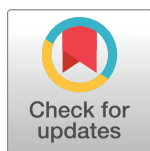


METHODS AND RESOURCES

# The *Plasmodium falciparum* transcriptome in severe malaria reveals altered expression of genes involved in important processes including surface antigen-encoding *var* genes

Gerry Q. Tonkin-Hill<sup>1,2</sup>, Leily Trianty<sup>3</sup>, Rintis Noviyanti<sup>3</sup>, Hanh H. T. Nguyen<sup>4,5</sup>, Boni F. Sebayang<sup>3</sup>, Daniel A. Lampah<sup>6</sup>, Jutta Marfurt<sup>7</sup>, Simon A. Cobbold<sup>8</sup>, Janavi S. Rambhatla<sup>5,9</sup>, Malcolm J. McConville<sup>8</sup>, Stephen J. Rogerson<sup>5,9</sup>, Graham V. Brown<sup>10</sup>, Karen P. Day<sup>4</sup>, Ric N. Price<sup>7,11</sup>, Nicholas M. Anstey<sup>7</sup>, Anthony T. Papenfuss<sup>1,2,12,13,14</sup>\*, Michael F. Duffy<sup>4</sup>✉\*



**1** Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia, **2** Department of Mathematics and Statistics, University of Melbourne, Victoria, Australia, **3** The Eijkman Institute for Molecular Biology, Jakarta, Indonesia, **4** School of Biosciences, Bio21 Institute, University of Melbourne, Melbourne, Victoria, Australia, **5** Department of Medicine and Radiology, Royal Melbourne Hospital, University of Melbourne, Melbourne, Victoria, Australia, **6** Timika Malaria Research Program, Papuan Health and Community Development Foundation, Timika, Papua, Indonesia, **7** Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia, **8** Department of Biochemistry and Molecular Biology, Bio21 Institute, University of Melbourne, Melbourne, Victoria, Australia, **9** Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia, **10** The Nossal Institute for Global Health, University of Melbourne, Parkville, Victoria, Australia, **11** Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom, **12** Peter MacCallum Cancer Centre, Victorian Comprehensive Cancer Centre, Melbourne, Australia, **13** Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia, **14** Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Victoria, Australia

✉ These authors contributed equally to this work.

\* [mduffy@unimelb.edu.au](mailto:mduffy@unimelb.edu.au) (MFD); [papenfuss@wehi.edu.au](mailto:papenfuss@wehi.edu.au) (ATP)

**OPEN ACCESS**

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

**Academic Editor:** David Schneider, Stanford University, United States of America

**Received:** September 25, 2017

**Accepted:** February 16, 2018

**Published:** March 12, 2018

**Copyright:** © 2018 Tonkin-Hill et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The datasets generated and/or analysed during the current study are available in the Arrayexpress repository accession: E-MTAB-5860 (sequenced libraries for each sample) and the European Nucleotide Archive (ENA) repository accession: PRJEB20632 (de novo *var* gene assemblies for combined and individual samples).

## Abstract

Within the human host, the malaria parasite *Plasmodium falciparum* is exposed to multiple selection pressures. The host environment changes dramatically in severe malaria, but the extent to which the parasite responds to—or is selected by—this environment remains unclear. From previous studies, the parasites that cause severe malaria appear to increase expression of a restricted but poorly defined subset of the PfEMP1 variant, surface antigens. PfEMP1s are major targets of protective immunity. Here, we used RNA sequencing (RNA-seq) to analyse gene expression in 44 parasite isolates that caused severe and uncomplicated malaria in Papuan patients. The transcriptomes of 19 parasite isolates associated with severe malaria indicated that these parasites had decreased glycolysis without activation of compensatory pathways; altered chromatin structure and probably transcriptional regulation through decreased histone methylation; reduced surface expression of PfEMP1; and down-regulated expression of multiple chaperone proteins. Our RNAseq also identified novel associations between disease severity and PfEMP1 transcripts, domains, and smaller sequence segments and also confirmed all previously reported associations between expressed PfEMP1 sequences and severe disease. These findings will inform efforts to

**Funding:** National Health and Medical Research Council <https://www.nhmrc.gov.au/> (grant number 1007954) project grant received by MFD and ATP. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. National Health and Medical Research Council <https://www.nhmrc.gov.au/> (grant number 1054618) program grant received by ATP. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. National Health and Medical Research Council <https://www.nhmrc.gov.au/> (grant number 1042072). Practitioner Fellowship received by MMA. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Wellcome Trust <https://wellcome.ac.uk> (grant number 200909). Senior Fellow in Clinical Science received by RNP. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Australian Department of Foreign Affairs and Trade <http://dfat.gov.au>. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRIISS received by ATP. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** acetyl-CoA, acetyl coenzyme A; adj-*p*, adjusted *p*; ATCase, aspartate carbamoyltransferase; ATS, acidic terminal sequence; BCKDH, branched chain ketoacid dehydrogenase; BLAST, basic local alignment search tool; CCT, choline-phosphate cytidyltransferase; CIDR, cysteine-rich interdomain region; CYP52, cytochrome P450 52; CyRPA, cysteine-rich protective antigen; DBL, Duffy binding-like; DC, domain cassette; DHPS, dihydropteroate synthetase; dNTP, deoxyribonucleoside triphosphate; ENA, European Nucleotide Archive; EPCR, endothelial cell protein C receptor; ER, endoplasmic reticulum; FKBP, FK506 binding protein; gDNA, genomic DNA; GO, Gene Ontology; GTP, guanosine triphosphate; GTPase, guanosine triphosphatase; Hb, haemoglobin; HIP, HSP70 interacting protein; HMM, Hidden Markov Model; HOP, hsp70/hsp90 organising protein; HRP2, Histidine Rich Protein 2; ICAM1, intercellular adhesion molecule 1; IE, infected erythrocyte; IQR, interquartile range; KEGG, Kyoto Encyclopedia of Genes and Genomes; LC-MS,

identify vaccine targets for severe malaria and also indicate how parasites adapt to—or are selected by—the host environment in severe malaria.

## Author summary

Infection by *Plasmodium falciparum*—the parasite responsible for malaria in humans—can result in a severe disease that can be fatal or in an uncomplicated disease that can be resolved by the host immune system. However, whether the parasites causing severe disease differ from those causing uncomplicated disease is unknown. Several strands of evidence have suggested that parasites causing severe disease may express a restricted set of the *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) proteins. PfEMP1 proteins are expressed on the surface of the infected red blood cells and elicit protective immunity. We compared the transcriptomes of parasites causing severe and uncomplicated malaria to determine whether these parasites differed in the genes they expressed. We found that the parasites causing severe malaria had altered expression of genes involved in basic metabolism, nuclear processes, and surface expression of PfEMP1. The parasites causing severe malaria had up-regulated expression of a set of PfEMP1 proteins. Some of these PfEMP1s had been previously implicated in severe malaria, lending support to our data. Multiple associations identified between severe malaria and expressed PfEMP1 sequences were novel. These novel, severe disease-associated PfEMP1 sequences could be useful for informing design of vaccines targeting severe malaria disease.

## Introduction

*P. falciparum* is the leading cause of fatal malaria and is responsible for the death of over 400,000 people annually, primarily in sub-Saharan Africa [1]. However, severe disease also occurs in Southeast Asia, and Papua is the Indonesian province with the highest prevalence of malaria [1]. Severe malaria due to *P. falciparum* can manifest as multiple, diverse clinical syndromes [2], but a critical common feature is the sequestration of erythrocytes infected with mature parasites in the microvasculature (reviewed in [3]).

Comparative genome-wide analyses of parasite isolates that cause severe and uncomplicated malaria can be used to identify genes associated with parasite virulence and pathology. This knowledge could inform therapy and the design of vaccines targeting severe disease. Although previous microarray studies of ring-stage—and/or ex vivo—cultivated mature parasites [4,5] showed no differences between parasites causing severe and uncomplicated malaria, transcriptomic differences indicative of fundamental metabolic variations between clinical isolates have been reported. These differences were apparent in clinical isolates segregated by transcriptional profile alone [6] or by direct [7] or surrogate measures of parasitemia [8]. Limitations of these studies included the need to cultivate the isolates prior to analysis, the absence of clinical severity phenotype [8], and the inability to directly compare severe and uncomplicated malaria in the same study population [7]. In the current study, we used massively parallel sequencing technology to undertake comparative analysis of transcriptomes from parasites associated with uncomplicated and severe malaria in the same population. We found a unique parasite transcriptional profile that was associated with severe malaria. While elements of this profile were congruent with reported profiles of clinical isolates [6,8], these have not been previously linked with severe malaria phenotype. Genes deregulated in severe malaria were

liquid chromatography–mass spectrometry; logFC, log fold-change; LPD1, dihydrolipoyl dehydrogenase; LyMP, lysine-rich, membrane-associated PHISTb protein; NDK, nucleoside diphosphate kinase; NDP, ribonucleoside diphosphate; NTP, ribonucleoside triphosphate; NO, nitric oxide; NTS, N-terminal sequence; PC, principal component; PCA, Principal Component Analysis; PfEMP1, *Plasmodium falciparum* Erythrocyte Membrane Protein 1; PfPRMT1, protein arginine N-methyltransferase 1; PfRH5, *Plasmodium falciparum* reticulocyte-binding protein homolog 5; qPCR, quantitative PCR; Q-RT-PCR, quantitative reverse transcription PCR; RNAseq, RNA sequencing; RPKM, Reads Per Kilobase of transcript per Million mapped reads; RQI, RNA Quality Index; RUV, Remove Unwanted Variation; SIAP-2, sporozoite invasion associated protein-2; SNARE, Soluble NSF (N-ethylmaleimide-sensitive factor) Attachment Protein Receptor; TCA, tricarboxylic acid; TM, transmembrane; TMM, trimmed mean of M values; TRAPP5, trafficking protein particle complex subunit 5; uncompl, uncomplicated; UPS, upstream sequence; WHO, World Health Organization.

involved in pathways including central carbon metabolism, folate biosynthesis, histone methylation, chaperone function, and surface expression of *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1).

PfEMP1 is the immunodominant, variant surface antigen of *P. falciparum* [9]. PfEMP1 is expressed on the surface of the infected erythrocyte (IE), where it mediates adhesion to diverse host receptors. PfEMP1 binding to receptors on endothelium leads to the pathogenic sequestration of IEs in the microvasculature (reviewed in [10]). The resulting obstruction is probably exacerbated by IEs binding receptors on uninfected erythrocytes to form ‘rosettes’ [11]. By switching between single, expressed PfEMP1 variants, the parasite can change receptor specificity and also avoid the acquired immune response, leading to chronic and recrudescing infections. A parasite’s genome contains approximately 60 *var* gene copies that code for PfEMP1 [12], and immune pressure has driven evolution of extreme diversity in PfEMP1s such that there is very little overlap in *var* repertoires [13–16].

Even with such large sequence diversity, *var* genes can be classified into three broad groups based on their upstream sequence (UPS; A, B, and C) [12,17]. Group A *var* genes appear to have diverged from groups B and C in their binding properties [18]. Expression of group A and B *var* genes has been associated with clinical malaria in Papua New Guinea [19,20], severe malaria in Africa [21], and cerebral malaria in Africa [22–24]. The PfEMP1 ectodomain contains multiple, semiconserved Duffy binding-like (DBL) domains and cysteine-rich interdomain regions (CIDRs) that mediate adhesion to host receptors [25]. These domains have been classified into major types—DBL $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ , and  $\chi$  and CIDR $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  [14,26,27]—and into 147 further subtypes, e.g., CIDR $\alpha$ 1.1 [14]. Multiple domain cassettes (DCs) containing conserved, sequential arrangements of 2 or more domain subtypes have also been identified [14].

A conserved group of variant surface antigens that are presumably a subset of PfEMP1s appear to be expressed by parasites causing severe disease. These antigens are encountered early in life and are recognised more widely by sera from semi-immune children than antigens expressed by parasites causing uncomplicated disease [28,29]. The expression of a conserved subset of PfEMP1s by parasites that cause severe malaria probably explains why immunity to severe malaria is acquired more rapidly than immunity to uncomplicated malaria [30,31]. The conserved PfEMP1 VAR2CSA is expressed by parasites causing malaria in pregnancy, but associations with entire PfEMP1s have not been detected for other severe malaria disease syndromes.

However, at a finer resolution than whole PfEMP1s, some of the PfEMP1 domains that bind specific host receptors and/or are expressed in severe malaria have been identified. CIDR $\alpha$ 1 binds endothelial cell protein C receptor (EPCR) [32], and its expression has been linked to severe malaria in children and adults [33–35], whilst rosetting is associated with severe malaria and expression of the DBL $\alpha$ 1-CIDR $\beta$ / $\gamma$ / $\delta$  head structure [11,36,37]. DBL $\beta$ 5 from group B *var* genes and specific motifs in DBL $\beta$ 1 and DBL $\beta$ 3 from group A *var* genes bind intercellular adhesion molecule 1 (ICAM1) [38–42], and cerebral malaria has been associated with ICAM1 binding [43,44] and expression of group A carrying tandem CIDR $\alpha$ 1-DBL $\beta$ 1/3 domains [42]. DBL $\beta$ 12 binds the host receptor gC1qR, and its expression is also associated with severe malaria [45]. Elevated expression of a number of DCs has also been associated with severe malaria; these included DC8 (DBL $\alpha$ 2-CIDR $\alpha$ 1.1-DBL $\beta$ 12-DBL $\gamma$ 4/6) [34,45–49], DC13 (DBL $\alpha$ 1.7-CIDR $\alpha$ 1.4) [32,49], DC4 (DBL $\alpha$ 1.4-CIDR $\alpha$ 1.6-DBL $\beta$ 3) [39], DC5 (DBL $\gamma$ 12-DBL $\delta$ 5-CIDR $\beta$ 3/4) [48], and DC6 (DBL $\gamma$ 14-DBL $\zeta$ 5-DBL $\epsilon$ 4) [34].

Due to the immense diversity seen in *var* gene domains, attempts have been made to investigate them by concentrating on more conserved sequence or homology blocks [14,50]. Few studies have attempted to link these conserved blocks with disease severity, although one study

found an association between homology blocks 219 and 486 and rosetting, whereas homology block 204 was associated with impaired consciousness [51].

All of the previously reported associations between severe disease and *var* gene expression relied on PCR using primers derived from *var* sequences of laboratory isolates. In contrast, RNA sequencing (RNAseq) of clinical samples can be used to assemble all expressed *var* sequences, regardless of their homology to the *var* genes of sequenced laboratory isolates. In the current study, innovative bioinformatic approaches were used to identify multiple novel associations between severe disease and differential expression of *var* gene sequences at the multi-, single-, and sub-domain levels. Furthermore, we recapitulated all previously described associations between expressed *var* gene sequences and severe malaria. These novel, severe malaria-associated *var* sequences have relevance to efforts to design vaccines targeting severe disease.

## Results

### Malaria patients

Parasites were isolated from the venous blood of 23 patients with severe malaria and 21 patients with uncomplicated malaria (Table 1). Patients with severe malaria tended to be older than those with uncomplicated malaria, but there were no significant differences in *P. falciparum* density, haemoglobin (Hb) concentration, or gender. Among patients with severe malaria, 19 had presented with a single diagnostic criterion [2], including 4 with cerebral malaria, 3 with jaundice, 8 with hyperparasitaemia, 3 with prostration, and 1 with acute renal failure. Four patients had 2 or more manifestations of severe malaria: 1 patient with jaundice and acute renal failure, 1 with acute renal failure and acute respiratory distress syndrome, 1 with jaundice and hyperparasitaemia, and 1 with hyperparasitaemia and prostration. The parasite biomass marker Histidine Rich Protein 2 (HRP2) was present at higher concentrations in the plasma of patients with severe malaria than those with uncomplicated malaria ( $p = 0.02$ ). None of the patients had severe malarial anaemia (defined as Hb < 5 g/dL in children <12 years old; Hb < 7 g/dL in adults; Table 1) [2]. These findings suggest that severe *P. falciparum* malaria in these patients was associated with sequestration rather than anaemia due to repeat infections [52].

### Preprocessing

RNA quality was assessed using the BioRad Experion system (Fig A in S1 Fig). The median RNA Quality Index (RQI) value was 7.75, and the interquartile range (IQR) was 7.175 to 8.55.

**Table 1. Patient data.**

Variable	Severe malaria median (IQR) <i>n</i> = 23	Uncomplicated malaria median (IQR) <i>n</i> = 21	<i>p</i> (Mann Whitney U test)
Patient age in years	28 (20, 36)	22 (18.5, 24.5)	0.0524
[Hb] g/dL	11.6 (9.6, 13.7)	12.4 (10.6, 14.3)	0.4246
Parasites/ul	39,400 (6,600; 259,378)	30,520 (15,690; 52,800)	0.8726
[Glucose] mg/dL	<sup>§</sup> 124 (101.5, 188.3)	nd	
Gender	M 13 F 10	M 9 F 12	0.5647*
HRP2 ng/ul	848 (244, 2398)	391 (114, 541)	0.0213
Number <i>P. fal</i> mapped reads	5.2e06 (4.8e06, 6.5e06)	6.2e06 (3.9e06, 9.4e06)	0.9

\* Fisher's exact test.

<sup>§</sup> *n* = 12

Abbreviations: Hb, haemoglobin; HRP2, histidine rich protein 2; IQR, interquartile range; nd, not done; *P. fal*, *P. falciparum*.

<https://doi.org/10.1371/journal.pbio.2004328.t001>



Transcriptome libraries were constructed for 44 patient samples (Arrayexpress accession: E-MTAB-5860). Library sizes ranged from 17,054 to 247,859,790 sequence reads (Fig B in [S1 Fig](#)). The libraries were aligned to the *Homo sapiens* (GRCh38), *P. vivax* (PlasmoDB-11.1 Sal1), and *P. falciparum* (PlasmoDB-11.1 3D7) reference genomes, and the proportion of *P. falciparum* in the libraries ranged from 0.11% to 88.44% ([S1 Table](#)). To identify significant features distinguishing severe and uncomplicated malaria transcriptomes, the transcriptome libraries were subjected to a series of sequence and expression analyses (Fig C in [S1 Fig](#)).

### De novo assembly of *var* genes

A pipeline for the de novo assembly of *var* genes from RNAseq data was developed and verified using a *P. falciparum* ItG clone (ItG is the parent line of the It4 sequenced clone) for which the *var* repertoire is known. Expression profiles of the assembled transcripts were compared with those obtained by quantitative PCR (qPCR) and were found to correlate significantly (Pearson correlation coefficient  $R = 0.88$ ) ([Fig 1A](#)). The pipeline used the SoapDeNovo-Trans/Cap3 method of [53], which is robust to chimeric assemblies and minimises redundant transcripts. Non-*var* *P. falciparum*, *P. vivax*, and *H. sapiens* reads were filtered out prior to assembly.

