7 domain (part of DC13) is the most differentially expressed in the CM samples compared to the UM samples.

Eleven domains subtypes were up-regulated in SA compared to UM (Fig 2B). These subtypes were different from those found in CM compared to UM and not correspond to domain cassettes. We found no significantly expressed domains subtype in the CM samples in comparison to SA samples.

Protein identification using LC-MS/MS

Protein identification was performed using a homemade database (reference sequences from human and *P. falciparum* repositories, and the assembled *var* from field samples) containing 295,601 protein sequences, among which 87,489 were *P. falciparum*-associated sequences. Overall, we identified 3300 proteins. A total of 1302 proteins were associated to the human proteome, and 1912 to *P. falciparum*'s. Among those later, 460/1912 proteins were identified as *P. falciparum* membrane-associated proteins, including 60.4% of hypothetical or putative, 12% of PfEMP1s, 3.5% of RIFINs, 0.9% of STEVORs, 1.5% of PHISTs and 21.7% belong to other protein families.

A total of 57 proteins associated with PfEMP1 were identified. Only 10 of the identified PfEMP1 using LC-MS/MS (as part of the identified isoforms) were known sequences from public database repository (Uniprot and PlasmoDB). All other identified PfEMP1 sequences resulted from the translation of the reconstructed *var* genes from our samples (S3 Table).

PfEMP1 identification and composition

Average molecular weight of the identified PfEMP1 was 228.3 kDa. Using the VarDom online server, we reconstructed the domain architecture from the identified proteins (S4 Table) and we were able to identify domains in 54 out of 57 sequences. NTS domain was found in 41/54 of the sequences (76%) and 39/41 (95%) of the sequences identified presenting an NTS domain displayed DBLα-CIDRα associated to the NTS domain. The three-major head-terminal domain organizations were the following: NTS-DBLα-CIDRα-DBLβ (n = 24/41; 59%), NTS-DBLα-CIDRα-DBLδ (n = 12/41; 29%) and NTS-DBLα-CIDRα-DBLγ (n = 1/41; 2,4%).

Considering the difficulties to attribute a given PfEMP1 protein to a sample in this experimental setting, we then focused our analysis on the peptides attributed to PfEMP1 proteins. We identified 147 peptides attributed to PfEMP1, among which 110 were unique peptides (S5 Table). Among these 147 peptides identification, 46 were peptides from the public data repositories, while the remaining ones were specific to the protein sequences identified using WGS. The peptides were distributed as following among the PfEMP1 domains: ATS 14/147 (10%), CIDRα 6/147 (4%), CIDRβ 22/147 (15%), DBLα 11/147 (7%), DBLβ 14/147 (9%), DBLδ 32/147 (22%), DBLε 2/147 (1%), DBLγ 9/147 (6%), DBLζ 1/147 (0.7%), NTS 21/147 (14%) and 15/147 (10%) of the peptides were unattributed. The two majors' domains identified with the peptides are CIDRβ and DBLδ which are in equivalent proportion in all clinical group (p = 0.41 and 0.21 respectively). Regarding the CM samples, no peptide associated to the DBLα was identified (Fig 3).

Discussion

The evolution of *P. falciparum* infection from uncomplicated forms of the disease to cerebral malaria, the most fatal, is a complex phenomenon [44]. There are strong evidences that the

