Volume 10:1-10

The Undergraduate Journal of Experimental Microbiology & Immunology (+Peer Reviewed)

RNA-seq Analysis of *Plasmodium falciparum* Shows Upregulated Genes Involved in Merozoite Egress and Erythrocyte Invasion

Ethan Clements[1, 2], Gabriella Pflederer[2], Lena Shehayber[3]

[1]Department of Chemistry, University of Wisconsin Parkside, Parkside, WI, USA
[2]Department of Biology and Chemistry, Greenville University, Greenville IL, USA
[3]Department of Biological Sciences, Benedictine University, Lisle, IL, USA.

SUMMARY Malaria is a febrile disease caused by parasites of the genus *Plasmodium*. The most severe cases and most of the deaths of malaria are due to *Plasmodium falciparum* infection. Investigating transcriptional differences in *P. falciparum* between severe and uncomplicated cases can shed light on malaria pathogenicity and virulence. In this study we used previously acquired publicly available RNA sequencing data from severe and uncomplicated cases of *P. falciparum* infection. Differential expression analysis of the *P. falciparum* transcriptome showed that a group of gene products involved in merozoite egress and erythrocyte invasion were significantly upregulated in severe malaria. More specifically, we identified that Subtilisin-like protease 1 (SUB1), merozoite surface proteins 1 and 2 (MSP1 and MSP2), and serine repeat antigen 5 (SERA5) were upregulated in severe malaria. Upregulation of genes in merozoite egress and erythrocyte invasion suggest molecular mechanisms of malaria severity specifically during the blood stages. In addition, upregulation of tyrosine kinase-like protein (TKL) transcription in conjunction with SERA5 regulatory proteins upregulation suggests additional mechanisms of SERA5-mediated malaria severity.

INTRODUCTION

Malaria is a parasitic infection which affects approximately 500 million people annually (1). Most malaria cases are found within Africa, though they are also commonly found in South-East Asia, Eastern Mediterranean, and Western Pacific. The disease is caused by the *Plasmodium* species, which is transmitted through the saliva of a female *Anopheles* mosquito. There are five species of *Plasmodium* which infect humans, *Plasmodium vivax* and *Plasmodium falciparum* being the most common ones (2). Within the South-East Asia region, about 50% of cases are coincidental with *P. falciparum* infection, and *P. vivax* infection is less common.

The parasite exists in several life cycle stages, in both mosquito and human hosts (Fig. 1). In this study we focus on the life cycle stages that exist within the human host, specifically the sporozoite and merozoite stages. The parasite is in the sporozoite stage when introduced to the human body through *Anopheles* saliva. The sporozoites travel to the liver, where they infect hepatocytes (liver cells) and evade filtering out of the bloodstream. After maturing into merozoites, the parasite egresses the liver and enters into the blood stage by attacking erythrocytes. Once in the blood stage, the parasite either remains in the asexual stage (merozoite) to continue infecting healthy erythrocytes (becoming a trophozoite) or progresses to the sexual stage (gametocyte) waiting to infect healthy mosquitos. In order to progress

Published Online: September 2024

Citation: Clements, Pflederer, and Shehayber. 2024. RNA-seq analysis of *Plasmodium falciparum* shows upregulated genes involved in merozoite egress and erythrocyte invasion. UJEMI+ 10:1-10