As proof of concept, the pipeline correctly assembled an ItG subclone E8B that expressed predominantly the IT4var04 *var* gene. Additionally, the ItG subclone CS2—with a recombination event between IT4var04 and IT4var08 *var* genes [54]—was correctly assembled (Figs A and B, respectively, in [S2 Fig](#)). Alternative approaches were investigated ([S2 Table](#)), with the SoapDeNovo-Trans/Cap3 pipeline chosen because it assembled the known samples correctly, was sensitive to low-expressed transcripts, and produced minimal redundancy. The pipeline is available at [https://github.com/PapenfussLab/assemble\\_var](https://github.com/PapenfussLab/assemble_var).

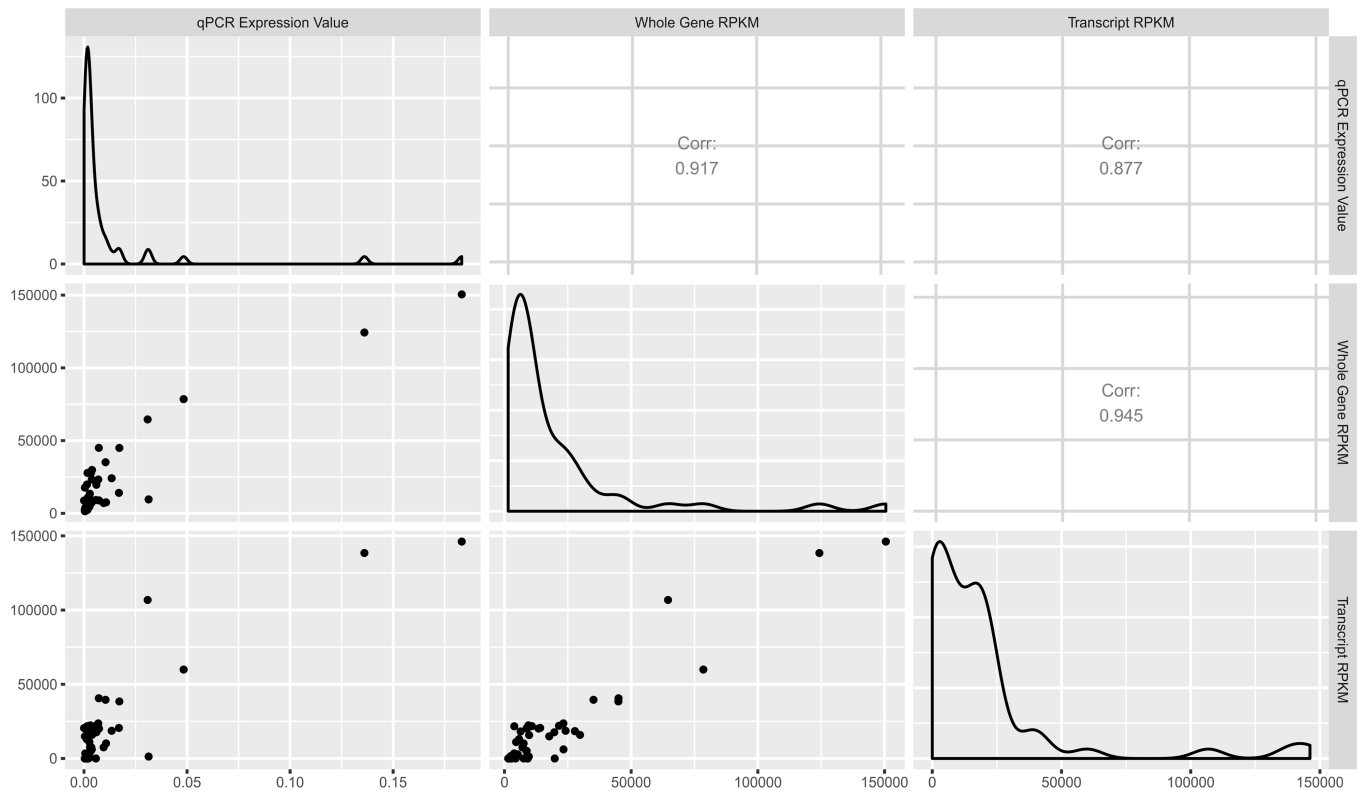
The assembly pipeline was run separately for each of the 44 patient samples in addition to a pooled sample assembly where all the reads from each patient sample were combined (European Nucleotide Archive [ENA] accession: PRJEB20632). [S3 Table](#) indicates the number of assembled transcripts constructed for each sample along with the major N50 and maximum-length values after discarding transcripts shorter than 500 nt in length. For the remainder of this paper, we refer to these 2 assemblies as the separate and combined assemblies, respectively. The assembled *var* genes were analysed at the transcript, domain, and segment or homology block level ([Fig 1B](#)). Three of the severe malaria samples had a low percentage of reads mapping to *P. falciparum*: SFC025, SFD001 (both cerebral malaria), and SFM009 (hyperparasitemia) (Fig B in [S1 Fig](#), [S1 Table](#)). These samples were used for *var* gene assemblies and sequence clustering but were omitted from the differential gene expression analysis, both for *var* and non-*var* genes.

### All gene expression analysis

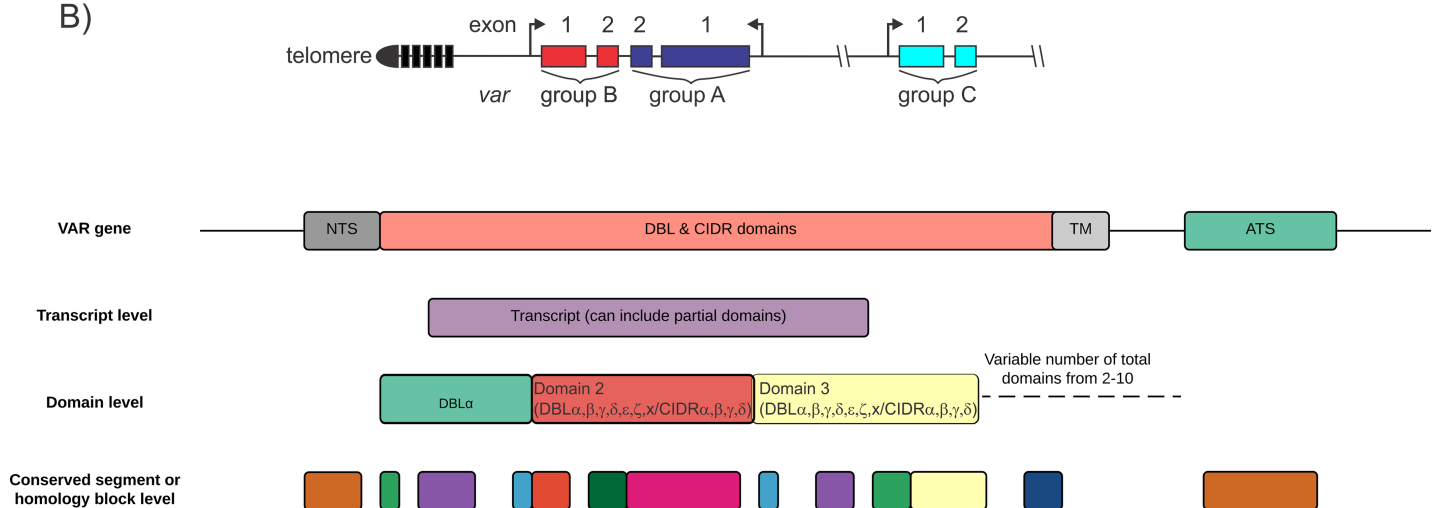
Two patients (SFU2 and SFU3) were drug treated at admission prior to blood collection, and 4 patients (SFC023, SFM007, IFM012, IFM021) were treated with antimalarials for previous *Plasmodium* infections more than 2 weeks but less than 4 weeks prior to admission. These patients were omitted from the differential expression analyses of the total transcriptomes. Significant differences were identified in the expression of genes between severe and uncomplicated cases of malaria. After accounting for library size, parasite life cycle, and other unwanted sources of variation, 358 genes were found to be differentially expressed after multiple testing correction ( $p = 0.1$ , limma/Voom pipeline [55,56]). A full list of genes with relevant log fold changes and  $p$ -values can be found in [S1 Data](#).

A mixture model was used to account for parasite life cycle. A constrained linear model was fit using published data [57] to estimate the proportion of ring, early trophozoite, late

A)



B)



**Fig 1. De novo assembly of *var* gene sequence.** (A) Expression profiles for the ItG subclone E8B. The assembled transcripts were annotated with their closest BLAST match to the IT4 (a clone of the ItG isolate) sequences from the database of [14]. The expression levels in RPKM are then compared to RPKM levels of reads annotated directly to the whole gene DNA sequences of [14] and to those obtained using qPCR. (B) *Var* gene chromosomal arrangements, group A and B *var* genes are present in subtelomeric clusters, group C *var* genes are present in chromosome internal *var* gene clusters. The different resolutions of *var* sequences investigated in this manuscript are illustrated. *Var* gene transcripts are obtained by de novo assembly of transcriptome data. Domain regions are then identified within these transcripts along with smaller subdomain segments and homology blocks. The number and order of both domains and segments varies between *var* genes. ATS, acidic terminal sequence;

BLAST, basic local alignment search tool; CIDR, cysteine-rich interdomain region; DBL, Duffy binding-like; NTS, N-terminal sequence; qPCR, quantitative PCR; RPKM, Reads Per Kilobase of transcript per Million mapped reads; TM, transmembrane.

<https://doi.org/10.1371/journal.pbio.2004328.g001>

trophozoite, schizont, and gametocyte stages present in each sample (Fig 2A, S2 Data). This approach returns similar results to the maximum likelihood approach of [5] and is comparable to the approach of [58], which focused on microarray data. The mixture model correctly identified sample SFC21 as having a higher proportion of gametocytes, a finding that was confirmed by microscopy. Trimmed mean of M values (TMM) normalisation [59] was used to account for library size, with samples SFC025, SFD001, and SFM009 excluded due to insufficient coverage.

The proportion of parasites present at the ring stage—as well as 3 factors of unwanted variation estimated using the R package *ruv* [60]—were used to account for life cycle and other unwanted batch effects. Differential expression testing was conducted using the *limma/Voom* pipeline [55,56]. The impact of including these covariates in the model is evident in the Principal Component Analysis (PCA) plots (Fig 2B and 2C, S2 Data). The choice of covariates strikes a balance between testing power and accounting for unwanted variation. The PCA plots indicate that the outlying SFC21 sample has been accounted for. Furthermore, the separation between the severe and uncomplicated cases shows that, after accounting for variations due to parasite life cycle, significant differences exist between the phenotypes.

### Differences between severe malaria transcriptomes

The severe malaria transcriptomes could be separated by profile of differentially expressed genes into 2 principal clusters—S1 and S2 (S3 Fig), which was consistent with previous reports of clinical isolates and severe malaria [6,7]. This suggests that severe malaria can be caused by parasites in different physiological states. A previous report also found that median parasitemias differed between severe malaria clusters [7]; the median parasitemias in the clusters in this study were also suggestive of a difference ( $p = 0.0755$  Mann Whitney test; parasites/ $\mu$ l median, IQR, S1: 43,040; 5,880; 259,378; S2: 786,316; 212,708; 1,095,789). However, the severe malaria transcriptomes did not cluster by clinical syndrome (Fisher's exact test, all  $p > 0.12$ ).

### Analysis of differential gene expression

The 358 genes differentially expressed in severe malaria were from diverse functional pathways and revealed a distinct severe malaria parasite transcriptome (S3 Fig, S1 Data). Biological pathways annotated as Gene Ontology (GO) biological process terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) were ranked using a hypergeometric test, and those with  $p < 0.1$  are considered. Terms relating to glycolysis, histone methylation, folate metabolism, and protein folding ranked highly and included genes down-regulated in severe malaria, whilst pathways relating to the tricarboxylic acid (TCA) cycle, nucleoside diphosphate (pyrimidine) metabolism, and regulation of guanosine triphosphatase (GTPase) activity included genes up-regulated in severe malaria (Fig 3A, S3 Data). In addition, genes involved in PfEMP1 transport and a gene involved in regulation of *var* genes were down-regulated in severe malaria. This suggested that *var* gene expression was modulated but PfEMP1 surface presentation was reduced. Several GO and KEGG pathways that ranked highly included deregulated genes that were not functionally related in a coherent manner and will not be discussed further.

### Parasite carbon metabolism

Parasites isolated from patients with severe malaria had significantly down-regulated genes included in the KEGG pathway 'Glycolysis/gluconeogenesis'. Significant decreases were





the outlying position of sample SFC21, which has been identified as having more late-stage and gametocyte parasites. (C) A PCA plot of read counts normalised for library size, staging effects, and other unwanted batch effects using the novel mixture model along with 3 unwanted factors of variation estimated by RUV4 (normalised read counts are available in [S2 Data](#)). Sample SFC21 has been appropriately dealt with and a better separation of the samples by disease phenotype can be observed. PC, principal component; PCA, principal component analysis; RUV, Remove Unwanted Variation.

<https://doi.org/10.1371/journal.pbio.2004328.g002>

observed in transcript levels of 3 glycolytic enzymes (0.52- to 0.62-fold the levels in parasites that caused uncomplicated malaria) (all adjusted  $p < 0.1$ ) ([Fig 3B](#), [Fig 3A](#), [S1 Data](#), [S3 Data](#)). These were aldolase, glyceraldehyde 3-phosphate dehydrogenase, and mitochondrial dihydrolipoyl dehydrogenase (LPD1) that converts glycolytic pyruvate to acetyl coenzyme A (acetyl-CoA). Expression of most other enzymes in this pathway trended down (with 3 adjusted  $p \leq 0.12$ ) ([Fig 3B](#)). The lactate transporter (also known as the formate nitrite transporter [61]) was also down-regulated in parasites causing severe malaria (0.33-fold  $p = 0.009$ ). Together, these data suggest that parasites associated with severe malaria have decreased transcription of genes involved in aerobic glycolysis.

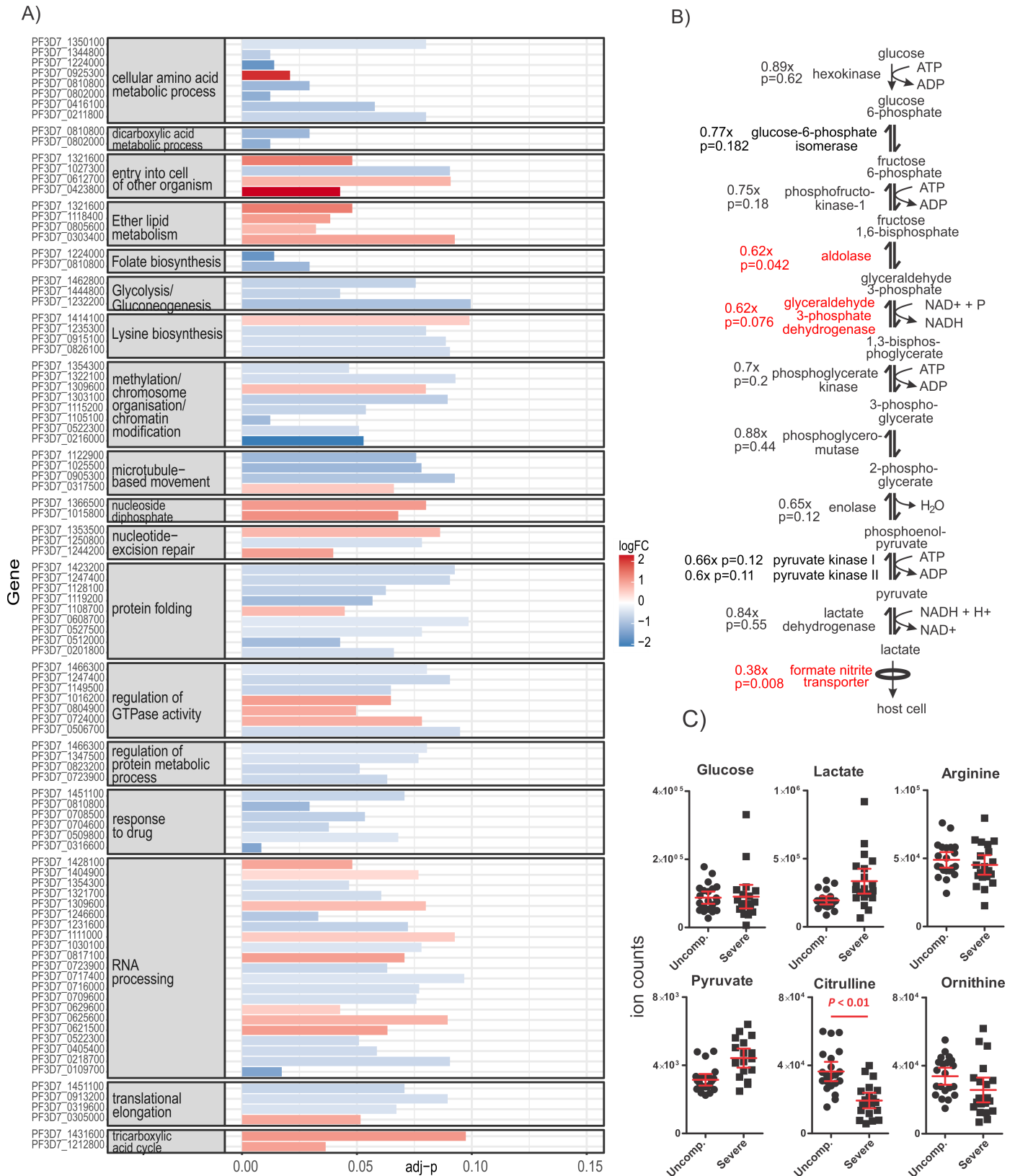
Our results confirm and extend previous analyses on the transcriptional regulation of enzymes involved in central carbon metabolism in clinical isolates [8]. In particular, Daily et al. [6] described a cluster of *P. falciparum* clinical isolates that exhibited a distinct, starvation-like response, characterised by decreased transcription of genes involved in glycolysis and increased transcription of genes encoding enzymes involved in the TCA cycle, which is similar to the transcriptional signature we observed from parasites linked to severe malaria samples ([S4 Table](#)). In contrast, parasites isolated from patients with severe malaria in a subsequent study had a transcriptional profile that was more consistent with a glycolytic phenotype [7]. However, neither study directly compared transcriptional profiles from parasites causing severe and uncomplicated malaria [6,7].

### Severe malaria patients' metabolic profile

To determine whether nutrient availability was contributing to the reduced expression of genes encoding glycolysis enzymes by the parasites causing severe malaria, metabolite levels in plasma samples from the severe and uncomplicated malaria patients were analysed by untargeted liquid chromatography–mass spectrometry (LC-MS) analysis. Thirty-five metabolite peaks differed significantly between the plasma of patients with severe and uncomplicated malaria ( $p < 0.01$ , Benjamini-corrected; [S4 Data](#)). These included 7 metabolites—provisionally identified as lipids—and citrulline (confirmed with an authentic standard), which was depleted in the patients with severe malaria. Citrulline recycling to arginine contributes significantly to nitric oxide (NO) synthase substrate availability and thereby NO bioavailability in malaria. Low citrulline is therefore likely to contribute to the hypoargininemia, impaired NO bioavailability and endothelial dysfunction found in both adults and children with severe malaria, in both Melanesian [62] and African [63–65] populations. The plasma levels of glucose and lactate were similar in patients with uncomplicated and severe malaria ([Fig 3C](#)), suggesting that the down-regulation of parasite glycolysis in the patients with severe malaria is not a direct response to reduced availability of blood glucose. Blood glucose concentrations were also similar in individuals harboring parasites with or without the proposed starvation transcriptional pattern described by Daily et al. [6].