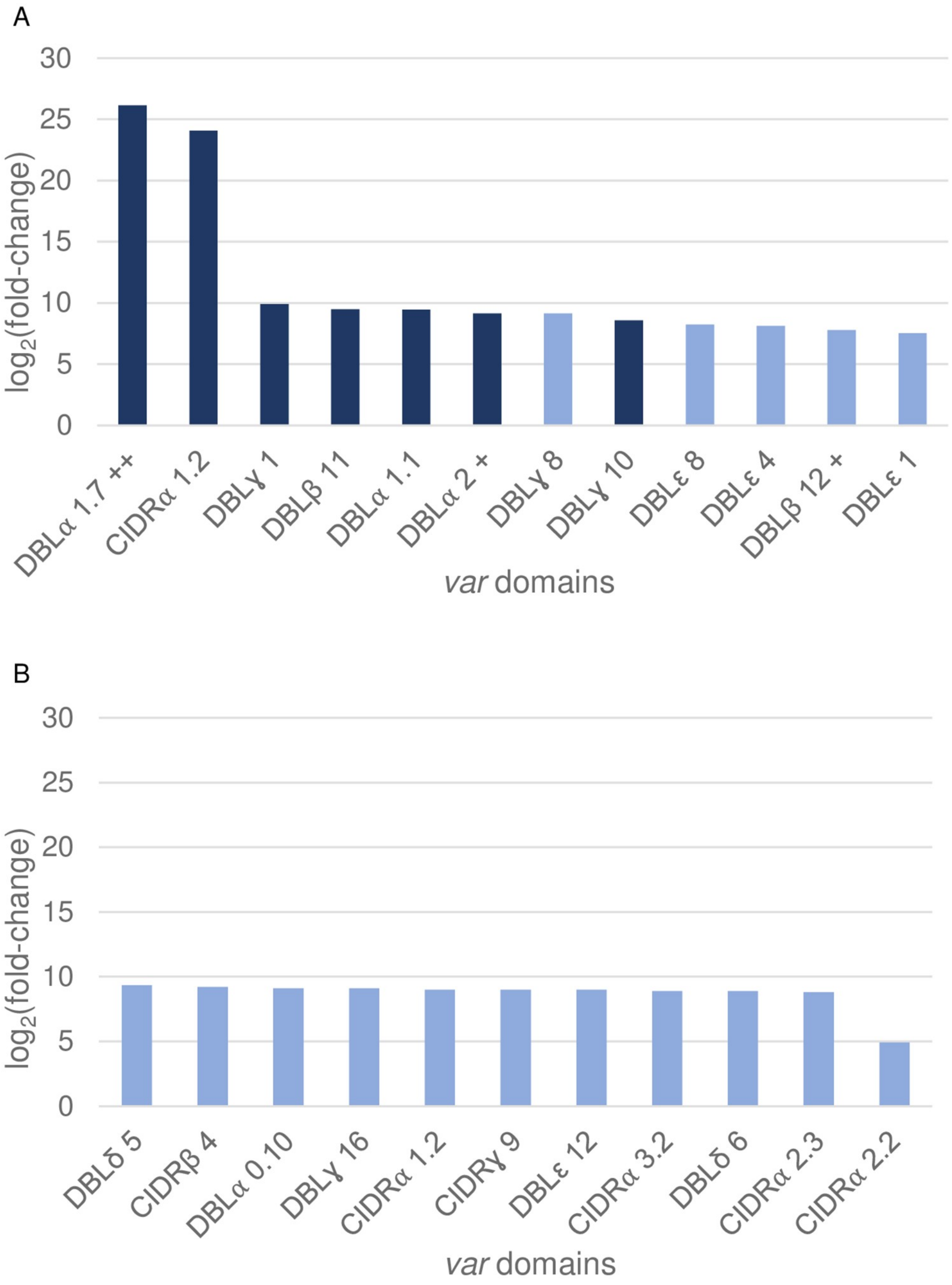

Fig 2. Var transcripts associated domains subtypes identified as up regulated. The bar graph represents the expression differential of the sub-domains realized with the package R DESeq2 (A) overexpressed in the CM in comparison to UM samples and (B) overexpressed in the SA in comparison to UM samples. The signs + and ++ represents respectively the subdomains of DC8 and DC13. Y axis plots the values of log2 (fold change) between the clinic groups by subdomain. X axis represents each domains subtype identified as up-regulated in clinic groups Two adjusted p-value thresholds are indicated: dark blue < 0.05 and light blue < 0.1.

https://doi.org/10.1371/journal.pone.0218012.g002

PfEMP1 proteins are involved in the disease progression since they allow the parasite to bind to host endothelium [10]. It is believed that a distinct subset of PfEMP1 proteins is involved in severe malaria [23,45], most likely by providing to the parasite the ability to sequester to a given receptor. However, PfEMP1 identification in natural infection remained challenging,