Editor: Shruti Sandilya and Ronja Kothe, University of British Columbia

Address correspondence to:
https://jemi.microbiology.ubc.ca/

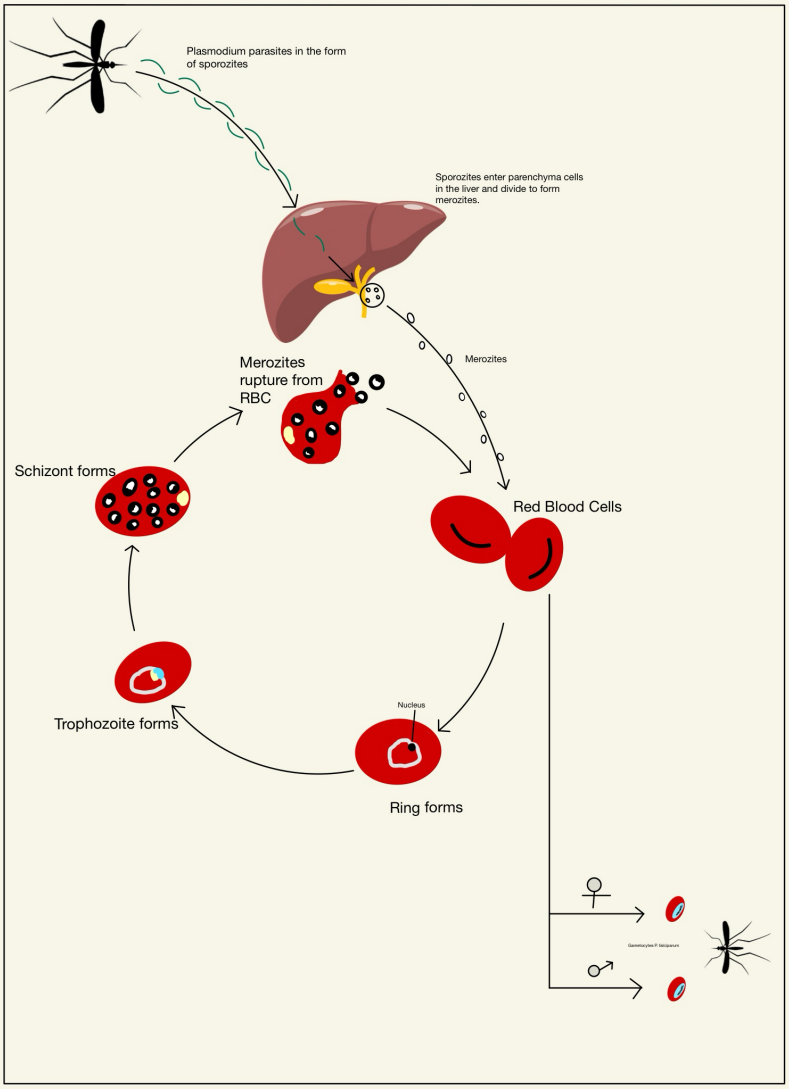

FIG. 1 *Plasmodium* life cycle within the human host. Sporozoites enter host in *Anopheles* saliva. Sporozoites then migrate to the liver, where they invade hepatocytes. While inside hepatocytes, sporozoites mature into merozoites. Mature merozoites egress the hepatocytes and enter the bloodstream, where they invade erythrocytes. In erythrocytes trophozoites develop. At this point, some of the trophozoites will enter the sexual stage of the life cycle and differentiate into gametocytes. Gametocytes then stay in the bloodstream until they enter the mosquito host. Gametocytes play no role in pathogenesis. The majority of trophozoites develop into merozoites, which then egress the erythrocyte, and invade new erythrocytes.

through these stages, the *Plasmodium* can change expression of many different genes over the course of its life. The predominant life stage present in the samples from this study are the erythrocytic ring and trophozoite stages (3), which are a part of the asexual reproductive blood cycle stages.

When malaria is diagnosed, it is categorized as being either an uncomplicated or a severe case. Severe malaria is categorized by clinical manifestations meeting certain criteria, as published by the World Health Organization (1). Some examples of these symptoms include cerebral malaria, impaired consciousness, and severe anemia (4). Most severe cases of malaria are caused by one species of *Plasmodium* in particular, *P. falciparum* (1, 2, 4). However, some cases of *P. falciparum* infection are severe while others are uncomplicated, and the genetic mechanisms behind the variety of clinical manifestation are not completely understood.

Previous genome-wide studies of gene expression in malaria conclude that surface proteins involved in the recognition of the parasite with the host are associated vary among *Plasmodium* strains and can be associated with disease symptoms (5–8). We sought to confirm and expand upon these finding using RNA-sequencing data, which is more sensitive compared to DNA microarray data that was previously used. Previous studies also focused on patients within Africa. Our dataset comes from Indonesia, where the *Plasmodium* strains seem to have unique gene expression profiles compared to African strains (9). In studying

this dataset, we seek to contribute to understanding how genetic regulation in *P. falciparum* might contain unique features from another part of the world.