### Alternative pathways of carbon metabolism

Glucose-starved yeast [66] and clinical *P. falciparum* isolates with the proposed starvation response-like transcriptome both increased transcription of TCA cycle enzymes [6]. Similarly,



**Fig 3. Gene sets enriched in deregulated genes in severe malaria.** (A) Summary of highly ranked GO and KEGG gene annotation pathways that included significantly deregulated genes in severe malaria. Only gene sets that contained more than 1 deregulated gene are shown; deregulated gene set data available in [S3](#)

Data, deregulated genes available in [S1 Data](#). (B) The glycolysis pathway in *P. falciparum* in severe malaria. Fold-change in gene expression in severe malaria relative to uncomplicated malaria (x) and *p*-value for the fold-change are indicated beside genes. Genes that were significantly (adjusted  $p < 0.1$ ) down-regulated in severe malaria are indicated in red. (C) LC-MS metabolomic analysis of plasma samples from patients with severe and uncomplicated malaria. Ion counts for metabolites commonly affected by malaria are presented; data available in [S4 Data](#). adj-p, adjusted *p*; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LC-MS, liquid chromatography–mass spectrometry; logFC, log fold-change; uncompl, uncomplicated.

<https://doi.org/10.1371/journal.pbio.2004328.g003>

in severe malaria, the GO category ‘tricarboxylic acid cycle’ included 2 genes up-regulated more than 2-fold in severe malaria (both adjusted  $p < 0.097$ ): the Fe<sub>2</sub>S subunit of the mitochondrial TCA cycle enzyme succinate dehydrogenase and the putative succinyl CoA synthetase β subunit. Aconitase was also up-regulated more than 1.6-fold (adjusted  $p = 0.156$ ); however, no significant differences in expression of the other TCA cycle enzymes were observed ( $p$ -values  $> 0.2$ ). We have previously shown that *P. falciparum* asexual blood stages primarily sustain TCA cycle fluxes and low-level oxidative phosphorylation by catabolizing glutamine [67] up-regulating the TCA cycle. However, key enzymes in glutamine utilisation were either down-regulated (glutamate dehydrogenase down 0.4-fold,  $p = 0.012$ ) or unchanged (NADP-specific glutamate dehydrogenase, aspartate transaminase, glutamate synthase, malate dehydrogenase, phosphoenolpyruvate carboxylase, and branched chain ketoacid dehydrogenase complex [BCKDH] subunits E1β and E2) in isolates from patients with severe malaria, indicating that these parasites are unlikely to exhibit a significant switch to mitochondrial respiration. Overall, these data suggest that parasites associated with severe malaria were not compensating for decreased glycolysis by increasing oxidation of pyruvate in the TCA cycle and may be metabolically less active than parasite isolates associated with uncomplicated malaria.

### Methylation and lysine degradation

The GO term ‘methylation’ and a number of subsidiary GO terms relating to histone methylation included genes down-regulated in parasites causing severe malaria. The down-regulated genes included a putative histone S-adenosyl methyltransferase and 2 of the 10 SET-domain lysine methyl transferases found in *P. falciparum* (SET2 and PfSET7). Levels of SET3 and a putative protein arginine N-methyltransferase 1 (PfPRMT1) were also suggestive of down-regulation (both adjusted  $p < 0.11$ ,  $< 0.73$ -fold). PfSETvs or SET2 plays an important role in regulating expression of *var* genes (see below). PfSET3 and PfSET7 appear essential for blood-stage growth [68], and PfSET7 can methylate H3 but is localised to the cytoplasm in asexual blood stages [69]. PfPRMT1 probably methylates cytoplasmic and nuclear proteins including methylations of histone 4 that are involved in gene activation [70]. These data suggest that histone methylation pathways involved in gene regulation were down-regulated in severe malaria. Genes involved in chromatin modification were also deregulated by parasites previously reported to have caused high parasitaemia infections [8]. Severe malaria is known to elicit gametocytogenesis, and heterochromatin structure dependent on histone methylation is known to repress the gametocytogenesis transcription factor ApiAP2G [71], so down-regulation of histone methylation would be consistent with induction of gametocytogenesis in severe malaria.

### Folate and nucleoside metabolism

The KEGG term ‘folate biosynthesis’ and a number of related GO terms included 2 genes down-regulated in severe malaria: dihydropteroate synthetase (DHPS) and guanosine triphosphate (GTP) cyclohydrolase. GTP cyclohydrolase is the first and rate-limiting enzyme in the folate pathway and therefore is essential for DNA and protein synthesis. Aspartate

carbamoyltransferase (ATCase) was also down-regulated; it is the second enzyme in the pyrimidine biosynthetic pathway; and whether it is rate limiting in *P. falciparum* is unknown, but it is so in bacteria [72]. These changes suggest that nucleoside biosynthesis may be decreased in severe malaria, consistent with lower growth rate and/or reduced metabolism. A number of GO pathways related to nucleoside diphosphate and pyrimidine metabolism included 2 genes up-regulated in severe malaria: nucleoside diphosphate kinase (NDK) and the putative small subunit of ribonucleotide reductase. These 2 genes are central to ribonucleoside triphosphate (NTP) and deoxyribonucleoside triphosphate (dNTP) synthesis. Although up-regulation of these enzymes suggests increased DNA synthesis, the down-regulation of key enzymes in the folate and pyrimidine pathways instead indicates that a diminished nucleoside pool is subject to increased flux through ribonucleoside diphosphate (NDP) to dNTP metabolism.

### Translation and protein folding

The GO term ‘translational elongation’ included 3 down-regulated elongation factor genes suggesting decreased protein production. The GO term ‘protein folding’ included 9 down-regulated genes, including the HSP70 interacting protein (HIP), the peptidyl-prolyl *cis-trans* isomerase cytochrome P450 52 (CYP52)—which has in vitro chaperone activity (Marin-Menendez, 2012)—an FK506 binding protein (FKBP)-type peptidyl-prolyl isomerase, and the PfEMP1 transport-associated KAHsp40. Functionally related proteins outside this pathway were also down-regulated, including the hsp70/hsp90 organising protein (HOP), HSP70-x (see below), and the essential PfHsp110c, which is important for preventing heat-induced aggregation of the many *P. falciparum* Asn repeat rich proteins during fever [73]. Overall, down-regulation of these genes indicated a decreased stress response or generalised, decreased protein processing.

### Regulation of GTPase activity

The GO category ‘regulation of GTPase activity’ and related GO categories included 3 GTPase-activating protein genes that were up-regulated in severe malaria, 2 of which were specific for Rab GTPases. This would be consistent with decreased Rab GTPase trafficking regulatory activity and therefore decreased vesicular transport. Two genes involved in vesicle transport were down-regulated; these were SNAP proteins, which is involved in dissociation of the Soluble NSF (N-ethylmaleimide-sensitive factor) Attachment Protein Receptor (SNARE) complex, and choline-phosphate cytidylyltransferase (CCT), which is rate limiting for synthesis of the major *P. falciparum* membrane phospholipid, phosphatidylcholine [74]. Four genes involved in vesicle transport were up-regulated. These included 2 proteins involved in endoplasmic reticulum (ER) to Golgi transport, the trafficking protein particle complex subunit 5 (TRAPPC5), and the SNARE protein PfGS27; the retrieval receptor for ER membrane proteins, which is required for anterograde vesicular transport; and the vacuolar protein sorting-associated protein 45 that is implicated in vesicle transport from the Golgi to endosomes or the food vacuole. Overall, the probable decreased trafficking activity of several Rabs and deregulated vesicle transport processes suggest deregulated protein trafficking in severe malaria.

### PfEMP1 and *var* regulation

Multiple genes involved in PfEMP1 biology were down-regulated in severe malaria. These included PfSETvs—which methylates lysine 36 on histone 3, is required for *var* gene silencing [68], and is involved in normal *var* switching [75]. The knob-localised KAHRP—which binds PfEMP1 and the cytoskeleton—and the lysine-rich, membrane-associated PHISTb protein (LyMP) (PF3D7\_0532400) are both required for optimal binding of PfEMP1 to (some)



receptors [76] and were amongst the most down-regulated genes in severe malaria. Also down-regulated were the following: the Maurers cleft proteins SBP1 and REX1, which are required for proper Maurer's cleft organization and PfEMP1 transport to the erythrocyte surface [77–79]; Heat shock protein 70-x, which forms a complex with Hsp40 in the red blood cell cytosol and is possibly involved in PfEMP1 transport [80]; and KAHsp40, which binds PfEMP3 and KAHRP and colocalises with knob-associated proteins [81]. Therefore, we observed probable mechanistic drivers of modulated *var* regulation and decreased transport of PfEMP1 to the parasite surface. GO categories that were highly ranked due primarily to inclusion of deregulated 3D7 *var* genes were not reported because 3D7 *var* genes were not present in the clinical isolates.

## Surface proteins

Several parasite surface proteins with established functions unrelated to ring-stage parasites were highly up-regulated in severe malaria. The second most up-regulated gene encoded the glycosylphosphatidylinositol-anchored cysteine-rich protective antigen (CyRPA) that anchors the critical invasion protein *Plasmodium falciparum* reticulocyte-binding protein homolog 5 (PfRh5) to the surface of the merozoite [82]; the seventh most up-regulated gene was the merozoite surface-located 6-cysteine protein P41 [83], and the 14th most up-regulated gene was sporozoite invasion-associated protein-2 (SIAP-2). This gene is expressed at low levels in blood-stage cultures but at high levels on the surface of sporozoites, and it appears to be important for hepatocyte traversal [84]. The serpentine receptor 10 was also up-regulated. It is most closely related to receptors that transduce external stimuli in other organisms [85].

## *var* gene expression analysis

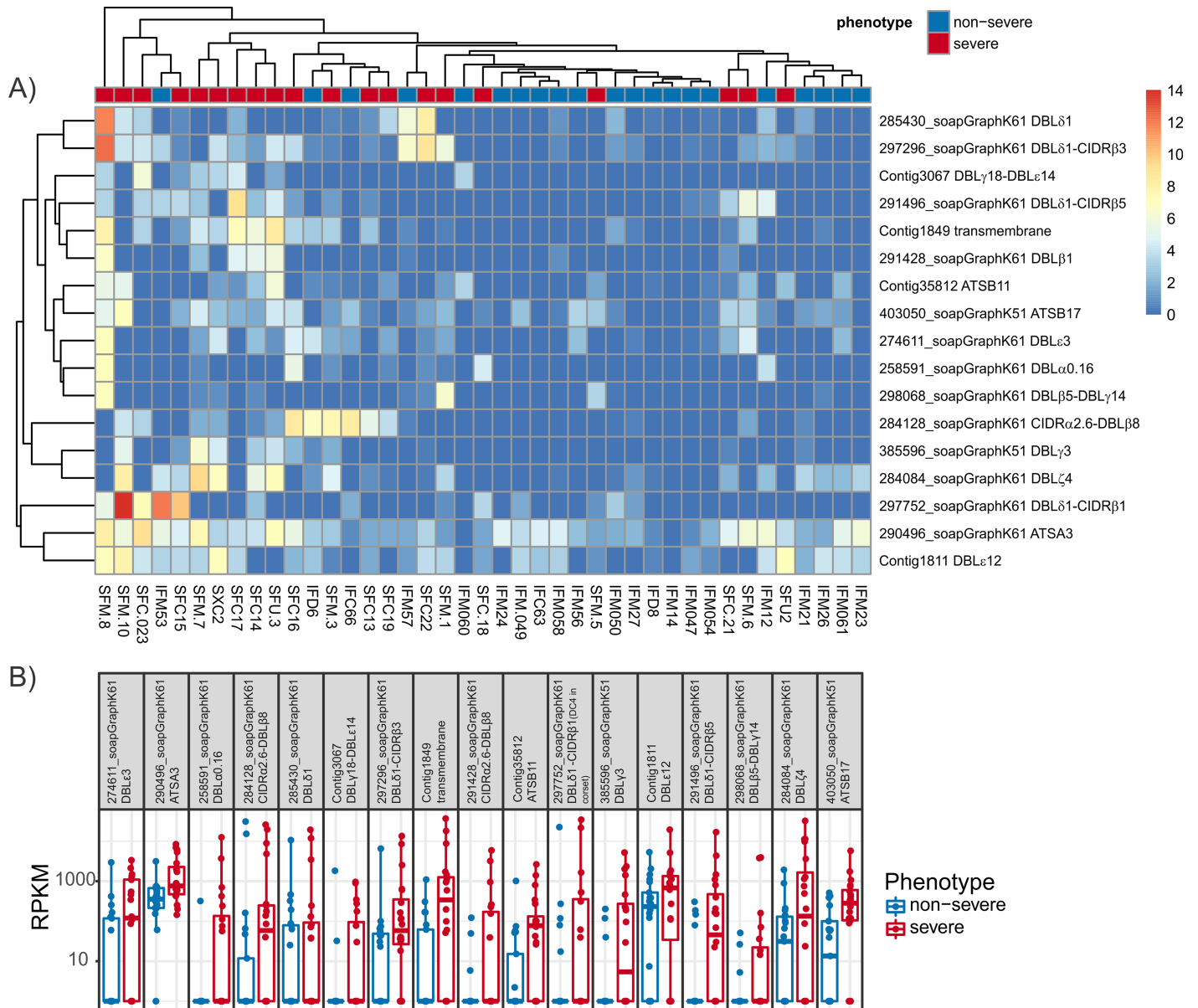
There was no difference between severe malaria and uncomplicated malaria in total *var* gene expression, i.e., the number of reads that mapped to de novo-assembled *var* genes (normalised for number of total reads that mapped to all genes; Welch 2-sample *t* test,  $p = 0.28$ ). Differential expression analysis was conducted at the *var* multidomain transcript, individual domain, and segment levels because associations between *var* expression and severe disease have been previously detected separately at each of these resolutions. At each level, significant, differentially expressed sequences were identified. Additionally, the resulting sequence transcripts, domain classification, and segments were found to better distinguish severe and nonsevere cases of malaria than previous *var* gene classifications [14,39,46,48].

Fig B in S4 Fig illustrates a PCA plot of normalised read counts annotated to the transcripts from the combined sample *var* gene assembly. By comparing it to the all-gene PCA plot (Fig 2C), it is evident that *var* gene expression differentiates severe cases of malaria. The severe cases are more tightly clustered together than the nonsevere.

S5 Data lists all the separate sample assembly transcripts along with whether they were significant at the transcript, domain, or segment level. A number of transcripts had domains and segments that were significantly associated with disease severity when the transcript itself was found not to be significantly associated with severe disease. This highlights the importance of investigating the *var* gene sequences at multiple resolutions.

## Transcript level

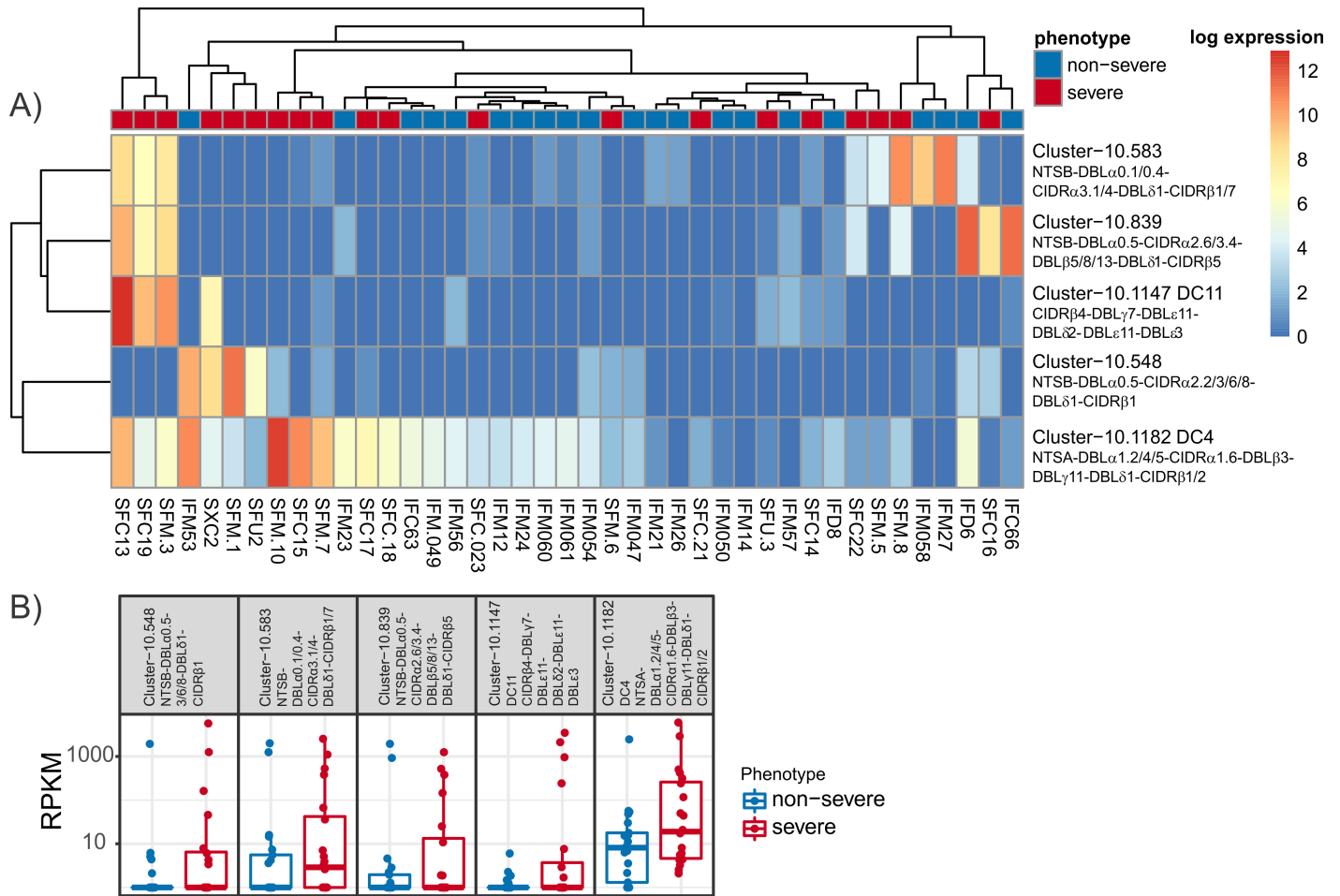
In the combined sample assembly (S6 Data), 53 transcripts were found to be differentially expressed using the default DESeq2 pipeline [86]. Of these, 17 are up-regulated in severe malaria ( $p < 0.05$ ) (Fig 4A). The expression profiles of the up-regulated transcripts from the combined assembly differentiated the samples based on severity (Fig 4A, S4 Fig, panel B). Amongst the transcripts up-regulated in severe malaria, the extracellular domains most highly



**Fig 4. Analysis of RNAseq data at the level of *var* gene transcripts: Combined assembly.** (A) Expression levels of transcripts from the combined sample assembly found to be up-regulated in severe disease. Samples and clusters have been grouped using complete linkage hierarchical clustering (raw read counts available in [S6 Data](#)). (B) Expression levels of transcripts from the combined sample assembly found to be up-regulated in severe disease; values for all samples and the IQR and median are indicated. RPKM is reads per kb of transcript per million reads mapped to total *var* transcripts (RPKM available in [S6 Data](#)). IQR, interquartile range; RPKM, Reads Per Kilobase of transcript per Million mapped reads.