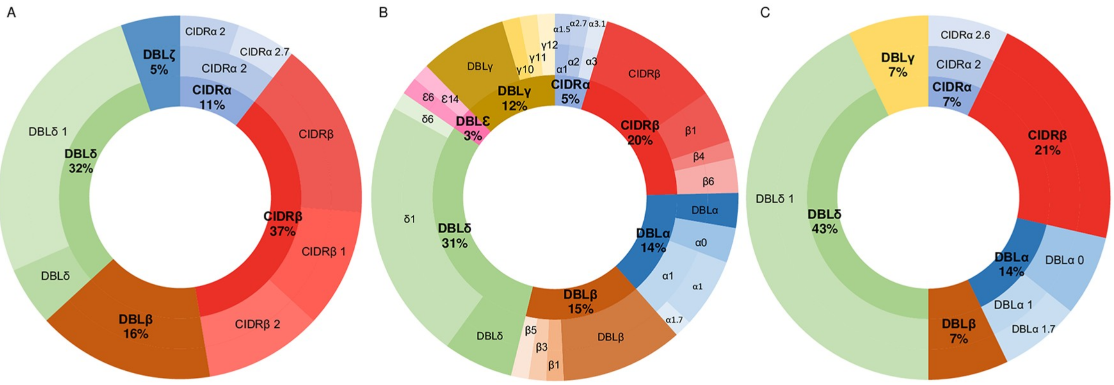

Fig 3. Proportion of peptide corresponding to each PfEMP1 domains and subdomains in association with the clinical outcome. Domains were attributed to each peptide identified in LC-MS/MS. The proportions are displayed for (A) CM patients, (B) SA patients and (C) UM patients.

https://doi.org/10.1371/journal.pone.0218012.g003

due to the large size of PfEMP1 and their high sequences diversity. Recently, Jespersen et *al* [23] provided a new insight towards *var* genes sequences expression analysis in patient's sample using transcript reconstruction after DBLα barcoding. They confirmed the preferential expression of CIDRα associated with EPCR binding in severe malaria patients. In addition, Tonkin Hill et *al* performed a *de novo* reconstruction of *var* genes from patient's isolates [28].

We used a mass spectrometry–based proteomic approach to analyse the *P. falciparum* proteome in the context of severe malaria (SA and CM) compared to UM. We aimed to accurately identify, at the protein level, the PfEMP1 sequence variants associated with diseases severity. To this end, we initiated a "proteogenomic" study of field samples (Fig 1).

Using reconstructed *var* genes obtained by WGS, we were able to identify the transcript expressed for each isolate among the one from the cognate genome. In addition, we performed a differential expression analysis of the *var* domains. We demonstrated that the domains DBLα1.7/2 and DBLβ12 are a signature of the CM sample. These domains are part of the DC8 and DC13, which are described as involved in the pathogenesis of cerebral malaria in patients from several endemic area [11,20,25,36]. The convergence of our results with the published results in the literature using targeted methods enforce the association of DBLα1.7/2 and DBLβ12 expression and cerebral malaria. We also demonstrated that the *var* expression pattern of the SA patients was distinct from the CM patients, in accordance with the specific sequestration pattern of *P. falciparum* in CM pathogenesis.

At the protein level, we were able to identify peptides associated with PfEMP1. As anticipated, most of the identified PfEMP1 came from the newly added sequences to the database (10/57 were known sequences), confirming the validity of our approach considering the high variability of PfEMP1 proteins.

Using peptides fractionation, we identified more proteins than previously published studies [12,46], with higher sequence coverage. We identified a set of 57 PfEMP1 in the studied samples and investigated the structure of theses sequences. Our finding revealed that the two main domain organisations were NTS-DBLα-CIDRα-DBLβ and NTS-DBLα-CIDRα-DBLδ. The high proportion of NTS-DBLα-CIDRα-DBLβ in our identified PfEMP1 proteins compared to genomic sequences within the same sample pool reflects the preferential expression of the PfEMP1 containing this domain association. The CIDRα-DBLβ tandem is associated with the potential "double binding" PfEMP1 [19,24], targeting both ICAM-1 (through DBLβ [19]) and EPCR (through CIDRα [22]) human endothelial receptors. Nevertheless, the highly recombinogenic nature of *var* genes means that the presence of a partial *var* sequence in a *var* gene from one isolate does not mean that if the sequence is present in another isolate that it is present in the same gene. Thus inferring the presence of entire PfEMP1s or domains for which peptides have not been directly obtained must be regarded with caution, with the exception of the atypically conserved var2csa, var1 and var3.