A study by Tonkin-Hill et al. has described the use of comparative transcriptomics to study the expression patterns of *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) in severe and uncomplicated malaria (3). Tonkin-Hill et al. found that PfEMP1 was upregulated in severe malaria and discovered novel highly-expressed sequences (3). Our study uses the same dataset as Tonkin-Hill et al., but with several different analytical approaches. We sought to reanalyze the data using a modified analysis pipeline that both validates and builds from the findings of Tonkin-Hill et al to focus specifically on genes relating to merozoite egress and erythrocyte invasion.

METHODS AND MATERIALS

All data for this study was collected by Tonkin-Hill et al. from malaria patients in Papua, Indonesia (3). A set of 44 transcriptomes was obtained from blood serum samples of 23 patients with severe malaria and 21 patients with uncomplicated malaria. The sequences were retrieved from the European Nucleotide Archive. The Galaxy cluster (10) was used to manipulate all transcriptomes, and a pipeline for the assembly of the *P. falciparum* genome was adapted from Tonkin Hill et al, 2018. Raw pair-end reads were trimmed with Trim Galore! (11) to remove Illumina primers that were still attached to the ends of sequences. HISAT2 (12) was then used to align the trimmed reads to three different reference genomes: build 63 of *P. vivax*, the hg38 human genome, and build 63 of *P. falciparum* 3D7. The human and *P. vivax* alignments were then converted to BED format (13). BAM filter (14) was used to remove any reads that were mapped to either *P. vivax* or the human genome. After filtering, featurecounts (15) was used to prepare a table containing counts of features such as genes, exons, non-coding regions, etc. The reads that mapped to either the *P. vivax* or the human genomes were then removed in a filtering process (Fig. 2). The samples from blood serum