<https://doi.org/10.1371/journal.pbio.2004328.g004>

expressed in severe malaria were a DBL $\zeta$ 4 (284084\_soapGraphK61), a DBL $\epsilon$ 3 (274611\_soapGraphK61), and a DBL $\epsilon$ 12 (Contig1811) (Fig 4B). The up-regulated transcripts included a transcript that contained DBL $\beta$ 5-DBL $\gamma$ 14 (298068\_soapGraphK61); DBL $\gamma$ 14 has only been found in DC6 [14], and its expression was recently associated with severe disease [34]. In 7 *P. falciparum* genomes [14], the tandem combination DBL $\beta$ 5-DBL $\gamma$ 14 was detected only in the 3D7 gene PFL0020w, which is expressed by 3D7 parasites selected for adhesion to ICAM1 [87]. Another of the up-regulated transcripts contained DBL $\gamma$ 18-DBL $\epsilon$ 14 (Contig3067); this tandem domain arrangement was only detected twice in the 7 sequenced genomes but was not part of any DCs.



**Fig 5. Analysis of RNAseq data at the level of *var* gene transcripts: Separate assembly.** (A) Expression levels of clusters identified by Corset found to be up-regulated in severe disease. Samples and clusters have been grouped using complete linkage hierarchical clustering. Raw read counts are available in S7 data. (B) Expression levels of clusters identified by Corset found to be up-regulated in severe disease. Values for all samples and the IQR and median are indicated and are available in S7 Data. RPKM is reads per kb of transcript per million reads mapped to total *var* transcripts. IQR, interquartile range; RPKM, Reads Per Kilobase of transcript per Million mapped reads.

<https://doi.org/10.1371/journal.pbio.2004328.g005>

The remaining transcripts were either single domains or common tandem domain arrangements. A transcript incorporating DC5 (DBLδ5-CIDRβ3-DBLβ7-[DBLγ4]) (contig12688) was also up-regulated in severe malaria ( $p = 0.0537$ ). DC5 was up-regulated in severe malaria in Africa [48] and expressed in a cerebral malaria case in Papua New Guinea [88].

Corset [89] groups transcripts together based on the number of reads that multi-map between them whilst ensuring transcripts are not combined if they have significantly different expression profiles. We used Corset to detect transcripts associated with severe disease in the separate sample assemblies. Associations between severe disease and *var* transcripts can be inferred with greater confidence if identified using multiple approaches. Corset identified 82 differentially expressed clusters in total, of which 5 were clearly up-regulated in severe disease (Fig 5A, S7 Data). These clusters included overlapping, multidomain contigs that spanned DC4 (N-terminal sequence A [NTSA]-DBLα1.2/1.5/1.4-CIDRα1.6-DBLβ3-[DBLγ11-DBLδ1-CIDRβ1/2]) (cluster-10.1182) and DC11 (CIDRβ4-DBLγ7-DBLε11-DBLζ2-DBLε11-DBLε3) (cluster-10.1147). The contigs spanning DC4 were the most abundantly expressed of the clustered contigs up-regulated in severe malaria (Fig 5B). DC4 expression has previously been

associated with severe malaria [39], and the DC4 cluster included 2 DC4 transcripts from the cerebral malaria sample SFC15. For each transcript in the separate assembly, its closest basic local alignment search tool (BLAST) [90] hit in the combined assembly was identified. Of the 5 Corset clusters up-regulated in severe malaria, 2 included transcripts with their closest BLAST hit in the 17 up-regulated transcripts from the combined assembly. These were the DC4 cluster—which was homologous to the DBL $\delta$ 1-CIDR $\beta$ 1 combined assembly transcript 297752\_soapGraphK61—and cluster-10.839 (N-terminal sequence B [NTSB]-DBL $\alpha$ 0.5-CIDR $\alpha$ 2.6/3.4-DBL $\beta$ 5/8/13-DBL $\delta$ 1-CIDR $\beta$ 5)—which was homologous to the combined assembly transcript 284128\_soapGraphK61 (CIDR $\alpha$ 2.6-DBL $\beta$ 8) (Fig 6). The 2 remaining clusters contained transcripts spanning NTSB-DBL $\alpha$ 0.1/0.4-CIDR $\alpha$ 3.1/4-DBL $\delta$ 1-CIDR $\beta$ 1/7 (cluster-10.583) and NTSB-DBL $\alpha$ 0.5-CIDR $\alpha$ 2.2/2.3/2.6/2.8-DBL $\delta$ 1-CIDR $\beta$ 1 (cluster-10.548). These 2 clusters and the DC4 and DC11 clusters were all homologous to additional transcripts that were up-regulated in the combined assembly at an adjusted *p*-value of no more than 0.153 (Fig 6). The elevated *p*-values in the combined assembly analysis can be explained by the heavier multiple testing penalty due to the larger number of transcripts.

### *var* UPS type level

For simplicity, we restricted our analysis at the type level to distinguishing between UPS types A and B/C combined. Expression of the conserved NTS segments allows for these 2 groups to be identified. HMMER3 [91] was used to align the profile hidden Markov models of the domains defined in [14] to the transcripts built from the separate assemblies. Reads that aligned to the regions annotated as either NTSA or NTSB were then used as counts for the respective *var* types. NTSA was more highly expressed in the severe malaria samples than in the uncomplicated malaria samples (Fig A in S4 Fig). This is consistent with previous studies [19–23,34,92].

### Domain level

The domain models of [14] were first investigated using the same approach as the type-level analysis. Of the 149 domain classifications identified in the transcripts, 16 were found to be significantly up-regulated in severe malaria using the default pipeline of DESeq2 [86] (S8 Data, Fig 7A). Some previously described associations between expressed *var* sequences and severe malaria were confirmed, adding confidence to our analysis. These included up-regulation in severe malaria of CIDR $\alpha$ 1.1 and CIDR $\alpha$ 1.6, which bind EPCR [35] and are often found in DC8 and DC4, respectively [32,34,45,46,48,49]; DBL $\alpha$ 2, which is restricted to DC8; and DBL $\beta$ 12, which binds gC1qR [45] and is invariably found in DC8. DBL $\beta$ 3 was also up-regulated; it can bind ICAM1 and is found in—but is not restricted to—severe malaria-associated DC4 [39] and DC8. The domain subtypes DBL $\beta$ 3, CIDR $\alpha$ 1.6, DBL $\alpha$ 1.2, DBL $\alpha$ 1.5, and DBL $\gamma$ 11 that were all up-regulated in this analysis were also part of the up-regulated DC4 transcript in the Corset analysis (Fig 6, Fig 5A, S5 Data, S7 Data). NTSA that is restricted to group A *var* genes was also up-regulated. Despite these clear differences, many of the domains were still expressed in a large number of the uncomplicated samples, e.g., DBL $\beta$ 3 was far more abundantly expressed than CIDR $\alpha$ 1.6 and DBL $\alpha$ 1.5, but the latter 2 domains were more clearly differentially expressed in severe malaria (Fig 7A, Fig 8B). Comparing the domain-level PCA plot in Fig 7B with the transcript-level PCA plot in Fig B in S4 Fig showed that the differentiation between severe and uncomplicated malaria samples was less evident at the domain level and suggested that a more accurate classification could be made.

A novel hierarchical approach was developed to identify domains that are associated with severe malaria. Domain regions were first defined using HMMER3 domain models based on



	COMBINED TRANSCRIPTS <sup>a</sup>	CORSET CLUSTERED SEPARATE ASSEMBLY TRANSCRIPTS	DOMAIN HMM	DOMAIN% ID	BLOCK HMM	SEGMENTS %ID
	DBLδ1 <sup>a</sup> CIDRβ7 <sup>b</sup>	NTSB DBLα0.1/0.4 CIDRα3.1/4 DBLδ1 CIDRβ1/7				
	CIDRα2.6/8 DBLβ8/10	NTSB DBLα0.5 CIDRα2.6/3.4 DBLβ5/8/13 DBLδ1 CIDRβ5				
DC11	DBLε3 DBLε11 <sup>b</sup> DBLζ2 <sup>b</sup> DBLε11 <sup>b</sup> DBLε3 <sup>b</sup>	CIDRβ4DBLY7 DBLε11 DBLζ2 DBLε11 DBLε3	DBLε3	DBLε3		0.8_DBLε_block8
	DBLδ1 <sup>a</sup> CIDRβ1 <sup>b</sup>	NTSB DBLα0.5 CIDRα2.2/2.3 /2.0/2.8 DBLδ1 CIDRβ1				
DC4	DBLδ1 CIDRβ1 DBLα1.4 <sup>b</sup> CIDRα1.6 <sup>b</sup> DBLβ <sup>b</sup>	NTSA DBLα1.2/1.4/1.5 CIDRα1.6 DBLβ3 DBLY11 DBLδ1 CIDRβ1/2	NTSA DBLα1.2 DBLα1.5 CIDRα1.6 DBLβ3 DBLY11		HB219 HB121 HB150 HB141 HB183	NTSA_block4 CIDRα_block6 DBLβ_block8 DBLβ_block4
DC9?	DBLY3/5 DBLζ4			DBLY3 DBLζ4	HB582	DBLY_block2 DBLζ_block15 DBLζ_block12
						CIDRY_block4 DBLδ_block11
DC8			DBLα2 CIDRα1.1 DBLβ12	DBLδ1 DBLβ12		DBLβ_block4
				DBLε9 DBLε2		DBLε_block3 DBLε_block1
				DBLδ1 <sup>c</sup> DBLδ1 <sup>c</sup> CIDRβ1 <sup>c</sup> DBLβ3 <sup>c</sup>		DBLβ_block5
DC6	DBLβ5 DBLY14					DBLY_block7
var2				DBLεPAM5	HB97	
var1			DBLε5		HB126 HB142	DBLε_block6 0.6_DBLε_block8
	ATSB11 ATSB17				HB47	
	DBLα0.16					0.9_DBLα_block5
	DBLδ1 CIDRβ3 DBLY4/18 DBLε4/14 DBLδ1 CIDRβ5 DBLδ1 DBLε12 DBLβ1 ATSA3 TM		CIDRY9 DBLζ3 DBLβ13 CIDRY12 DBLδ7	CIDRα2.4 DBLδ1		0.6_DBLα_block5 DBLY_block5 DBLα_block6 DBLα_block1
DC5	DBLδ5 CIDRβ3 DBLβ7 DBLY4					

**Fig 6. Summary of PfEMP1 transcripts, domains, and segments that were up-regulated in severe malaria.** Sequences up-regulated in severe malaria are organised in columns for each analysis method separated by grey bars. Multiple domains found in the same single transcripts from the combined or separate assemblies are on a single row. Closely related sequences found in multiple analyses are colour coded for each of the major domain types and are grouped together across analyses by unbroken horizontal lines. Domains and/or segments that clustered together by expression profile in multiple individuals within a single analysis are also grouped by unbroken horizontal lines. Grey shaded sequences at the bottom of the diagram are unrelated to each other. For example, in the case of DC4, 2 transcripts from the combined assembly were amongst the closest BLAST hits to the DC4-like transcripts from the CORSET cluster of the separate assembly; 6 domains and 5 blocks identified by HMM in the separate assembly are found in DC4 domains; and clusters for 1 domain and 4 segments identified by hierarchical analysis contained DC4 domain sequences, including those from the DC4-like transcripts from the CORSET cluster of the separate assembly. <sup>a</sup>Combined assembly transcripts up-regulated in severe malaria were all adjusted  $p < 0.05$  except for domains marked <sup>b</sup> (adjusted  $p < 0.153$ ). Domains HMM and blocks HMM were identified using the HMM of [14]. Domains and segments %ID were identified using the novel hierarchical approach developed for this study. <sup>c</sup>Non-DC8-like DBLδ1 and non-DC4-like DBLβ3 that clustered by expression profile in the same patients with a highly conserved CIDRβ1. A dashed line separates DBLβ12 from DC8 because DC8 typically contain DBLβ12, but these DBLβ12 formed a phylogenetic cluster with non-DC8 DBLβ12. Dashed lines separate putative DC9 components because transcripts containing all

components were not up-regulated in the combined assembly or the Corset analysis, but the clusters from which the up-regulated segments were identified contained multiple transcripts carrying the DC9 domains. ATS, acidic terminal sequence; CIDR, cysteine-rich interdomain region; DBL, Duffy binding-like; DC, domain cassette; HMM, Hidden Markov Model; PfEMP1, *Plasmodium falciparum* Erythrocyte Membrane Protein 1; TM, transmembrane.

<https://doi.org/10.1371/journal.pbio.2004328.g006>

the domain sequences identified in [14]. The identified domain regions were then hierarchically clustered using USEARCH [93] as described in the Materials and methods section at sequence identity levels 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 97. The counts for each cluster were aggregated up the hierarchical tree, and differential expression was tested at each node. Benjamini-Yekutieli [94] multiple testing correction was performed before the most significant node was successively chosen and added to the list of significant clusters. The children and parents of each node were removed from the list of potential clusters before the next node was chosen in the tree. A more detailed description is given in the Materials and methods section.

This approach attempts to identify the point, or node, in the hierarchical tree that best distinguishes domains associated with severe disease from those that are not. By looking at different levels of the tree, we are able to identify potentially important domain groups that would otherwise remain elusive. Additionally, by grouping domains at various identity levels, we are increasing the sensitivity to domain groups with higher sequence variation such as the DBL $\delta$  domain class.

Fig 7C illustrates the advantage of this approach by focusing on a tree related to the DBL $\epsilon$ 3 domain of [14]. At 70% identity, no clusters are significantly associated with severe disease. However, at 65%, 1 cluster (DBL $\epsilon$ 3.s.1) becomes significant. This difference is then lost at 50% identity. A PCA plot of the clusters at 50% identity is shown in Fig 7D. It provides a much clearer grouping of the severe samples than the previous domain definitions. Similar groupings are seen at all levels of the tree (S5 Fig).

The tree-building approach identified 70 differentially expressed domain clusters, of which 15 are up-regulated in severe malaria (S9 Data). To investigate possible associations between these 15 clusters, they were grouped based on their expression across the samples using average linkage hierarchical clustering. The dendrogram was then cut using the dynamic height-cutting algorithm of [95]. This identified 4 groups, labelled in the heatmap diagram of Fig 8A. Compared to Fig 7A, these domains provide a clearer distinction between parasites causing severe and uncomplicated malaria. However, the expression levels of these domain clusters were, in general, lower than the levels of the domains identified using the HMMER3 approach of Rask et al. (Fig 8B). This is consistent with the higher level of sequence identity within a domain cluster identified by the hierarchical tree approach. Therefore, fewer sequences per isolate were captured by each domain cluster than were assigned to each domain by the HMMER3 domain models, and the more diverse sequences captured by a HMMER3 model included many that were highly expressed in uncomplicated malaria (Fig 8B).

Examining Fig 8A in more detail reiterated elements of the original domain analysis (Fig 7A and Fig 6) and confirmed previous associations between expressed PfEMP1 domains and severe malaria. Two domain clusters similar to DBL $\beta$ 3 were up-regulated in severe disease. Domain transcripts from the DBL $\beta$ 3 clusters were aligned with published DBL $\beta$ 3 sequences [39] using MUSCLE [96] and clustered using FastTree [97]. Notably, the DBL $\beta$ 3 domain 4 from DC4 in the *var* gene PFD1235w [39] clustered tightly with 2 of the 4 transcripts from the DBL $\beta$ 3s.2 cluster we identified but none of the 4 transcripts from the DBL $\beta$ 3s.1 cluster. The turquoise group of domain clusters includes a cluster with sequences similar to DBL $\beta$ 12 domains. These DBL $\beta$ 12 sequences were compared with domains from DC8 but formed a separate cluster to the 2 clusters formed by the DBL $\beta$ 12 domains found in DC8 [14] (S2 Text).



HMMER3 models. There is less separation by phenotypes in this plot than was observed at the whole-transcript-and all-gene-analysis levels. Read count data for Fig 7 is available in [S8 Data](#). (C) An example of the hierarchical clustering tree. Colours represent significance, with red indicating a significant difference in expression after multiple testing correction and blue indicating not significant. Nodes are coloured grey if there is insufficient evidence for them to be considered in the testing either because they have less than 5 samples present or they are marked by DESeq2's prefilter step. At the 60% identity level, cluster 670\_X0.6 becomes significant. This significance is then obscured at the 50% identity level, demonstrating the importance of considering different levels of the hierarchy. (D) Clustering the domain level counts at 50% sequence identity rather than using the previous classifications of [14] improves the grouping of severe samples. At 50% identity, the severe samples are grouped more closely together, suggesting that they have more in common than the nonsevere samples; transformed read count data available in [S9 Data](#). PCA, principal component analysis; RNAseq, RNA sequencing.