Focusing on the identified peptides, we were able to identify peptides as a signature of a PfEMP1 specific domain. Even though the peptide length might seem short, this is equivalent to the length of the PCR products used in the conventional qPCR approaches to assess specific domain expression in field samples [11,21].

In conclusion, we identified PfEMP1 proteins expressed by parasite in patients presenting several forms of malaria. This is one of the first proteomic report of full PfEMP1 protein direct identification and is providing insight towards malaria pathogenesis understanding. The high proportion of CIDRα among the identified sequences enforce the idea that iE sequestration occurs either through CD36 binding, or EPCR binding, pending of clinical presentation [22,47]. We also preferentially identified PfEMP1 protein harbouring DBLβ, among which 20% (6/30 identified DBLβ) displayed the binding pattern for ICAM-1. In addition, the proportion of peptides corresponding to DBLβ was higher in the severe malaria patients compared to the uncomplicated malaria patients. These strengthen the hypothesis that DBLβ is involved in the disease development, as demonstrated with antibodies against DBLβ in Tanzania [48] and Papua New Guinea [49]. However, the technical limitation of bottom-up approach in LC-MS/MS does not allow for an optimal sequence coverage for precise PfEMP1 variants identification.

Both RNAseq and LC-MS/MS analysis showed that *var* and PfEMP1 involved in CM and SA are distinct. This enforce the necessity to study well characterized clinical group. In addition, severe anaemia is a common complication of *P. falciparum* infection in endemic areas [50]. The dedicated *P. falciparum var* and PfEMP1 associated phenotype should be further investigated. However, severe anaemia associated malaria is multi-factorial and the clinical outcome might not be solely related to a dedicated *var*/PfEMP1 subtype.

Our study opens opportunities to identify PfEMP1 variants and later implement these newly identified sequences in PfEMP1 based vaccine development strategies [51,52].

Further studies should include patients from various *P. falciparum* endemic areas to better represent PfEMP1 associated within *P. falciparum* disease in general and specifically to severe malaria.

Supporting information

S1 Table. List of the ENA accession number for the WGS data file.
(DOCX)

S2 Table. List of the *var* transcripts identified using RNAseq for each sequenced sample with the corresponding RPKM values. For each identified transcript, the domain combination obtained using phylogenic analysis is displayed. The signs + and ++ represents respectively the subdomains of DC8 and DC13.
(XLSX)

S3 Table. Sequences of the PfEMP1 proteins identified in LC-MS/MS.
(FASTA)

S4 Table. Structure of the PfEMP1 proteins identified in LC-MS/MS.
(XLSX)

S5 Table. List of the peptides identified per sample. The PEP score reflects the probability of true identification of a given peptide. Domain type has been attributed as described in the methods section.
(XLSX)

Acknowledgments

The authors would like to thank patients who participate in the study, the clinicians and nurses who were involved in patient's inclusion. We especially acknowledge Dr Nadine Fievet help and counselling during the field study. We would like to thank Matt Berriman, Chris Newbold and Thomas Otto from the Welcome Trust Sanger Institute for the access to the unpublished PfEMP1 sequences. We also would like to thank Patrick Mayeux and Emilie-Fleur Gautier for their helpful advices in mass spectrometry and hematology, and Evangeline Bennana for her technical support in sample preparation. The authors thank Antoine Claessens for his help in WGS. The Orbitrap Fusion mass spectrometer was acquired with funds from the FEDER through the "Operational Program for Competitiveness Factors and employment 2007–2013" and from the "Canceropole Ile de France"

Author Contributions

Conceptualization: Claire Kamaliddin, Agnès Aubouy, François Guillonneau, Philippe Deloron, Gwladys I. Bertin.

Data curation: Claire Kamaliddin, Emilie Guillochon, Stéphanie Huguet, Sayeh Guemouri, Céline Peirera, Cédric Broussard.

Formal analysis: Emilie Guillochon, Romain Coppée, Cédric Broussard.

Funding acquisition: Philippe Deloron, Gwladys I. Bertin.

Investigation: Jade Royo, Sem Ezinmegnon, Gino Agbota, Sayeh Guemouri, Jules M. Alao.

Methodology: Claire Kamaliddin, David Rombaut, Romain Coppée, François Guillonneau, Gwladys I. Bertin.