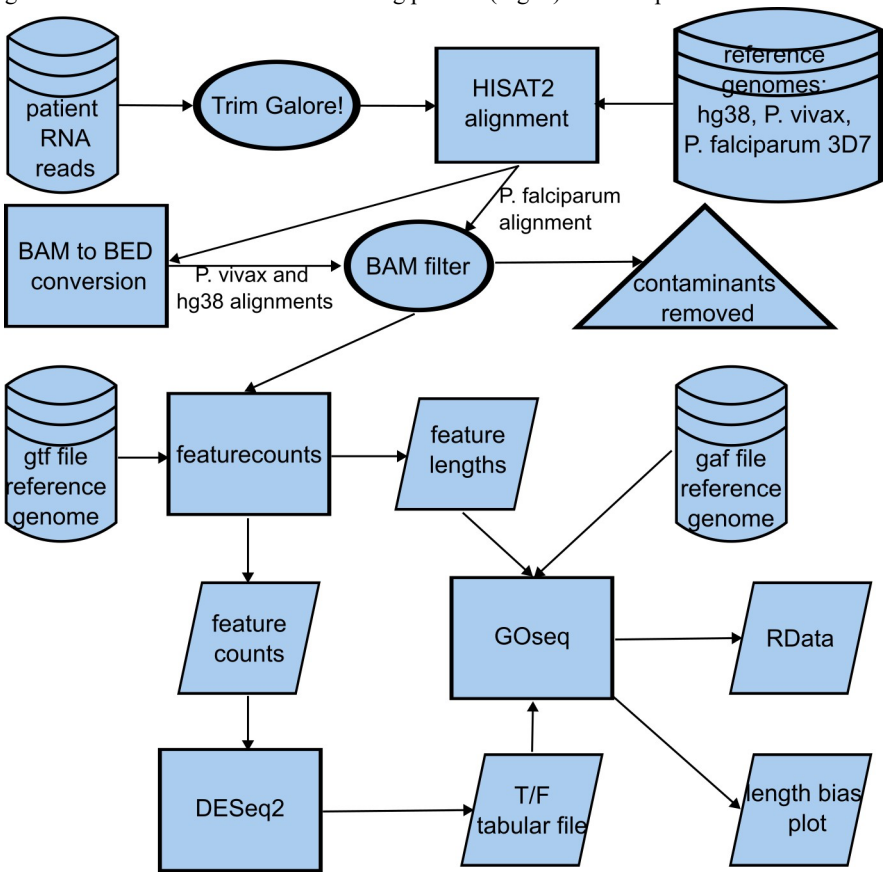

FIG. 2 Differential expression analysis workflow. The tools (rectangles and ovals) used to process RNA-sequencing reads are depicted according to how the work was completed. Tools used in the Galaxy platform include: Trim Galore (trims reads); HISAT2 (alignment tools); BAM to BED (file converter); BAM filter (removes reads according to a formula); FeatureCounts (quantifies gene expression); DEseq2 (tests differentially expressed features); GOseq (tests for overrepresented functional gene categories). Cylinders represent data sources of raw RNA-sequencing reads, human, *P. vivax*, and *P. falciparum* reference genomes, and genome feature files. Parallelograms depict specific output files that were used for later analysis.

were largely contaminated with other genetic material, such as the human transcriptome or other strains of *Plasmodium* (Table 1). We fed individual filtered sample reads through our

TABLE. 1 Removal of Contaminant Sequences. Total sample sequencing reads in severe and uncomplicated malaria blood serum samples are listed as they went through filtering. Sequences that were kept after the filtering aligned to the *P. falciparum* genome and not to the contaminant genomes. Sequences that were filtered out and removed as contaminants aligned to either the *P. vivax* or the human genomes.

	Kept	Removed	Total	Percent Removed
Total Severe Reads	289,349,422	63,560,923	352,910,345	18.01
Total Uncomplicated reads	294,357,145	99,549,008	393,906,153	25.27

pipeline and compared them together when inputting into the DESeq2 tool (severe versus uncomplicated, Fig. 2) (16). DESeq2 (16) outputs a list of genes which are differentially expressed, using the Wald Test to create adjusted p-values according to the logarithm base 2 of the fold-change (Log2FC) between severe and uncomplicated cases of malaria and the lengths of the genes (Suppl. Table S1). DESeq2 performs the differential analysis and outputs a dispersion plot to ensure that the assumptions of using the tool were met (Suppl. Fig. S1). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database categorizes genes according to their biological function. The functions of the differentially expressed genes were found using the KEGG database (Suppl. Table S2). Data analysis and visualization that compares different levels of RNA expression was done in R. When no KEGG term was available for a gene, a Refseq name was used. We also used the Refseq name to individually search the literature for biological functions where the other database information could not be found (Suppl. Table S2). This comprehensive categorization of differentially expressed genes allowed us to identify patterns of biological significance.

RESULTS

Sequencing reads were aligned and filtered before differential-expression analysis. The focus of our study was to determine what gene expression differences were present in the main parasite for malaria, *P. falciparum,* that is associated with severity of the disease. In a malaria-infected population, individuals may contain parasites of multiple species, such as *P. vivax*, which is common in Indonesia. Therefore, we used HISAT2 to align reads to several reference genomes (Fig. 2). The reads that mapped to either the *P. vivax* genome or the human genome were then removed in a filtering process (Fig. 2). The samples from blood serum were largely contaminated with other genetic material, such as the human transcriptome or other strains of *Plasmodium* (Table 1). A total of 289,349,422 contaminants were removed from the transcriptome of severe malaria samples, which was 18.01% of the total reads (Table 1). A total of 294,357,145 contaminants were removed from the transcriptome of uncomplicated malaria samples, which was 25.27% of the total reads (Table 1).

DEseq2 results show 1073 upregulated genes in severe malaria. The featurecounts tool counts the features present in the assembled genome, such as genes and exons (Fig. 2, 16). We fed individual filtered sample reads through our pipeline and synthesized them together when inputting into the DESeq2 tool (Fig. 2). DESeq2 performs the differential analysis and outputs a dispersion plot (Suppl. Fig. S1) which shows that the assumptions of using the tool were met. DESeq2 outputted 1174 genes that were found to have a significant change ($p < 0.05$) in expression between severe and uncomplicated malaria samples (Suppl. Table S1). There are significantly more genes which are upregulated than downregulated in severe malaria cases by a factor of 10 (Suppl. Table S1, Suppl. Table S1, X^2 test $p < 0.0001$).