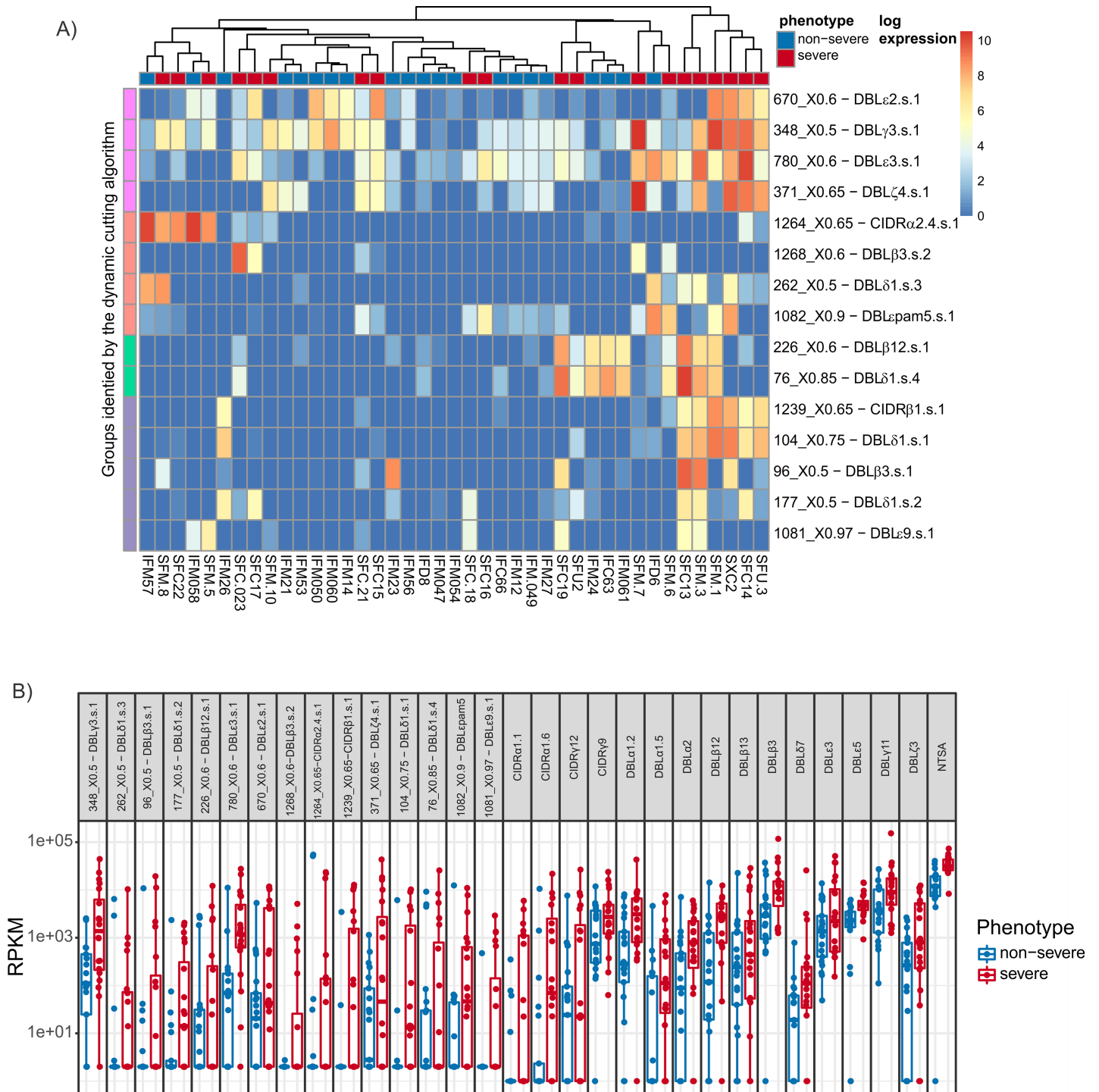
<https://doi.org/10.1371/journal.pbio.2004328.g007>

The turquoise group of domain clusters in [Fig 8A](#) also contained the domain cluster DBL $\delta$ 1.s.4, which did cluster closely with 2 DC8 DBL $\delta$ 1 domains, although DC8 DBL $\delta$ 1 domains cannot be differentiated from other DBL $\delta$ 1 domains. Nonetheless, the clustering by expression profile of DBL $\delta$ 1.s.4 and DBL $\beta$ 12.s.1 (turquoise group in [Fig 8A](#)) suggests that these domains could be part of a single DC8-like *var* in these patients. Unlike all the other analyses employed in this study, the hierarchical approach did not associate up-regulated CIDR $\alpha$ 1 sequences with severe malaria. This is presumably a consequence of the low conservation of the CIDR $\alpha$ 1 sequences [35] that would not be grouped at the minimum 50% identity threshold employed in the hierarchical analysis. All alignments and trees made with published sequences described above are further described in [S2 Text](#) and are available in the Github repository [https://github.com/gtonkinhill/falciparum\\_transcriptome\\_manuscript](https://github.com/gtonkinhill/falciparum_transcriptome_manuscript).

[Fig 8A](#) also identifies domain clusters that have not been previously associated with severe disease. These included DBL $\gamma$ 3.s.1 and DBL $\zeta$ 4.s.1 from the pink group of domain clusters in [Fig 8A](#). Four of the 5 DBL $\gamma$ 3 and 11 of the 12 DBL $\zeta$ 4 sequences previously described were found in tandem in DC9 [14]. These domain clusters were among the most abundant up-regulated in severe malaria ([Fig 8B](#)) and were also up-regulated in the combined transcript assembly ([Fig 6](#)), although not as a single transcript, so there is no direct evidence that DC9 itself is associated with severe malaria. DBL $\epsilon$ 3 was also amongst the most abundant domain clusters up-regulated in severe malaria in this analysis ([Fig 8B](#)). DBL $\epsilon$ 3 was also up-regulated in every analysis we performed ([Fig 6](#)) and was part of DC11 in the separate assembly transcript Corset analysis, although the DC11 transcripts were all different from the up-regulated DBL $\epsilon$ 3 transcripts from the individual domain analysis. A cluster similar to DBL $\epsilon$ pam5 from the pregnancy malaria-associated gene *var2csa* was also up-regulated.

DBL $\delta$ 1-CIDR $\beta$ 1/2/3/5/7 are common arrangements and were up-regulated in severe malaria in both the combined assembly and the Corset analysis of separately assembled transcripts. These domain subclasses are highly variable and thus difficult to distinguish using the previous classifications of [14]. However, the hierarchical approach identified the CIDR $\beta$ 1.s.1 domain cluster that includes a number of identical domain sequences from different isolates. This finding differs from the high variability noted in previous studies [14]. A single CIDR $\beta$ 1 from a gene containing DC8 formed a phylogenetic cluster with the conserved severe malaria-associated CIDR $\beta$ 1 sequence, but CIDR $\beta$ 1 from other DC8 genes did not ([S2 Text](#)). The DBL $\delta$ 1.s.1 and DBL $\delta$ 1.s.2 domain clusters did not form phylogenetic clusters with DBL $\delta$ 1 sequences from DC8 genes ([S2 Text](#)) but did cluster by expression profile with the uniquely conserved CIDR $\beta$ 1.s.1 (purple group [Fig 8A](#)) and with the non-DC4-like DBL $\beta$ 3.s.1, suggesting the existence of *var* genes that carry a unique pathogenic arrangement of domains, including a highly conserved CIDR $\beta$ 1 sequence. These 4 domains were not the most abundantly expressed in severe malaria but did discriminate strongly between severe and uncomplicated malaria, suggesting that they were strongly associated with severe malaria in a subset of cases ([Fig 8A](#), [Fig 8B](#)). Images of the hierarchical trees that make up each of these newly identified domain clusters are provided in the Github repository along with their respective multiple sequence alignments.





**Fig 8. Analysis of RNAseq data via de novo assembly at the level of var gene domains: Hierarchical analysis.** (A) Expression levels of the domain clusters identified using the hierarchical approach. Samples and domains are grouped using complete linkage hierarchical clustering. The colourings on the left indicate notable groups identified using the hierarchical cutting algorithm of [95]. The clusters are also annotated with the domain model of [14] that they most closely resemble. Raw read counts are available in S9 Data. (B) Expression levels of the domain clusters identified using the hierarchical approach, values for all samples, and the IQR and median are indicated. RPKM is reads per kb of domain per million reads mapped to total var domains (RPKM data available in S9 Data). IQR, interquartile range; RPKM, Reads Per Kilobase of transcript per Million mapped reads.

<https://doi.org/10.1371/journal.pbio.2004328.g008>

The sequence of domains associated with severe malaria by RNAseq was confirmed by Sanger sequencing 34 sequences that were cloned from genomic DNA (gDNA) of patient samples. These included domains identified by hierarchical analysis (13 domains up-regulated and 11 down-regulated in severe malaria), HMMER3 analysis (9 domains up-regulated in severe malaria), or corset analysis (the CIDR $\alpha$ 2.6-DBL $\beta$ 8 tandem arrangement up-regulated in severe malaria) (S10 Data). Every one of these cloned sequences was 100% identical to the cognate sequence assembled from RNAseq. RNAseq quantitation of domains in severe malaria was corroborated by quantitative reverse transcription PCR (Q-RT-PCR) of 10 up-regulated and 3 down-regulated domains. Insufficient RNA was available to test all patient samples, so a subset of patients was tested that included several patients for each domain that had high levels of RNAseq expression of that domain. Q-RT-PCR data correlated with RNAseq Reads Per Kilo-base of transcript per Million mapped reads (RPKM) for 12 of the 13 domains (Spearman  $r$  all greater than 0.53, all  $p < 0.03$ ) (S6 Fig). Q-RT-PCR and RNAseq of DBL $\epsilon$ 3.s.1 did not correlate because a number of uncomplicated malaria samples contained high levels of expression by Q-RT-PCR but not by RNAseq. Because Q-RT-PCR works best on small sequences (in this case a 64 bp product) and detects hybridisation rather than actual sequence, the most probable explanation for this discordance is cross-reactive amplification of non-DBL $\epsilon$ 3.s.1 by the Q-RT-PCR.

### Segment level

Due to the highly variable nature of *var* genes, it is common to focus on the most conserved segments—or blocks—of *var* gene sequence [14,50]. To investigate these conserved regions, we examined 628 homology blocks that were previously defined [14]. Of these, 613 were available for download from the VARDOM server. An approach was also developed comparable to that of [98] to divide the *var* sequences into conserved and variable regions.

The previously defined Homology blocks [14] were clustered and examined using the same approach as for the previously defined domains [14]. HMMER3 [91] profile hidden Markov models were used to annotate the separate assembly transcripts, and read counts were obtained from the aggregate of these annotations for each block. Overall, 16 homology blocks were identified as being differentially expressed (S11 Data). Ten of these (homology blocks 47, 97, 121, 126, 141, 142, 150, 183, 219, and 582) were up-regulated in severe disease. The heatmap in Fig A in S7 Fig indicates that homology blocks 219 and 582 are the most distinct in their expression profiles. Homology block 219 is located in the DBL $\alpha$ 1 domain class found in group A *var* genes and has previously been associated with severe malaria and rosetting [51]. Block 582 is usually found after a DBL $\zeta$ 4 domain in DC9. DBL $\zeta$ 4 domains were found to be up-regulated in severe disease in the domain-level analysis. Homology blocks 126 and 142 are found mainly within DBL $\epsilon$ 5 but also other DBL $\epsilon$  subtypes, whilst block 97 is found in DBL $\epsilon$ 4–8,12,14,PAM5 and DBL $\gamma$ 6,12,16,17 domains. Some of these DBL $\epsilon$  subtypes were also identified in the domain-level analysis. Blocks 121 and 150 are found in CIDR $\alpha$ 1 domains, and homology block 141 is found at the junction between CIDR $\alpha$ 1 and DBL $\beta$ 1,3,7,12 domains, whilst block 183 is found in DBL $\beta$ 1,3–5,10–12, domains. As mentioned previously, DBL $\beta$ 3, DBL $\beta$ 12, and CIDR $\alpha$ 1 domains are associated with DC4 and DC8, which have been associated with severe disease. Finally, homology block 47 is found within the acidic terminal sequence (ATS) of *var* genes. This region does not code for the extracellular part of the protein.

Although differentially expressed blocks are identified, it likely that, as in the domain analysis, a better classification can be made by making use of the novel transcripts. The homology blocks used for this analysis were defined based on conserved recombining regions in the *var* gene genome [14] and not on their relationship to disease severity. This may have obscured

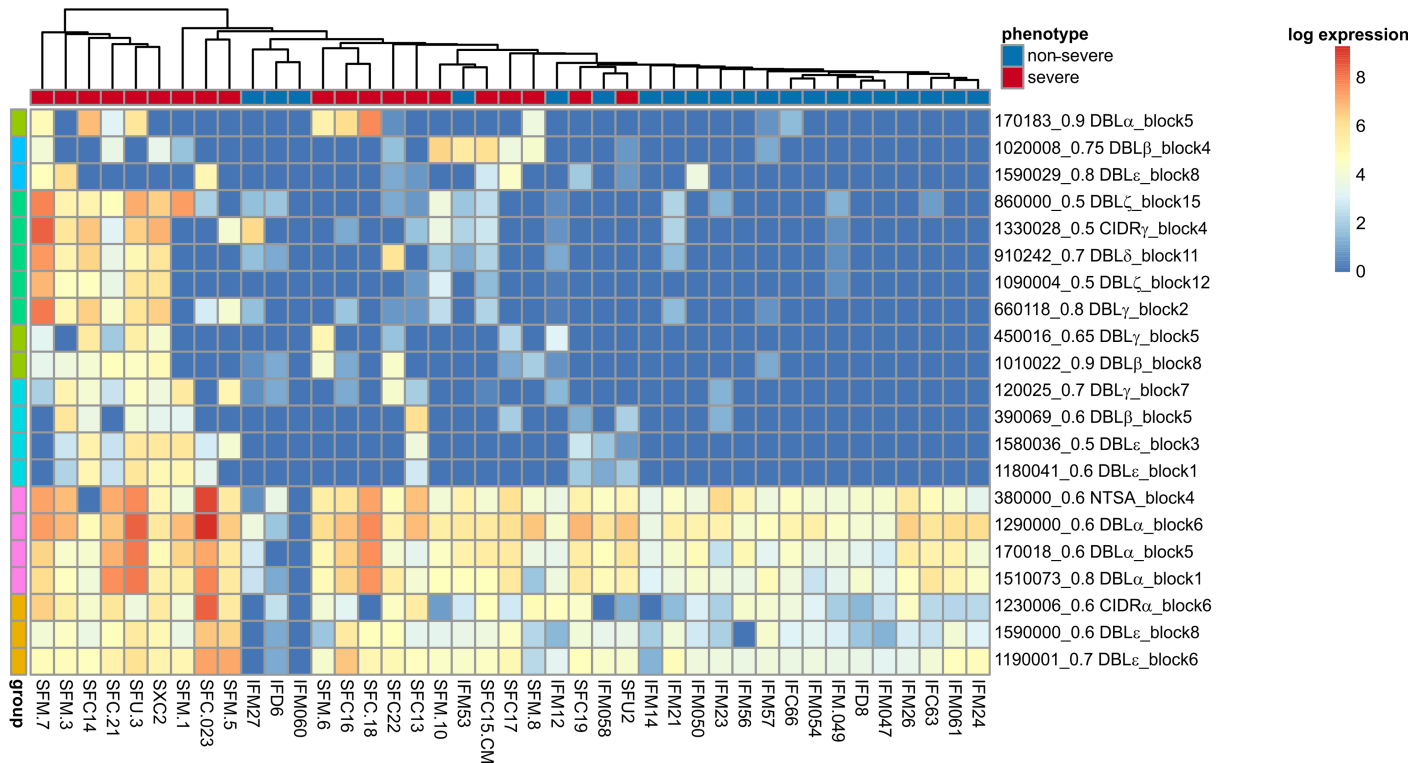
conserved regions that are related to severe disease. Furthermore, by focusing on only the most conserved regions, we are potentially ignoring informative—but more variable—regions. Finally, the homology blocks of [14] were inferred from laboratory strains and may not include conserved segments that are unique to severe disease types.

As an alternative, an approach similar to [98] was used to divide up multiple sequence alignments of the major domain classes. Domains identified using HMMER3 [99] were grouped into their major domain classes and aligned using Gismo [100]. Sequence logos of the resulting alignments were generated using skylign [101] (see S8 Fig). Gismo [100] was found to handle the large diversity in the *var* domains better than other aligners. The resulting alignments were then segmented into regions of high and low occupancy. If 7 or more consecutive columns within an alignment had an occupancy greater than 95%, these columns were considered a conserved region. The columns in between these conserved regions were considered variable regions. Regions of high variability are harder to align and consequently result in more gapped alignments. The results were found to be robust to the choice of the occupancy threshold as well as the choice for the number of consecutive conserved columns. This approach produces interleaved regions of higher conservation and diversity. The approach is similar to that proposed by [98]; however, we focus on both the conserved and variable regions. Each domain sequence was then split into segments based on the regions identified. We refer to these segments by their location within the domain from which they originate. For example, DBL $\alpha$ \_block2 is the second interleaved region of the DBL $\alpha$  domain class. The segments were then hierarchically clustered within their respective regions and analysed for differential expression in a similar manner to that used for the domains. Due to the short nature of these segments, CD-HIT [102] was used in place of USEARCH [93] because it accounts for the terminal gaps in its definition of pairwise sequence identity. Aside from identifying segments associated with severe disease, an advantage of this approach is that the resulting segments can easily be understood in terms of their relationship to the *var* domains and gene sequence.

DESeq2 [86] was used to investigate the differential expression of the segments, and Benjamini-Yekutieli [94] correction was used to correct for the multiple dependent tests. Overall, 26 clusters of segments were identified as being differentially expressed, of which 21 were up-regulated in severe disease (S12 Data). Fig 9 indicates the expression levels for each segment cluster across the samples. One lies in segment 4 of the NTSA region and 3 in regions 1, 5, and 6 of the DBL $\alpha$ 1 domain class of group A *var* genes. A single up-regulated cluster (170183\_0.9 DBL $\alpha$ \_block5) lies in segment 5 of DBL $\alpha$ 0.1 from non-group A *var* genes.

A cluster from region 6 of the CIDR $\alpha$  domain class had a similar expression profile to those segment clusters from the DBL $\alpha$  domain class. The cluster was also mostly made up of segments from CIDR $\alpha$ 1.4 and 1.8 domains that have been associated with severe disease [32–34,46,49]. Two DBL $\epsilon$  segment clusters from regions 6 and 8 derived primarily from DBL $\epsilon$ 5 subtypes also had similar expression profiles to the DBL $\alpha$  segment clusters. Furthermore, the clusters from region 6 of the DBL $\epsilon$  domain often appear in conjunction with homology blocks 126 and 142, suggesting they are identifying similar domains. The DBL $\epsilon$ 5 subtype has only been described in *var1*.

A striking difference in expression profile was observed between the moderately high levels of expression in both uncomplicated and severe malaria samples of the grouping of clusters from NTSA, DBL $\alpha$ 1, CIDR $\alpha$ 1, and the DBL $\epsilon$  region 6 and 8 compared to the markedly lower levels of expression in the uncomplicated samples of all other segment clusters that were up-regulated in the severe samples (Fig 9). This is consistent with the presence of NTSA and DBL $\alpha$ 1 on all group A *var* genes and therefore their expression in both uncomplicated and severe malaria. Similarly, *var1* is ubiquitously expressed by laboratory isolates and is not subject to the same program of gene regulation as other *var* genes, and so might be expected to be



**Fig 9. Analysis of RNAseq data via de novo assembly at the level of *var* gene segments.** Expression levels of novel conserved segment clusters found to be up-regulated in severe disease. Samples and segment clusters have been grouped using complete linkage hierarchical clustering. The raw read counts that were transformed for this figure are available in [S12 Data](#).

<https://doi.org/10.1371/journal.pbio.2004328.g009>

expressed in both uncomplicated and severe malaria. However, these data also suggest that expression of CIDR $\alpha$ 1 that can bind EPCR does not distinguish between severe and uncomplicated malaria as well as other *var* regions, which presumably mediate adhesion to other receptors.

The relationships between the segments, homology blocks, domains, and transcripts are illustrated in Fig 6. The Clusters from region 2 of DBL $\gamma$  and regions 12 and 15 of DBL $\zeta$  contained 15 transcripts, including 13 from the domain clusters DBL $\gamma$ 3.s.1 and DBL $\zeta$ 4.s.1, suggesting that we have identified similar disease-associated sequences at both the domain and segment level. Five of the transcripts included region 2 of DBL $\gamma$  and both or either of regions 12 and 15 of DBL $\zeta$  and 5 of the transcripts included regions 4 of CIDR $\gamma$  and/or region 11 of the DBL $\delta$ . DBL $\delta$ -CIDR $\gamma$  tandem domains invariably precede the DBL $\gamma$ -DBL $\zeta$  tandem domains of DC9. Seven of the 11 transcripts containing regions 12 and 15 of the DBL $\zeta$  domain class also contained published homology block 582 [14], indicating they may be detecting similar signals (S5 Data).