Project administration: Philippe Deloron, Gwladys I. Bertin.

Resources: Stéphanie Huguet, Cédric Broussard, Jules M. Alao.

Software: David Rombaut, Emilie Guillochon.

Supervision: François Guillonneau, Gwladys I. Bertin.

Writing – original draft: Claire Kamaliddin.

Writing – review & editing: Claire Kamaliddin, François Guillonneau, Philippe Deloron, Gwladys I. Bertin.

References

1. Kessler A, Dankwa S, Bernabeu M, Harawa V, Danziger SA, Duffy F, et al. Linking EPCR-Binding PfEMP1 to Brain Swelling in Pediatric Cerebral Malaria. Cell Host Microbe. 2017 Nov 8; 22(5):601–614. e5. https://doi.org/10.1016/j.chom.2017.09.009 PMID: 29107642
2. Leech JH, Barnwell JW, Miller LH, Howard RJ. Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. J Exp Med. 1984 Jun 1; 159 (6):1567–75. https://doi.org/10.1084/jem.159.6.1567 PMID: 6374009

3. Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, et al. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell. 1995 Jul 14; 82(1):89–100. PMID: 7606788
4. Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, et al. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell. 1995 Jul 14; 82(1):77–87. PMID: 7541722
5. Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, Peterson DS, et al. Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell. 1995 Jul 14; 82(1):101–10. PMID: 7606775
6. Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, et al. Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. Wellcome Open Res [Internet]. 2018 May 3 [cited 2019 Jan 17]; 3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5964635/
7. Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes—divide and conquer. PLoS Comput Biol. 2010 Sep 16; 6(9).
8. Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, et al. A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. Proc Natl Acad Sci U S A. 2012 Jun 26; 109(26):E1772–1781. https://doi.org/10.1073/pnas.1120461109 PMID: 22619330
9. Smith JD. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. Mol Biochem Parasitol. 2014 Jul; 195(2):82–7. https://doi.org/10.1016/j.molbiopara.2014.07.006 PMID: 25064606
10. Wahlgren M, Goel S, Akhouri RR. Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. Nat Rev Microbiol. 2017 Jun 12;
11. Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, et al. Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. Proc Natl Acad Sci U S A. 2012 Jun 26; 109(26):E1791–1800. https://doi.org/10.1073/pnas.1120455109 PMID: 22619319
12. Bertin GI, Lavstsen T, Guillonneau F, Doritchamou J, Wang CW, Jespersen JS, et al. Expression of the domain cassette 8 Plasmodium falciparum erythrocyte membrane protein 1 is associated with cerebral malaria in Benin. PloS One. 2013; 8(7):e68368. https://doi.org/10.1371/journal.pone.0068368 PMID: 23922654
13. Baruch DI, Gormely JA, Ma C, Howard RJ, Pasloske BL. Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. Proc Natl Acad Sci U S A. 1996 Apr 16; 93(8):3497–502. https://doi.org/10.1073/pnas.93.8.3497 PMID: 8622965
14. Smith JD, Craig AG, Kriek N, Hudson-Taylor D, Kyes S, Fagan T, et al. Identification of a Plasmodium falciparum intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci U S A. 2000 Feb 15; 97(4):1766–71. https://doi.org/10.1073/pnas.040545897 PMID: 10677532
15. Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JEV, Avril M, et al. Severe malaria is associated with parasite binding to endothelial protein C receptor. Nature. 2013 Jun 27; 498(7455):502–5. https://doi.org/10.1038/nature12216 PMID: 23739325
16. Hess DC, Bhutwala T, Sheppard JC, Zhao W, Smith J. ICAM-1 expression on human brain microvascular endothelial cells. Neurosci Lett. 1994 Feb 28; 168(1–2):201–4. PMID: 7913216
17. Bengtsson A, Joergensen L, Rask TS, Olsen RW, Andersen MA, Turner L, et al. A novel domain cassette identifies Plasmodium falciparum PfEMP1 proteins binding ICAM-1 and is a target of cross-reactive, adhesion-inhibitory antibodies. J Immunol Baltim Md 1950. 2013 Jan 1; 190(1):240–9.
18. Madkhali AM, Alkurbi MO, Szestak T, Bengtsson A, Patil PR, Wu Y, et al. An analysis of the binding characteristics of a panel of recently selected ICAM-1 binding Plasmodium falciparum patient isolates. PloS One. 2014; 9(10):e111518. https://doi.org/10.1371/journal.pone.0111518 PMID: 25360558
19. Lennartz F, Adams Y, Bengtsson A, Olsen RW, Turner L, Ndam NT, et al. Structure-Guided Identification of a Family of Dual Receptor-Binding PfEMP1 that Is Associated with Cerebral Malaria. Cell Host Microbe. 2017 Mar 8; 21(3):403–14. https://doi.org/10.1016/j.chom.2017.02.009 PMID: 28279348
20. Storm J, Jespersen JS, Seydel KB, Szestak T, Mbewe M, Chisala NV, et al. Cerebral malaria is associated with differential cytoadherence to brain endothelial cells. EMBO Mol Med. 0(0):e9164.
21. Mkumbaye SI, Wang CW, Lyimo E, Jespersen JS, Manjurano A, Mosha J, et al. The Severity of Plasmodium falciparum Infection Is Associated with Transcript Levels of var Genes Encoding Endothelial