Graphical depictions of the results in terms of magnitude and significance of differential expression are shown in the MA plot (Suppl. Fig. S1) and in a volcano plot (Fig. 3) created to better visualize the data. The difference between the $\log_2$ intensities (severe minus uncomplicated) of each genetic feature (M) is plotted on the Y-axis. Averages of the gene expression $\log_2$ intensity (A) are plotted on the X-axis. Most of the data points are centralized around y=0, which signifies that the majority of analyzed genes have no major difference between them (Suppl. Fig. 2, Suppl. Table S1). The distance between the data points and the

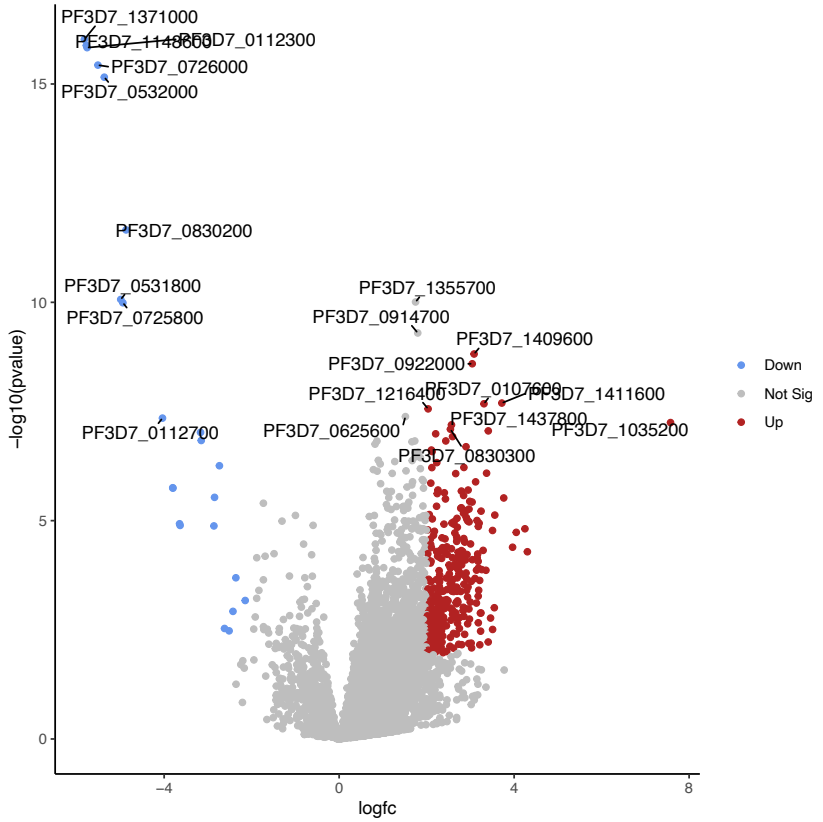

FIG. 3 Volcano plot. The volcano plot shows the significance of difference in expression (severe minus uncomplicated) versus the $\log_2$ of the fold change. Red points indicate significantly upregulated genes, while blue points represent significantly down regulated genes, at the $p<0.05$ (DESeq2) level. The GeneIDs with the lowest p-values are labeled. NA (uncolored) denotes that for some genes, either the featurecounts tool was unable to find values in at least one sample, or the adjusted p-value could not be calculated due to variable gene lengths.

y = 0 line increases, which indicates the magnitude of expression difference. Genes that are significantly ($p \leq 0.01$) differentially expressed are shown in red. Points above the y = 0 line are genes that are upregulated, and points below are downregulated.