Significant clusters from regions 1 and 3 of the DBL $\epsilon$  domain class often appear in both of the previously identified domain clusters DBL $\epsilon$ 2.s.1 and DBL $\epsilon$ 9.s.1 as well as a number of other DBL $\epsilon$ 9 domain sequences. As the segment clusters collapse, 2 previous domain clusters along with a number of other sequences; this indicates that these segments may have better captured the sequence elements associated with severe disease. Due to the high diversity of the DBL $\epsilon$  domain class at the domain level, it is hard to accurately define which sequences are associated with severe disease, and consequently this highlights the virtue of investigating these sequences at multiple resolutions.

Eight out of 11 clustered DBL $\beta$  region 4 sequences are also part of the DBL $\beta$ 12 domain class. In 4 occurrences, it appears in a transcript that includes DC8. Two of the 10 DC4 transcripts clustered by Corset also contained DBL $\beta$  region 4 sequences that were part of DBL $\beta$ 3. Therefore, the up-regulated DBL $\beta$  region 4 sequence collapses 2 DBL $\beta$  subtypes that are independently associated with severe malaria and implicated in different adhesion phenotypes. Five of the 13 transcripts from the cluster containing DBL $\beta$  region 8 also contained the DC4-like homology block 141, whilst 2 of the 8 transcripts from the cluster containing region 5 of DBL $\beta$  also contain the non-DC4-like DBL $\beta$ 3.s.1 domain.

Transcripts from clusters of regions 5 and 7 of the DBL $\gamma$  domain class don't appear with other segments significantly associated with severe malaria, with the exception of 2 region 7 segments that appear in transcripts that include DC6. These segments may represent signal lost in the analysis of larger sequence elements.

To investigate the utility of using the different feature levels to differentiate severe and non-severe disease, we fit a logistic regression model with lasso regularisation. A model was generated for each level of the *var* gene analysis (transcript, domain, and segment) using the features found to be up-regulated in severe disease. We made use of crossvalidation to determine the optimal lambda value for the regularisation and to give an indication of how well the features distinguish severe and nonsevere disease. Overall, the segment level provided the best discrimination, with misclassification error of 9.76% and 12.20% for the homology block and segment clusters, respectively. The misclassification for the domain-level analysis was 21.95% when using either the Rask et al. domains or the hierarchically clustered domains as features. Notably, by making use of the domains defined using the novel hierarchical approach, fewer features were required to achieve a similar classification accuracy. Distinguishing between phenotypes using a smaller number of features is important when investigating possible targets for vaccines. The transcript-level features provided the least discrimination, giving misclassification errors of 31.71% and 43.90% for the combined assembly transcripts and transcript clusters, respectively. It should be noted that these classification rates cannot be generalised to new samples because the cross validation was used to determine the lambda value as well as the misclassification rates. The code for this regression analysis is available in the Github repository.

The relationship between the segments, domains, and transcripts discussed is available in [S5 Data](#) and [Fig 6](#). Tree diagrams, like those produced for the domains, are available for each significant segment cluster in the Github repository.

## Discussion

Transcriptional profiling of parasites isolated from patients with severe malaria indicated a shift towards a less glycolytic phenotype. Previous studies have also reported decreases in glycolytic transcripts in some clinical isolates [6], including those from patients with higher temperatures [8] as well as in parasites cultivated in vitro at a high density that inhibits subsequent growth [103]. Down-regulation of genes encoding key enzymes in folate and pyrimidine biosynthesis is also consistent with decreased nucleotide production and reduced parasite growth. The down-regulation of genes involved in histone methylation was similar to deregulation of genes involved in chromatin and RNA biology that was observed in clinical isolates from patients with an elevated surrogate measure of parasitaemia [8].

Our data suggest that parasites causing severe malaria have a more metabolically quiescent phenotype than parasites causing uncomplicated malaria. It remains to be determined whether parasites with the severe malaria transcriptional profile are more resilient and therefore able to cause severe malaria, or whether the host environment in either severe malaria or



uncomplicated malaria could have selected or elicited the differing transcriptional profiles. Modulation of parasite growth in response to host environment might be consistent with previous reports of *P. falciparum* density sensing in malaria [103–105] and protracted maturation of *P. berghei* and *P. yoelli* in response to an acute host immune response [106]. In the latter study, more mature, circulating *P. berghei* and *P. yoelli* were detected in semi-immune than naive mice. This was consistent with our observation that the circulating parasites were older in the uncomplicated than the severe malaria patients because we previously showed that the uncomplicated malaria patients had more immunity to PfEMP1 than the severe malaria patients [88].

The parasites causing severe malaria had also down-regulated genes involved in PfEMP1 surface expression. This differed from the reported increased expression of genes encoding exported proteins involved in PfEMP1 surface expression in severe malaria from a posthoc comparison [107] of separately published transcriptomes of parasites causing severe [7] and uncomplicated [58] malaria. This difference probably relates to the difficulty of posthoc inference of differential gene expression when the compared samples are from different populations and different studies and were analysed using different microarrays. None of the 87 genes identified in up-regulated gene sets by Pelle et al. were up-regulated in severe malaria in the current study; however, 17 of these genes were down-regulated, including skeleton-binding protein 1, which was the only gene directly involved in PfEMP1 surface expression identified by Pelle et al. We previously showed that the severe malaria patients in the current study had antibodies to PfEMP1 that were generally present at lower levels and that recognised fewer PfEMP1s than the antibodies from the uncomplicated malaria patients [88]. This suggests that humoral immunity to PfEMP1 did not select for decreased PfEMP1 surface expression in the parasites causing severe malaria. However, loss or decrease of many of the proteins involved in PfEMP1 surface expression causes decreased cytoadherence [76,108,109], so the parasites infecting patients with severe malaria at the time of sampling might have had a decreased cytoadherent capacity.

The unique *var* transcriptional profile we describe in severe malaria recapitulates all of the previously described associations as well as uncovering multiple, novel sequence associations. These findings are remarkable considering that all of the associations that have been observed previously in children with severe malaria from multiple sites across Africa were found in 23 adults with severe malaria from Papua. This suggests that the same conserved *var* genes are associated with severe disease in nonimmune individuals regardless of geography or patients' age. Furthermore, a consistent pattern of expression of restricted subsets of *var* genes, domains, and/or segments was observed despite heterogeneous presentations of severe disease. Similarly, the severe malaria non-*var* transcriptome clusters also did not segregate by specific severe malaria syndromes. These observations suggest that common mechanisms of disease may cause the varied syndromes of severe malaria. This could have therapeutic implications, although the analyses should be confirmed with larger sample sets.

These findings emphasise the strength of the association between severe malaria and DC8, DC4, DC6 CIDR $\alpha$ 1, DBL $\beta$ 3, and DBL $\beta$ 12 sequences, which were each shown to be up-regulated in multiple analyses of the de novo *var* assemblies. However, they also uncover significant, novel associations with other *var* sequences at the transcript, domain, and segment level. Some of these were found at multiple levels of analysis, e.g., DC11, CIDR $\alpha$ 2.6-DBL $\beta$ 8, DBL $\epsilon$ 3, DBL $\gamma$ 3, DBL $\zeta$ 4, and DBL $\epsilon$ 2/9, and in the individual domain analysis, the latter 4 domains were expressed at least as highly in severe malaria as the EPCR-binding CIDR $\alpha$ 1 sequences. We cannot exclude the possibility that some of these domains were present on the same PfEMP1 as a CIDR $\alpha$ 1 sequence; however, CIDR $\alpha$ 1 was not present on the significantly up-regulated transcripts that carried these other domains in either the combined assembly or the core-set analysis of the individual isolate *var* assemblies.

We developed a novel analytical approach testing sequences for associations with disease at multiple levels of sequence homology. This revealed domain subtypes that were strongly associated with disease, including a highly conserved CIDR $\beta$ 1 subtype and a DBL $\delta$ 1 subtype that clustered in the same patients. The diversity of the parent CIDR $\beta$ 1 and DBL $\delta$ 1 subtypes prevented detection of an association using the established subtype classifications. Finally, we revealed striking associations between smaller *var* sequence segments and severe disease again by testing for associations at multiple levels of sequence identity. These small segments were limited in number, and many of the findings recapitulated our domain analysis. Some of these segments collapsed multiple domain subtypes, e.g., DBL $\beta$ \_block4 collapsed DBL $\beta$ 3 and DBL $\beta$ 12, raising the possibility that a single segment may elicit cross-reactive immunity against different domain subtypes that are independently associated with severe disease. These segments may help identify critical, fine-scale details of the *var* sequences expressed by parasites that cause disease and may be of great utility in designing vaccines for severe malaria. The association of these sequences with severe malaria should be validated in other populations from across the world and the encoded proteins tested for adhesion phenotype and for seroreactivity consistent with protection from severe malaria.

## Materials and methods

### Ethics statement

Written, informed consent was provided by all participants. The study was approved in Indonesia by the Eijkman Institute Research Ethics Commission (project number 46), in Australia by the Melbourne Health Human Research Ethics Committee (project number 2010.284) and Human Research Ethics Committee of the NT Department of Health & Families and Menzies School of Health Research, Darwin, Australia (HREC 2010–1396).

### Data sets

The data sets generated and/or analysed during the current study are available in the Arrayexpress repository accession: E-MTAB-5860 (sequenced libraries for each sample) and the ENA repository accession: PRJEB20632 (de novo *var* gene assemblies for combined and individual samples).

### Sample collection

Venous samples were collected from patients with severe ( $n = 23$ ) and uncomplicated ( $n = 21$ ) malaria attending a healthcare facility in Timika, Papua Province, Indonesia. This area has unstable malaria transmission, with estimated annual parasite incidence of 450 per 1,000 population and symptomatic illness in all ages [110]. Severe malaria was defined as peripheral parasitaemia with at least one modified World Health Organization (WHO) criterion of severity [111]. All of the 23 patients with severe malaria had parasitemias greater than 1,000 per  $\mu$ L, which is a previously derived threshold that predicts clinical disease in northern Papua [49]. Therefore, incidental parasitaemia is unlikely in these severe malaria patients.

### RNA extraction and RNAseq

White blood cells were depleted from the blood by retention on CF11 cellulose (Whatman-no longer available) using a modification of a previously described protocol [112] (S1 Text Supplementary methods). RNA was extracted from erythrocytes in TRIzol using a modified RNeasy mini (QIAGEN, Hilden, Germany) protocol (S1 Text Supplementary methods). Purified RNA 1 to 3  $\mu$ g was depleted of Hb mRNA using the GlobinClear human Hb RNA

depletion kit (Ambion, Thermo Fisher Scientific, Waltham, MA) and a modified protocol (S1 Text supplementary methods).

## Sequencing

mRNA was oligo dT purified from the total RNA using the NEBNext Poly(A) mRNA magnetic isolation module (New England Biolabs, Ipswich, MA) and mRNA fragmented, reverse transcribed, and used for library synthesis using the NEBnext ultra directional RNA library prep kit for Illumina (New England Biolabs) as per the manufacturer's instructions but with modifications (S1 Text supplementary methods), including a high AT tolerant PCR amplification [113]. Libraries were 100 bp paired end sequenced on a 2500-HT Hiseq (Illumina, San Diego, CA) using RapidRun chemistry (Illumina).

## De novo assembly of *var* genes

Briefly, de novo assembly of *var* genes was performed by running the SoapDeNovo-Trans [114] and Cap3 [115] pipeline described in [53] (Fig C in S2 Fig). Non-*var* reads were first filtered out by removing reads that aligned to the *H. sapiens*, *P. vivax*, and non-*var P. falciparum* reference genomes. The resulting contigs were filtered for contaminants and translated into the correct reading frame. A more thorough description of the assembly methods is available in S1 Text. Additionally, the code used to run the pipeline is available on Github at [https://github.com/PapenfussLab/assemble\\_var](https://github.com/PapenfussLab/assemble_var).

## All gene expression analysis

Reads were first aligned to the *H. sapiens* and *P. falciparum* reference genomes using Subread-align v1.4.6 [97] with parameters -u -H. FeatureCounts v1.20.2 [116] was used to obtain read counts for each gene. To account for parasite life cycle, each sample is estimated as a mixture of 6 parasite life cycle stages from [57], excluding the ookinete stage. We aimed to choose the proportions  $\pi$  for each sample to minimise

$$\sum_{i=1}^N (g_{i,sample} - \sum_{s \in S} \pi_s g_{i,s})^2$$

subject to the constraints

$$\sum_{s \in S} \pi_s = 1$$

And

$$\pi_s \geq 0$$

such that  $g_{i,s}$  represents the expression of the  $i$ th gene in stage  $s$  of the [57] data.

Three factors of unwanted variation were estimated using the RUV4 function from the R package ruv v0.9.6 [117] using the 1,009 genes with the lowest  $p$ -values from [118] as controls. The choice of control genes was compared to using the least differentially expressed genes of [8], which was found to give similar results. Finally, the gene counts along with the estimated ring-stage factor, and 3 factors of unwanted variation estimated by RUV4 were fed into the Limma/Voom [55,56] differential analysis pipeline. For a detailed outline of the specific commands run in the all-gene analysis, see S1 Text and rmarkdown S1 Text available in the Github repository [https://github.com/gtonkinhill/falciparum\\_transcriptome\\_manuscript](https://github.com/gtonkinhill/falciparum_transcriptome_manuscript).

## *var* gene expression analysis

For all levels of *var* expression analysis, library size was normalised using the median ratio method [119] in the default DESeq2 pipeline [86].

### Transcript level

Differential expression analysis at the *var* transcript level was performed using 2 distinct approaches. The first made use of the separately assembled transcripts by first aligning reads to the transcripts allowing for multiple mapping using Bowtie v0.12.9 [120]. The transcripts were then clustered using Corset v1.03 [89]. The resulting cluster read counts were analysed using the default DESeq2 pipeline [86]. An alternative strategy made use of the combined assembly transcripts. Reads were aligned and transcript level counts obtained using Subread and FeatureCounts, respectively, before analysing differential expression using DESeq2. Rmarkdown texts S2 and S3 in the Github repository provide a more thorough description of this analysis along with the code [https://github.com/gtonkinhill/falciparum\\_transcriptome\\_manuscript](https://github.com/gtonkinhill/falciparum_transcriptome_manuscript).

### Type/domain level

HMMER3's `hmmsearch` v3.1b1 [91] was used to search the NTS, DBL, and CIDR domain models of [14] against the assembled transcripts from each sample. The most significant domain was annotated first and then successively less significant domains, with the requirement that 2 domains do not overlap. An E-value threshold of 1e-8 was chosen to minimise spurious annotations.

FeatureCounts was used to allocate reads to domains using a SAF file built from the HMMER3 annotations. The resulting counts were then aggregated using the previous domain classification of [14] as well as a novel hierarchical approach. The annotated domains were hierarchically clustered using USEARCH [93] by first clustering by length and then by successively lower identity thresholds. The read counts for each domain are then aggregated up this hierarchical tree, and the default DESeq2 pipeline was used to identify differentially expressed nodes. After multiple testing correction [94], we iteratively reject the null hypothesis ( $p < 0.05$ ) of the most significant node in the hierarchy before removing its ancestor and children nodes. This ensures that we select the most significant grouping of domains from which to form clusters. DESeq2 was also run on the domains aggregated using the previous classification.

### Segment level

**Homology block analysis.** In a similar fashion to the domain-level analysis, 613 of the possible 628 homology blocks of [14] were aligned to the separate assembly transcripts using a bit score cutoff of 9.97 as is described in [14]. Read counts were again aggregated for each homology block before the default DESeq2 pipeline was used to analyse differential expression.

**Identification of novel, differentially expressed segments.** To identify novel segments, *var* domains were initially identified as in the domain section. Major domain classes were aligned using Gismo v2.0 [100]. The resulting alignments were then segmented into regions of high and low occupancy in an approach comparable to [98]. If 7 or more consecutive columns within an alignment had an occupancy greater than 95%, these columns were considered a conserved region. Terminal gaps were not counted in the occupancy calculations. The columns in between these conserved regions were considered variable regions. Each domain sequence was then split into segments based on the regions, and the segments were clustered hierarchically within each region using CD-HIT v4.6 [102]. CD-HIT was chosen because it

takes the terminal ends into account when calculating pairwise identity, making it more appropriate for the short segments being considered here. Finally, as was done at the domain level, differential expression analysis was performed on each node of the hierarchy, and the most significant nodes were successively chosen.

A more thorough description of the methods for identifying the differentially expressed domain and homology block segments is given in [S1 Text](#) as well as in the markdown texts S4–S6 available in the Github repository.

**Untargeted LC-MS profiling.** Plasma samples (5  $\mu$ L) were extracted with 80% acetonitrile containing 1  $\mu$ M of  $^{13}\text{C}^{15}\text{N}$  aspartate (internal standard) and LC-MS analyses performed as described previously [121] ([S1 Text](#) supplementary methods). Data were converted to mzXML and analysed using the MAVEN software package [122]. Comparison between severe and uncomplicated samples was performed using a Benjamini Hochberg–corrected  $t$  test with significance set at  $p < 0.01$ , and significant features were searched in the METLIN database for putative identification.

## Supporting information

**S1 Fig. RNA quality and summary statistics on raw RNAseq data.** (A) RQI values from BioRad Experion automated RNA electrophoresis system. RQI values can range from 1 (low) to 10 (high); 28S/18S rRNA ratios are also provided; N/A indicates samples for which values could not be interpolated because the molecular weight standard ladder failed, though RNA quality for these samples could still be assessed visually from the electrophoretograms. The rRNA profile differs from the typical 2 peaks because it is a mixture of *H. sapiens* and *P. falciparum* 28S and 18S rRNAs; the *P. falciparum* rRNAs migrate as the 2 inner peaks. (B) Number of fragments (read pairs) assigned to genes of the *P. falciparum* reference genome. A sample was required to have at least 1 million fragments to be included in the rest of the analysis. (C) Summary diagram of the approaches taken to analyse the RNAseq data. RNAseq, RNA sequencing; RQI, RNA Quality Index.

(PDF)

**S2 Fig. De novo var gene assembly.** (A) Assembled transcripts greater than 500 nt in length were aligned using BLAST to the sequence database from [14]. Multiple alignments were allowed. The resulting alignments are plotted by reference sequence, including the percentage identity of the alignment. The alignment is coloured in bold, whilst sequence not aligned to the reference is translucent. The vertical black line indicates the end of the reference sequences. Fig A in S2 Fig indicates the alignment of the simpler *P. falciparum* ItG subclone E8B. (B) Similar to panel A; however, the ItG subclone CS2—in which there has been a recombination between IT4var04 and IT4var08 var genes—is displayed. (C) A flow diagram of the final assembly pipeline used to generate var gene transcripts from RNAseq data. RNAseq, RNA sequencing.