Protein C Receptor-Binding P. falciparum Erythrocyte Membrane Protein 1. Infect Immun. 2017 Apr 1; 85(4):e00841–16. https://doi.org/10.1128/IAI.00841-16 PMID: 28138022

22. Lau CKY, Turner L, Jespersen JS, Lowe ED, Petersen B, Wang CW, et al. Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. Cell Host Microbe. 2015 Jan 14; 17(1):118–29. https://doi.org/10.1016/j.chom.2014.11.007 PMID: 25482433

23. Jespersen JS, Wang CW, Mkumbaye SI, Minja DT, Petersen B, Turner L, et al. Plasmodium falciparum var genes expressed in children with severe malaria encode CIDRα1 domains. EMBO Mol Med. 2016 Aug 1; 8(8):839–50. https://doi.org/10.15252/emmm.201606188 PMID: 27354391

24. Tuikue Ndam N, Moussiliou A, Lavstsen T, Kamaliddin C, Jensen ATR, Mama A, et al. Parasites Causing Cerebral Falciparum Malaria Bind Multiple Endothelial Receptors and Express EPCR and ICAM-1-Binding PfEMP1. J Infect Dis. 2017 Jun 15; 215(12):1918–25. https://doi.org/10.1093/infdis/jix230 PMID: 28863469

25. Argy N, Bertin GI, Milet J, Hubert V, Clain J, Cojean S, et al. Preferential expression of domain cassettes 4, 8 and 13 of Plasmodium falciparum erythrocyte membrane protein 1 in severe malaria imported in France. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2017 Mar; 23(3):211.e1–211.e4.

26. Carrington E, Otto TD, Szestak T, Lennartz F, Higgins MK, Newbold CI, et al. In silico guided reconstruction and analysis of ICAM-1-binding var genes from Plasmodium falciparum. Sci Rep [Internet]. 2018 Feb 19 [cited 2018 Mar 18]; 8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5818487/

27. Dara A, Drábek EF, Travassos MA, Moser KA, Delcher AL, Su Q, et al. New var reconstruction algorithm exposes high var sequence diversity in a single geographic location in Mali. Genome Med. 2017 Mar 28; 9(1):30. https://doi.org/10.1186/s13073-017-0422-4 PMID: 28351419

28. Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, et al. The Plasmodium falciparum transcriptome in severe malaria reveals altered expression of genes involved in important processes including surface antigen-encoding var genes. PLoS Biol. 2018 Mar; 16(3):e2004328. https://doi.org/10.1371/journal.pbio.2004328 PMID: 29529020

29. Ponts N, Chung D-WD, Le Roch KG. Strand-specific RNA-seq applied to malaria samples. Methods Mol Biol Clifton NJ. 2012; 883:59–73.

30. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006; 7:3. https://doi.org/10.1186/1471-2199-7-3 PMID: 16448564

31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014 Aug 1; 30(15):2114–20.

32. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinforma Oxf Engl. 2012 Dec 15; 28(24):3211–7.

33. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr; 12(4):357–60. https://doi.org/10.1038/nmeth.3317 PMID: 25751142

34. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinforma Oxf Engl. 2015 Jan 15; 31(2):166–9.

35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

36. Bertin GI, Sabbagh A, Guillonneau F, Jafari-Guemouri S, Ezinmegnon S, Federici C, et al. Differential protein expression profiles between Plasmodium falciparum parasites isolated from subjects presenting with pregnancy-associated malaria and uncomplicated malaria in Benin. J Infect Dis. 2013 Dec 15; 208 (12):1987–97. https://doi.org/10.1093/infdis/jit377 PMID: 23901091

37. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods. 2014 Mar; 11(3):319–24. https://doi.org/10.1038/nmeth.2834 PMID: 24487582

38. Gautier E-F, Leduc M, Cochet S, Bailly K, Lacombe C, Mohandas N, et al. Absolute proteome quantification of highly purified populations of circulating reticulocytes and mature erythrocytes. Blood Adv. 2018 Oct 23; 2(20):2646–57. https://doi.org/10.1182/bloodadvances.2018023515 PMID: 30327373

39. Gautier E-F, Ducamp S, Leduc M, Salnot V, Guillonneau F, Dussiot M, et al. Comprehensive Proteomic Analysis of Human Erythropoiesis. Cell Rep. 2016 Aug 2; 16(5):1470–84. https://doi.org/10.1016/j.celrep.2016.06.085 PMID: 27452463

40. Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, et al. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 2002 Sep; 3(3):265–74. https://doi.org/10.1093/bib/3.3.265 PMID: 12230035

41. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15; 30(14):3059–66. https://doi.org/10.1093/nar/gkf436 PMID: 12136088

42. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003 Oct; 52(5):696–704. https://doi.org/10.1080/10635150390235520 PMID: 14530136

43. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016 8; 44(W1):W242–245. https://doi.org/10.1093/nar/gkw290 PMID: 27095192

44. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature. 2002 Feb 7; 415 (6872):673–9. https://doi.org/10.1038/415673a PMID: 11832955

45. Smith JD, Rowe JA, Higgins MK, Lavstsen T. Malaria's deadly grip: cytoadhesion of Plasmodium falciparum-infected erythrocytes. Cell Microbiol. 2013 Dec; 15(12):1976–83. https://doi.org/10.1111/cmi.12183 PMID: 23957661

46. Bertin GI, Sabbagh A, Argy N, Salnot V, Ezinmegnon S, Agbota G, et al. Proteomic analysis of Plasmodium falciparum parasites from patients with cerebral and uncomplicated malaria. Sci Rep. 2016 Jun 1; 6:26773. https://doi.org/10.1038/srep26773 PMID: 27245217

47. Avril M, Bernabeu M, Benjamin M, Brazier AJ, Smith JD. Interaction between Endothelial Protein C Receptor and Intercellular Adhesion Molecule 1 to Mediate Binding of Plasmodium falciparum-Infected Erythrocytes to Endothelial Cells. mBio. 2016 Sep 7; 7(4):e00615–16. https://doi.org/10.1128/mBio.00615-16 PMID: 27406562

48. Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, et al. Hierarchical, domain type-specific acquisition of antibodies to Plasmodium falciparum erythrocyte membrane protein 1 in Tanzanian children. Infect Immun. 2010 Nov; 78(11):4653–9. https://doi.org/10.1128/IAI.00593-10 PMID: 20823214

49. Tessema SK, Utama D, Chesnokov O, Hodder AN, Lin CS, Harrison GLA, et al. Antibodies to Intercellular Adhesion Molecule 1-Binding Plasmodium falciparum Erythrocyte Membrane Protein 1-DBLβ Are Biomarkers of Protective Immunity to Malaria in a Cohort of Young Children from Papua New Guinea. Infect Immun. 2018 Aug; 86(8).

50. WHO. Severe malaria. Trop Med Int Health TM IH. 2014 Sep; 19 Suppl 1:7–131.

51. Bull PC, Abdi AI. The role of PfEMP1 as targets of naturally acquired immunity to childhood malaria: prospects for a vaccine. Parasitology. 2016 Feb; 143(2):171–86. https://doi.org/10.1017/S0031182015001274 PMID: 26741401

52. Hviid L, Lavstsen T, Jensen AT. A vaccine targeted specifically to prevent cerebral malaria—is there hope? Expert Rev Vaccines. 2018 Jun 14;