(PDF)

**S3 Fig. Differential expression analysis of all genes.** Average linkage heatmap of the expression levels of the 358 deregulated genes from the differential analysis results for all genes with sufficient coverage. Two clusters of severe malaria transcriptomes that differ by expression profile are indicated as ‘S1’ and ‘S2’. Normalised CPM data are available in [S1 Data](#). CPM, counts per million mapped reads.

(PDF)

**S4 Fig. Differential expression analysis of var gene features.** (A) Comparison of expression levels of NTSA and NTSB segments for severe and nonsevere phenotypes. NTSA was found to be significantly up-regulated in severe disease. (B) PCA plot of the read counts associated with

transcripts from the combined sample assembly. The severe samples are clustered more tightly than the nonsevere. This is consistent with a more conserved set of features that describe severe disease. The less tightly clustered nonsevere samples are consistent with the difficulty in obtaining complete immunity to malaria. Raw read counts used for S4 Fig are available in [S9 Data](#). NTSA, N-terminal sequence A; NTSB, N-terminal sequence B; PCA, Principal Component Analysis.

(PDF)

**S5 Fig. PCA plots of read counts for different identity levels of the *var* domain hierarchical clustering.** These plots are similar to the PCA plot in Fig B in [S4 Fig](#), indicating that we are still capturing similar differences between severe and nonsevere cases using these domain clusters. Raw read counts transformed for S5 Fig are available in [S9 Data](#). PCA, Principal Component Analysis.

(PDF)

**S6 Fig. Transformed data, scatter plots, and Spearman correlations for *var* domain expression measured by Q-RT-PCR ( $2^{-\Delta C_p}$ ) and RNAseq (RPKM).** Red dots are severe malaria samples SFC12, SFC14, SFC15, SFC17, SFC19, SFC22, SFM1, SFM3, SFU2, and SXC2; blue dots are uncomplicated malaria samples IFM049, IFM047, IFM050, IFM054, IFM12, IFM27, IFM53, and IFM56; black dots are samples that were excluded from the correlation analysis because they had high RPKM values, but the relevant transcripts lacked the Q-RT-PCR primer binding sites (670\_X0.6—DBL $\epsilon$ 2, 348\_X0.5—DBL $\gamma$ 3, 345\_X0.5\_DBL $\gamma$ 13.ns.1), or the PCR product had a different dissociation curve to all other products amplified with those primers (226\_X0.6—DBL $\beta$ 12). DBL, Duffy binding-like; Q-RT-PCR, quantitative reverse transcription PCR; RNAseq, RNA sequencing; RPKM, Reads Per Kilobase of transcript per Million mapped reads.

(PDF)

**S7 Fig. Expression levels of homology blocks as defined in [14] found to be up-regulated in severe disease.** (A) Heatmap of expression level with patients' samples and homology blocks grouped using complete linkage hierarchical clustering. (B) PCA plot of read counts annotated to homology blocks identified in transcripts from the separate sample assembly. This plot indicates that homology blocks do not separate severe disease as distinctly as the domain clusters. Raw read counts that were transformed for these figures are provided in [S11 Data](#).

(PDF)

**S8 Fig. Multiple sequence alignments used to define *var* domain segments.** Logos of major domain classes produced using Skylign [101] on the output of Gismo [100]. Hierarchical clustering of segments was used for differential expression analysis.

(PDF)

**S1 Text. Supplementary methods.**

(DOCX)

**S2 Text. Phylogenetic analysis of *var* domains with known DC4 and DC8 DBL $\beta$ 3, DBL $\beta$ 12, DBL $\delta$ 1, and CIDR $\beta$ 1 sequences.** CIDR, cysteine-rich interdomain region; DBL, Duffy binding-like; DC, domain cassette.

(PDF)

**S1 Table. Raw read counts by sample for *H. sapiens* and *P. falciparum*. %Pf, percentage of reads that mapped to *P. falciparum*.**

(PDF)



**S2 Table. Comparison of approaches for *var* gene de novo assembly.** EIC and ECS are separate subcultures of the E8B clone of the ItG isolate; CS2 is a subclone of E8B.  
(PDF)

**S3 Table. De novo-assembled *var* transcript statistics by sample.**  
(PDF)

**S4 Table. Deregulation in severe malaria of *P. falciparum* gene sets previously reported to be deregulated in vivo.** FDR, false discovery rate; Prp, proportion.  
(DOCX)

**S1 Data. All *P. falciparum* genes differentially expressed in severe malaria.** Chr, chromosome; CI, confidence interval; logFC, log fold-change.  
(XLSX)

**S2 Data. *P. falciparum* gene expression analysis, proportion by life cycle stage and read counts normalised for library size and/or for variations in life cycle staging and other batch effects.**  
(XLSX)

**S3 Data. GO and KEGG classifications that were enriched in genes that were differentially expressed in severe malaria.** GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.  
(XLSX)

**S4 Data. Mz and Rt of metabolites enriched or depleted from plasmas of severe malaria patients.** Av, average; Mz, mass charge; Rt, retention time; SD, standard deviation.  
(XLSX)

**S5 Data. Assembled *var* transcripts from the separate assemblies.** Includes the translated protein sequences for all the transcripts as well as the annotated domain, domain clusters, homology blocks, and segment clusters. For each of the separate assembly transcripts, its closest BLAST hit in the combined assembly is included as well as the cluster to which it belongs from the Corset analysis. The results of the differential expression analyses are also summarised. HB, homology block; LFC, log fold-change; Orf, open reading frame.  
(XLSX)

**S6 Data. Assembled *var* transcripts from the combined assembly.** Includes the translated protein sequences for all the transcripts as well as the annotated domain and homology blocks. The results of the differential expression analyses are also summarised. Raw read counts used for Fig 4A heatmap and RPKM used for Fig 4B boxplot are also included. HB, homology block; LFC, log fold-change; Orf, open reading frame; RPKM, Reads Per Kilobase of transcript per Million mapped reads.  
(XLSX)

**S7 Data. *Var* transcripts from the separate assemblies clustered by CORSET.** The results of the differential expression analyses are also summarised.  
(XLSX)

**S8 Data. *Var* domains classified by HMMER [14] that were significantly deregulated in severe malaria.** The results of the differential expression analyses are also summarised.  
(XLSX)

**S9 Data. *Var* domain clusters at different percent identities that were significantly deregulated in severe malaria.** The domains from the separate assembly *var* transcripts that are present in each cluster are identified. The results of the differential expression analyses are also summarised. Transformed read counts at 50% identity, RPKM for all deregulated domains, and raw read counts for each domain from the separate de novo assemblies are also provided. RPKM, Reads Per Kilobase of transcript per Million mapped reads.

(XLSX)

**S10 Data. Thirty-four *var* domain sequences from the separate sample assemblies that were cloned and sanger sequenced from patient isolate gDNA.** gDNA, genomic DNA.

(XLSX)

**S11 Data. *Var* sequence homology blocks [14] that were significantly deregulated in severe malaria.** Differential expression analyses and raw read count data are provided.

(XLSX)

**S12 Data. *Var* sequence segment clusters that were significantly deregulated in severe malaria.** The sequence blocks within domains from the separate assembly *var* transcripts that are present in each cluster are identified. The results of the differential expression analyses and the raw read counts that were transformed for Fig 9 are also provided.

(XLSX)

## Acknowledgments

We acknowledge Kim Piera for technical assistance.

## Author Contributions

**Conceptualization:** Michael F. Duffy.

**Data curation:** Gerry Q. Tonkin-Hill, Anthony T. Papenfuss, Michael F. Duffy.

**Formal analysis:** Gerry Q. Tonkin-Hill, Hanh H. T. Nguyen, Simon A. Cobbold, Anthony T. Papenfuss, Michael F. Duffy.

**Funding acquisition:** Anthony T. Papenfuss, Michael F. Duffy.

**Investigation:** Gerry Q. Tonkin-Hill, Leily Trianty, Rintis Noviyanti, Hanh H. T. Nguyen, Boni F. Sebayang, Daniel A. Lampah, Simon A. Cobbold, Janavi S. Rambhatla, Anthony T. Papenfuss, Michael F. Duffy.

**Methodology:** Gerry Q. Tonkin-Hill, Rintis Noviyanti, Jutta Marfurt, Simon A. Cobbold, Ric N. Price, Nicholas M. Anstey, Anthony T. Papenfuss, Michael F. Duffy.

**Project administration:** Leily Trianty, Rintis Noviyanti, Jutta Marfurt, Ric N. Price, Nicholas M. Anstey, Anthony T. Papenfuss, Michael F. Duffy.

**Resources:** Leily Trianty, Rintis Noviyanti, Jutta Marfurt, Simon A. Cobbold, Malcolm J. McConville, Graham V. Brown, Karen P. Day, Ric N. Price, Nicholas M. Anstey, Anthony T. Papenfuss, Michael F. Duffy.

**Software:** Gerry Q. Tonkin-Hill, Anthony T. Papenfuss.

**Supervision:** Anthony T. Papenfuss, Michael F. Duffy.

**Validation:** Gerry Q. Tonkin-Hill, Anthony T. Papenfuss, Michael F. Duffy.

**Visualization:** Rintis Noviyanti, Stephen J. Rogerson, Graham V. Brown, Anthony T. Papenfuss, Michael F. Duffy.

**Writing – original draft:** Gerry Q. Tonkin-Hill, Michael F. Duffy.

**Writing – review & editing:** Gerry Q. Tonkin-Hill, Simon A. Cobbold, Malcolm J. McConville, Stephen J. Rogerson, Graham V. Brown, Karen P. Day, Ric N. Price, Nicholas M. Anstey, Anthony T. Papenfuss, Michael F. Duffy.

## References

1. World Health Organization. World Malaria Report 2015. World Health Organization; 2015. ISBN: 978 92 4 156515 8. Available from: [http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf).
2. WHO. Severe malaria. *Trop Med Int Health*. 2014; 19 Suppl 1: 7–131.
3. White NJ, Turner GDH, Day NPJ, Dondorp AM. Lethal malaria: Marchiafava and Bignami were right. *J Infect Dis*. 2013; 208: 192–198. <https://doi.org/10.1093/infdis/jit116> PMID: 23585685
4. Almelli T, Nuel G, Bischoff E, Aubouy A, Elati M, Wang CW, et al. Differences in gene transcriptomic pattern of *Plasmodium falciparum* in children with cerebral malaria and asymptomatic carriers. *PLoS ONE*. 2014; 9: e114401. <https://doi.org/10.1371/journal.pone.0114401> PMID: 25479608
5. Lemieux JE, Gomez-Escobar N, Feller A, Carret C, Amambua-Ngwa A, Pinches R, et al. Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in *ex vivo* culture. *Proceedings of the National Academy of Sciences*. 2009; 106: 7559–7564.
6. Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, Kamal M, et al. Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. *Nature*. 2007; 450: 1091–1095. <https://doi.org/10.1038/nature06311> PMID: 18046333
7. Milner DA Jr, Pochet N, Krupka M, Williams C, Seydel K, Taylor TE, et al. Transcriptional profiling of *Plasmodium falciparum* parasites from patients with severe malaria identifies distinct low vs. high parasitemic clusters. *PLoS ONE*. 2012; 7: e40739. <https://doi.org/10.1371/journal.pone.0040739> PMID: 22815802
8. Yamagishi J, Natori A, Tolba MEM, Mongan AE, Sugimoto C, Katayama T, et al. Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. *Genome Res*. 2014; 24: 1433–1444. <https://doi.org/10.1101/gr.158980.113> PMID: 25091627
9. Howard RJ, Barnwell JW, Rock EP, Neequaye J, Ofori-Adjei D, Maloy WL, et al. Two approximately 300 kilodalton *Plasmodium falciparum* proteins at the surface membrane of infected erythrocytes. *Mol Biochem Parasitol*. 1988; 27: 207–223. PMID: 3278227
10. Chan J-A, Fowkes FJI, Beeson JG. Surface antigens of *Plasmodium falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. *Cell Mol Life Sci*. 2014; 71: 3633–3657. <https://doi.org/10.1007/s00018-014-1614-3> PMID: 24691798
11. Carlson J, Helmby H, Hill AV, Brewster D, Greenwood BM, Wahlgren M. Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet*. 1990; 336: 1457–1460. PMID: 1979090
12. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 2002; 419: 498–511. <https://doi.org/10.1038/nature01097> PMID: 12368864
13. Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, et al. Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics*. 2007; 8: 45. <https://doi.org/10.1186/1471-2164-8-45> PMID: 17286864
14. 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; 6: e1000933. <https://doi.org/10.1371/journal.pcbi.1000933> PMID: 20862303
15. Day KP, Artzy-Randrup Y, Tiedje KE, Rougeron V, Chen DS, Rask TS, et al. Evidence of strain structure in *Plasmodium falciparum* var gene repertoires in children from Gabon, West Africa. *Proc Natl Acad Sci U S A*. 2017; 114: E4103–E4111. <https://doi.org/10.1073/pnas.1613018114> PMID: 28461509
16. Chen DS, Barry AE, Leliwa-Sytek A, Smith T-A, Peterson I, Brown SM, et al. A molecular epidemiological study of var gene diversity to characterize the reservoir of *Plasmodium falciparum* in humans in Africa. *PLoS ONE*. 2011; 6: e16629. <https://doi.org/10.1371/journal.pone.0016629> PMID: 21347415

17. Lavstsen T, Salanti A, Jensen ATR, Arnot DE, Theander TG. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. *Malar J*. 2003; 2: 27. <https://doi.org/10.1186/1475-2875-2-27> PMID: 14565852
18. Smith JD. The role of PfEMP1 adhesion domain classification in *Plasmodium falciparum* pathogenesis research. *Mol Biochem Parasitol*. 2014; 195: 82–87. <https://doi.org/10.1016/j.molbiopara.2014.07.006> PMID: 25064606
19. Kaestli M, Cockburn IA, Cortés A, Baea K, Rowe JA, Beck H-P. Virulence of malaria is associated with differential expression of *Plasmodium falciparum var* gene subgroups in a case-control study. *J Infect Dis*. 2006; 193: 1567–1574. <https://doi.org/10.1086/503776> PMID: 16652286
20. Falk N, Kaestli M, Qi W, Ott M, Baea K, Cortés A, et al. Analysis of *Plasmodium falciparum var* genes expressed in children from Papua New Guinea. *J Infect Dis*. 2009; 200: 347–356. <https://doi.org/10.1086/600071> PMID: 19552523
21. Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen ATR, Müller D, et al. Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infect Immun*. 2006; 74: 3904–3911. <https://doi.org/10.1128/IAI.02073-05> PMID: 16790763
22. Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, et al. Differential *var* gene transcription in *Plasmodium falciparum* isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol*. 2006; 150: 211–218. <https://doi.org/10.1016/j.molbiopara.2006.08.005> PMID: 16996149
23. Warimwe GM, Fegan G, Musyoki JN, Newton CRJC, Opiyo M, Githinji G, et al. Prognostic indicators of life-threatening malaria are associated with distinct parasite variant antigen profiles. *Sci Transl Med*. 2012; 4: 129ra45. <https://doi.org/10.1126/scitranslmed.3003247> PMID: 22496547
24. Kalmbach Y, Rottmann M, Kombila M, Kremsner PG, Beck H-P, Kun JFJ. Differential *var* gene expression in children with malaria and antidiromic effects on host gene expression. *J Infect Dis*. 2010; 202: 313–317. <https://doi.org/10.1086/653586> PMID: 20540611
25. Su X-Z, 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; 82: 89–100. PMID: 7606788
26. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. Classification of adhesive domains in the *Plasmodium falciparum* Erythrocyte Membrane Protein 1 family. *Mol Biochem Parasitol*. 2000; 110: 293–310. PMID: 11071284
27. Larremore DB, Sundararaman SA, Liu W, Proto WR, Clauset A, Loy DE, et al. Ape parasite origins of human malaria virulence genes. *Nat Commun*. 2015; 6: 8368. <https://doi.org/10.1038/ncomms9368> PMID: 26456841
28. Bull PC, Kortok M, Kai O, Ndungu F, Ross A, Lowe BS, et al. *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. *J Infect Dis*. 2000; 182: 252–259. <https://doi.org/10.1086/315652> PMID: 10882604
29. Nielsen MA, Staalsoe T, Kurtzhals J, Goka BQ, Doodoo D, Alifrangis M, et al. *Plasmodium falciparum* variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. *The Journal of Immunology*. *Am Assoc Immunol*; 2002; 168: 3444–3450. PMID: 11907103
30. Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C. Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med*. 1999; 5: 340–343. <https://doi.org/10.1038/6560> PMID: 10086393
31. Gupta S, Snow RW, Donnelly C, Newbold C. Acquired immunity and postnatal clinical protection in childhood cerebral malaria. *Proc Biol Sci*. 1999; 266: 33–38. <https://doi.org/10.1098/rspb.1999.0600> PMID: 10081156
32. 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; 498: 502–505. <https://doi.org/10.1038/nature12216> PMID: 23739325
33. 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; 8: 839–850. <https://doi.org/10.15252/emmm.201606188> PMID: 27354391
34. Bernabeu M, Danziger SA, Avril M, Vaz M, Babar PH, Brazier AJ, et al. Severe adult malaria is associated with specific PfEMP1 adhesion types and high parasite biomass. *Proc Natl Acad Sci U S A*. 2016; <https://doi.org/10.1073/pnas.1524294113> PMID: 27185931
35. 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; 17: 118–129. <https://doi.org/10.1016/j.chom.2014.11.007> PMID: 25482433
36. Ghumra A, Semblat J-P, Ataide R, Kifude C, Adams Y, Claessens A, et al. Induction of strain-transcending antibodies against Group A PfEMP1 surface antigens from virulent malaria parasites. *PLoS Pathog*. 2012; 8: e1002665. <https://doi.org/10.1371/journal.ppat.1002665> PMID: 22532802
  37. Rowe JA, Moulds JM, Newbold CI, Miller LH. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature*. 1997; 388: 292–295. <https://doi.org/10.1038/40888> PMID: 9230440
  38. 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; 97: 1766–1771. <https://doi.org/10.1073/pnas.040545897> PMID: 10677532
  39. 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*. 2013; 190: 240–249. <https://doi.org/10.4049/jimmunol.1202578> PMID: 23209327
  40. Howell DP-G, Levin EA, Springer AL, Kraemer SM, Phippard DJ, Schief WR, et al. Mapping a common interaction site used by *Plasmodium falciparum* Duffy binding-like domains to bind diverse host receptors. *Mol Microbiol*. 2008; 67: 78–87. <https://doi.org/10.1111/j.1365-2958.2007.06019.x> PMID: 18047571
  41. Oleinikov AV, Amos E, Frye IT, Rosnagle E, Mutabingwa TK, Fried M, et al. High throughput functional assays of the variant antigen PfEMP1 reveal a single domain in the 3D7 *Plasmodium falciparum* genome that binds ICAM1 with high affinity and is targeted by naturally acquired neutralizing antibodies. *PLoS Pathog*. 2009; 5: e1000386. <https://doi.org/10.1371/journal.ppat.1000386> PMID: 19381252
  42. 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; 21: 403–414. <https://doi.org/10.1016/j.chom.2017.02.009> PMID: 28279348
  43. Ochola LB, Siddondo BR, Ocholla H, Nkya S, Kimani EN, Williams TN, et al. Specific receptor usage in *Plasmodium falciparum* cytoadherence is associated with disease outcome. *PLoS ONE*. 2011; 6: e14741. <https://doi.org/10.1371/journal.pone.0014741> PMID: 21390226
  44. Turner GD, Morrison H, Jones M, Davis TM, Looareesuwan S, Buley ID, et al. An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J Pathol*. 1994; 145: 1057–1069. PMID: 7526692
  45. Magallón-Tejada A, Machevo S, Cisteró P, Lavstsen T, Aide P, Rubio M, et al. Cytoadhesion to gC1qR through *Plasmodium falciparum* Erythrocyte Membrane Protein 1 in Severe Malaria. *PLoS Pathog*. 2016; 12: e1006011. <https://doi.org/10.1371/journal.ppat.1006011> PMID: 27835682
  46. Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL, et al. A restricted subset of *var* genes mediates adherence of *Plasmodium falciparum*-infected erythrocytes to brain endothelial cells. *Proc Natl Acad Sci U S A*. 2012; 109: E1782–E1790. <https://doi.org/10.1073/pnas.1120534109> PMID: 22619321
  47. 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; 109: E1772–81. <https://doi.org/10.1073/pnas.1120461109> PMID: 22619330
  48. 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; 109: E1791–E1800. <https://doi.org/10.1073/pnas.1120455109> PMID: 22619319
  49. Avril M, Brazier AJ, Melcher M, Sampath S, Smith JD. DC8 and DC13 *var* genes associated with severe malaria bind avidly to diverse endothelial cells. *PLoS Pathog*. 2013; 9: e1003430. <https://doi.org/10.1371/journal.ppat.1003430> PMID: 23825944
  50. Bull PC, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, et al. *Plasmodium falciparum* antigenic variation. Mapping mosaic *var* gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol Microbiol*. 2008; 68: 1519–1534. <https://doi.org/10.1111/j.1365-2958.2008.06248.x> PMID: 18433451
  51. Rorick MM, Rask TS, Baskerville EB, Day KP, Pascual M. Homology blocks of *Plasmodium falciparum* *var* genes and clinically distinct forms of severe malaria in a local population. *BMC Microbiol*. 2013; 13: 244. <https://doi.org/10.1186/1471-2180-13-244> PMID: 24192078
  52. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM. Malaria. *Lancet*. 2014; 383: 723–735. [https://doi.org/10.1016/S0140-6736\(13\)60024-0](https://doi.org/10.1016/S0140-6736(13)60024-0) PMID: 23953767



53. Yang Y, Ya Y, Smith SA. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics*. 2013; 14: 328. <https://doi.org/10.1186/1471-2164-14-328> PMID: 23672450
54. Duffy MF, Byrne TJ, Carret C, Ivens A, Brown GV. Ectopic recombination of a malaria *var* gene during mitosis associated with an altered *var* switch rate. *J Mol Biol*. 2009; 389: 453–469. <https://doi.org/10.1016/j.jmb.2009.04.032> PMID: 19389407
55. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014; 15: R29. <https://doi.org/10.1186/gb-2014-15-2-r29> PMID: 24485249
56. Smyth GK. *limma: Linear Models for Microarray Data*. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer New York; 2005. pp. 397–420.
57. López-Barragán MJ, Lemieux J, Quiñones M, Williamson KC, Molina-Cruz A, Cui K, et al. Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics*. 2011; 12: 587. <https://doi.org/10.1186/1471-2164-12-587> PMID: 22129310
58. Joice R, Narasimhan V, Montgomery J, Sidhu AB, Oh K, Meyer E, et al. Inferring Developmental Stage Composition from Gene Expression in Human Malaria. *PLoS Comput Biol*. 2013; 9: e1003392. <https://doi.org/10.1371/journal.pcbi.1003392> PMID: 24348235
59. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010; 11: R25. <https://doi.org/10.1186/gb-2010-11-3-r25> PMID: 20196867
60. Gagnon-Bartsch J, Jacob L, Speed T. Removing Unwanted Variation from High Dimensional Data with Negative Controls. University of California Berkeley; 2013 Dec.
61. Marchetti RV, Lehane AM, Shafik SH, Winterberg M, Martin RE, Kirk K. A lactate and formate transporter in the intraerythrocytic malaria parasite, *Plasmodium falciparum*. *Nat Commun*. 2015; 6: 6721. <https://doi.org/10.1038/ncomms7721> PMID: 25823844
62. Yeo TW, Lampah DA, Gitawati R, Tjitra E, Kenangalem E, McNeil YR, et al. Impaired nitric oxide bioavailability and L-arginine reversible endothelial dysfunction in adults with *falciparum* malaria. *J Exp Med*. 2007; 204: 2693–2704. <https://doi.org/10.1084/jem.20070819> PMID: 17954570
63. Anstey NM, Weinberg JB, Hassanali MY, Mwaikambo ED, Manyenga D, Misukonis MA, et al. Nitric oxide in Tanzanian children with malaria: inverse relationship between malaria severity and nitric oxide production/nitric oxide synthase type 2 expression. *J Exp Med*. 1996; 184: 557–567. PMID: 8760809
64. Lopansri BK, Anstey NM, Weinberg JB, Stoddard GJ, Hobbs MR, Levesque MC, et al. Low plasma arginine concentrations in children with cerebral malaria and decreased nitric oxide production. *Lancet*. 2003; 361: 676–678. [https://doi.org/10.1016/S0140-6736\(03\)12564-0](https://doi.org/10.1016/S0140-6736(03)12564-0) PMID: 12606182
65. Yeo TW, Florence SM, Kalingonji AR, Chen Y, Granger DL, Anstey NM, et al. Decreased Microvascular Function in Tanzanian Children With Severe and Uncomplicated *Falciparum* Malaria. *Open Forum Infectious Diseases*. Oxford University Press US; 2017. p. ofx079. <https://doi.org/10.1093/ofid/ofx079> PMID: 28852670
66. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997; 278: 680–686. PMID: 9381177
67. MacRae JI, Dixon MW, Dearnley MK, Chua HH, Chambers JM, Kenny S, et al. Mitochondrial metabolism of sexual and asexual blood stages of the malaria parasite *Plasmodium falciparum*. *BMC Biol*. 2013; 11: 67. <https://doi.org/10.1186/1741-7007-11-67> PMID: 23763941
68. Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, et al. PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature*. 2013; 499: 223–227. <https://doi.org/10.1038/nature12361> PMID: 23823717
69. Chen PB, Ding S, Zanghi G, Soulard V, DiMaggio PA, Fuchter MJ, et al. *Plasmodium falciparum* PfSET7: enzymatic characterization and cellular localization of a novel protein methyltransferase in sporozoite, liver and erythrocytic stage parasites. *Sci Rep*. 2016; 6: 21802. <https://doi.org/10.1038/srep21802> PMID: 26902486
70. Fan Q, Miao J, Cui L, Cui L. Characterization of PRMT1 from *Plasmodium falciparum*. *Biochem J*. 2009; 421: 107–118. <https://doi.org/10.1042/BJ20090185> PMID: 19344311
71. Brancucci NMB, Bertschi NL, Zhu L, Niederwieser I, Chin WH, Wampfler R, et al. Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe*. 2014; 16: 165–176. <https://doi.org/10.1016/j.chom.2014.07.004> PMID: 25121746
72. Krungkrai SR, Krungkrai J. Insights into the pyrimidine biosynthetic pathway of human malaria parasite *Plasmodium falciparum* as chemotherapeutic target. *Asian Pac J Trop Med*. 2016; 9: 525–534. <https://doi.org/10.1016/j.apjtm.2016.04.012> PMID: 27262062
73. Muralidharan V, Oksman A, Pal P, Lindquist S, Goldberg DE. *Plasmodium falciparum* heat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. *Nat Commun*. 2012; 3: 1310. <https://doi.org/10.1038/ncomms2306> PMID: 23250440



74. Nagy GN, Marton L, Krámos B, Oláh J, Révész Á, Vékey K, et al. Evolutionary and mechanistic insights into substrate and product accommodation of CTP:phosphocholine cytidyltransferase from *Plasmodium falciparum*. *FEBS J.* 2013; 280: 3132–3148. <https://doi.org/10.1111/febs.12282> PMID: 23578277
75. Ukaegbu UE, Kishore SP, Kwiatkowski DL, Pandarinath C, Dahan-Pasternak N, Dzikowski R, et al. Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in *P. falciparum*. *PLoS Pathog.* 2014; 10: e1003854. <https://doi.org/10.1371/journal.ppat.1003854> PMID: 24391504
76. Oberli A, Zurbrugg L, Rusch S, Brand F, Butler ME, Day JL, et al. *Plasmodium falciparum* PHIST Proteins Contribute to Cytoadherence and Anchor PfEMP1 to the Host Cell Cytoskeleton. *Cell Microbiol.* 2016; 18: 1415–1428. <https://doi.org/10.1111/cmi.12583> PMID: 26916885
77. Dixon MW, Kenny S, McMillan PJ, Hanssen E, Trenholme KR, Gardiner DL, et al. Genetic ablation of a Maurer's cleft protein prevents assembly of the *Plasmodium falciparum* virulence complex. *Mol Microbiol.* 2011; 81: 982–993. <https://doi.org/10.1111/j.1365-2958.2011.07740.x> PMID: 21696460
78. Spycher C, Rug M, Pachlatko E, Hanssen E, Ferguson D, Cowman AF, et al. The Maurer's cleft protein MAHRP1 is essential for trafficking of PfEMP1 to the surface of *Plasmodium falciparum*-infected erythrocytes. *Mol Microbiol.* 2008; 68: 1300–1314. <https://doi.org/10.1111/j.1365-2958.2008.06235.x> PMID: 18410498
79. Kats LM, Proellocks NI, Buckingham DW, Blanc L, Hale J, Guo X, et al. Interactions between *Plasmodium falciparum* skeleton-binding protein 1 and the membrane skeleton of malaria-infected red blood cells. *Biochim Biophys Acta.* 2015; 1848: 1619–1628. <https://doi.org/10.1016/j.bbamem.2015.03.038> PMID: 25883090
80. Kulzer S, Charnaud S, Dagan T, Riedel J, Mandal P, Pesce ER, et al. *Plasmodium falciparum*-encoded exported hsp70/hsp40 chaperone/co-chaperone complexes within the host erythrocyte. *Cell Microbiol.* 2012; 14: 1784–1795. <https://doi.org/10.1111/j.1462-5822.2012.01840.x> PMID: 22925632
81. Acharya P, Chaubey S, Grover M, Tatu U. An exported heat shock protein 40 associates with pathogenesis-related knobs in *Plasmodium falciparum* infected erythrocytes. *PLoS ONE.* 2012; 7: e44605. <https://doi.org/10.1371/journal.pone.0044605> PMID: 22970262
82. Reddy KS, Amlabu E, Pandey AK, Mitra P, Chauhan VS, Gaur D. Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for *Plasmodium falciparum* erythrocyte invasion. *Proc Natl Acad Sci U S A.* 2015; 112: 1179–1184. <https://doi.org/10.1073/pnas.1415466112> PMID: 25583518
83. Taechalerpaisarn T, Crosnier C, Bartholdson SJ, Hodder AN, Thompson J, Bustamante LY, et al. Biochemical and functional analysis of two *Plasmodium falciparum* blood-stage 6-cys proteins: P12 and P41. *PLoS ONE.* 2012; 7: e41937. <https://doi.org/10.1371/journal.pone.0041937> PMID: 22848665
84. Siau A, Silvie O, Franetich JF, Yalaoui S, Marinach C, Hannoun L, et al. Temperature shift and host cell contact up-regulate sporozoite expression of *Plasmodium falciparum* genes involved in hepatocyte infection. *PLoS Pathog.* 2008; 4: e1000121. <https://doi.org/10.1371/journal.ppat.1000121> PMID: 18688281
85. Madeira L, Galante PA, Budu A, Azevedo MF, Malnic B, Garcia CR. Genome-wide detection of serpentine receptor-like proteins in malaria parasites. *PLoS ONE.* 2008; 3: e1889. <https://doi.org/10.1371/journal.pone.0001889> PMID: 18365025
86. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
87. Petter M, Lee CC, Byrne TJ, Boysen KE, Volz J, Ralph SA, et al. Expression of *P. falciparum* var genes involves exchange of the histone variant H2A.Z at the promoter. *PLoS Pathog.* 2011; 7: e1001292. <https://doi.org/10.1371/journal.ppat.1001292> PMID: 21379342
88. Duffy MF, Noviyanti R, Tsuboi T, Feng Z-P, Trianty L, Sebayang BF, et al. Differences in PfEMP1s recognized by antibodies from patients with uncomplicated or severe malaria. *Malar J.* 2016; 15: 258. <https://doi.org/10.1186/s12936-016-1296-4> PMID: 27149991
89. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* 2014; 15: 410. <https://doi.org/10.1186/s13059-014-0410-6> PMID: 25063469
90. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
91. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 2009; 23: 205–211. PMID: 20180275
92. Warimwe GM, Keane TM, Fegan G, Musyoki JN, Newton CRJC, Pain A, et al. *Plasmodium falciparum* var gene expression is modified by host immunity. *Proc Natl Acad Sci U S A.* 2009; 106: 21801–21806. <https://doi.org/10.1073/pnas.0907590106> PMID: 20018734

93. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010; 26: 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461> PMID: 20709691
94. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann Stat*. 2001; 29: 1165–1188.
95. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*. 2008; 24: 719–720. <https://doi.org/10.1093/bioinformatics/btm563> PMID: 18024473
96. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
97. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013; 41: e108. <https://doi.org/10.1093/nar/gkt214> PMID: 23558742
98. Larremore DB, Clauset A, Buckee CO. A network approach to analyzing highly recombinant malaria parasite genes. Antia R, editor. *PLoS Comput Biol*. 2013; 9: e1003268. <https://doi.org/10.1371/journal.pcbi.1003268> PMID: 24130474
99. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol*. 2011; 7: e1002195. <https://doi.org/10.1371/journal.pcbi.1002195> PMID: 22039361
100. Neuwald AF, Altschul SF. Bayesian Top-Down Protein Sequence Alignment with Inferred Position-Specific Gap Penalties. *PLoS Comput Biol*. 2016; 12: e1004936. <https://doi.org/10.1371/journal.pcbi.1004936> PMID: 27192614
101. Wheeler TJ, Clements J, Finn RD. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics*. 2014; 15: 7. <https://doi.org/10.1186/1471-2105-15-7> PMID: 24410852
102. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22: 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158> PMID: 16731699
103. Chou ES, Abidi SZ, Teye M, Leliwa-Sytek A, Rask TS, Cobbold SA, et al. A high parasite density environment induces transcriptional changes and cell death in *Plasmodium falciparum* blood stages. *FEBS J*. 2017; <https://doi.org/10.1111/febs.14370> PMID: 29281179
104. Bruce MC, Donnelly CA, Alpers MP, Galinski MR, Barnwell JW, Walliker D, et al. Cross-species interactions between malaria parasites in humans. *Science*. 2000; 287: 845–848. PMID: 10657296
105. Rovira-Graells N, Aguilera-Simón S, Tintó-Font E, Cortés A. New Assays to Characterise Growth-Related Phenotypes of *Plasmodium falciparum* Reveal Variation in Density-Dependent Growth Inhibition between Parasite Lines. *PLoS ONE*. 2016; 11: e0165358. <https://doi.org/10.1371/journal.pone.0165358> PMID: 27780272
106. Khoury DS, Cromer D, Akter J, Sebina I, Elliott T, Thomas BS, et al. Host-mediated impairment of parasite maturation during blood-stage *Plasmodium* infection. *Proc Natl Acad Sci U S A*. 2017; 114: 7701–7706. <https://doi.org/10.1073/pnas.1618939114> PMID: 28673996
107. Pelle KG, Oh K, Buchholz K, Narasimhan V, Joice R, Milner DA, et al. Transcriptional profiling defines dynamics of parasite tissue sequestration during malaria infection. *Genome Med*. 2015; 7: 19. <https://doi.org/10.1186/s13073-015-0133-7> PMID: 25722744
108. Crabb BS, Cooke BM, Reeder JC, Waller RF, Caruana SR, Davern KM, et al. Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell*. 1997; 89: 287–296. PMID: 9108483
109. Maier AG, Rug M, O'Neill MT, Beeson JG, Marti M, Reeder J, et al. Skeleton-binding protein 1 functions at the parasitophorous vacuole membrane to traffic PfEMP1 to the *Plasmodium falciparum*-infected erythrocyte surface. *Blood*. 2007; 109: 1289–1297. <https://doi.org/10.1182/blood-2006-08-043364> PMID: 17023587
110. Karyana M, Burdarm L, Yeung S, Kenangalem E, Wariker N, Maristela R, et al. Malaria morbidity in Papua Indonesia, an area with multidrug resistant *Plasmodium vivax* and *Plasmodium falciparum*. *Malar J*. 2008; 7: 148. <https://doi.org/10.1186/1475-2875-7-148> PMID: 18673572
111. Yeo TW, Lampah DA, Kenangalem E, Tjitra E, Price RN, Anstey NM. Impaired skeletal muscle microvascular function and increased skeletal muscle oxygen consumption in severe *falciparum* malaria. *J Infect Dis*. 2013; 207: 528–536. <https://doi.org/10.1093/infdis/jis692> PMID: 23162136
112. Goldman IF, Qari SH, Skinner J, Oliveira S, Nascimento JM, Póvoa MM, et al. Use of glass beads and CF 11 cellulose for removal of leukocytes from malaria-infected human blood in field settings. *Mem Inst Oswaldo Cruz*. 1992; 87: 583–587. PMID: 1343674
113. Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, et al. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics*. 2012; 13: 1. <https://doi.org/10.1186/1471-2164-13-1> PMID: 22214261

114. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. 2014; 30: 1660–1666. <https://doi.org/10.1093/bioinformatics/btu077> PMID: 24532719
115. Huang X, Madan A. CAP3: A DNA Sequence Assembly Program. *Genome Res*. 1999; 9: 868–877. PMID: 10508846
116. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30: 923–930. <https://doi.org/10.1093/bioinformatics/btt656> PMID: 24227677
117. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics*. 2012; 13: 539–552. <https://doi.org/10.1093/biostatistics/kxr034> PMID: 22101192
118. Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, et al. NSR-seq transcriptional profiling enables identification of a gene signature of *Plasmodium falciparum* parasites infecting children. *J Clin Invest*. 2011; 121: 1119–1129. <https://doi.org/10.1172/JCI43457> PMID: 21317536
119. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11: R106. <https://doi.org/10.1186/gb-2010-11-10-r106> PMID: 20979621
120. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10: R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: 19261174
121. Cobbold SA, Chua HH, Nijagal B, Creek DJ, Ralph SA, McConville MJ. Metabolic Dysregulation Induced in *Plasmodium falciparum* by Dihydroartemisinin and Other Front-Line Antimalarial Drugs. *J Infect Dis*. 2016; 213: 276–286. <https://doi.org/10.1093/infdis/jiv372> PMID: 26150544
122. Clasquin MF, Melamud E, Rabinowitz JD. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinformatics*. 2012;Chapter 14: Unit14.11.