The volcano plot is a scatterplot of DESeq2 outputs according to statistical significance and log fold change ($\log_2$FC) values of severe versus uncomplicated sample expression and (Fig. 3, blue and red dots) (17). Grey points towards the bottom of the plot do not have statistically significant differential expression ($p \geq 0.05$, Fig. 3). The twenty genes with the most significant changes in expression have their GeneIDs labeled. There are significantly more genes which are up-regulated (Fig. 3, red dots) than down-regulated (Fig. 3, blue dots) in severe malaria.

Categorization of upregulated genes by KEGG database identifies patterns of biological significance. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database categorizes genes according to their biological function. We found that the KEGG database provided more functional identifications to our differentially expressed genes than the Gene Ontology terms for *P. faciparum*.

We also used the Refseq name to individually search the literature for biological functions where the other database information could not be found (Suppl. Table S2). This comprehensive categorization of differentially expressed genes allowed us to identify patterns of biological significance. Upon conducting further literature searches, we found that several upregulated gene products in severe malaria interact with each other *in vitro*. We summarize our results of a possible biological mechanism of malaria virulence regulated by SUB1 through the upregulation of other *P. falciparum* genes in Figure 4.

Proteins involved in merozoite egress are upregulated in severe malaria. Efficient egress from the infected erythrocyte and subsequent invasion of new erythrocytes are important components of the *Plasmodium* life cycle. Merozoite egress and erythrocyte invasion are especially important with regards to the severity of malaria. The blood stage of the *Plasmodium* is the cause of all clinical symptoms (18). As such, an increase in the ability of the parasite to persist in the blood stages would be expected to increase the severity of the disease. Consistent with this hypothesis, we have observed increased transcription of four out of at least eight critical proteins (SUB1, MSP1, MSP6, MSP7, MSP8, SERA5, SERA6, and

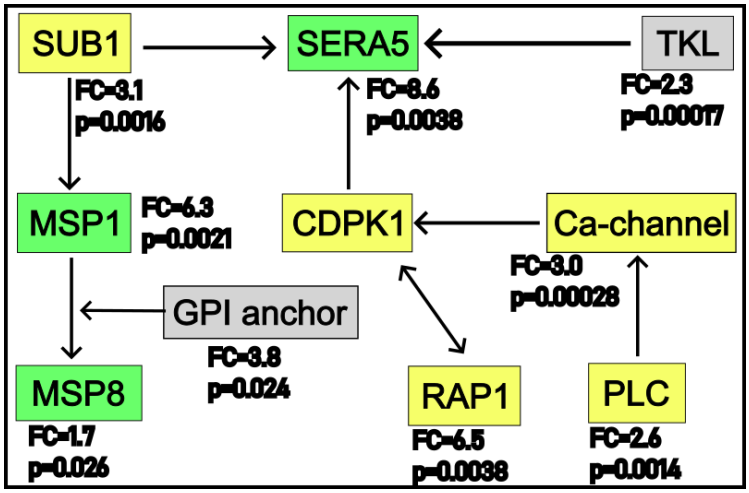

FIG. 4 Upregulated Genes Involved in Erythrocyte Egress & Invasion Genes. The green boxes indicate genes which play a direct role in erythrocyte egress and invasion. Yellow boxes indicate regulatory proteins which interact with genes involved in merozoite invasion or egress. The boxes below the gene names denote the fold change (FC) in expression of the gene in severe malaria compared to uncomplicated cases and its associated p-value, which was obtained from DESeq2. Arrows indicate positive regulation between gene products based on the current understanding of the literature.

Rh5) associated with erythrocyte egress and subsequent erythrocyte invasion in severe cases of malaria (Suppl. Table S1).

Subtilisin-like protease 1 (SUB1) is a regulatory protease, which functions to remove regulatory domains from other proteins when secreted into the parasitophorous vacuole (19). Specifically, SUB1 activates serine repeat antigen 5 (SERA5), SERA6, merozoite surface protein 1 (MSP1), MSP6, and MSP7 (19). These proteins are involved in egress from the infected erythrocytes and invasion of uninfected erythrocytes by cleaving the spectrin exoskeleton of red blood cells (20–22). SUB1 (PF3D7_0507500) is upregulated 3.1-fold ($p = 0.001558$) between severe and uncomplicated malaria blood samples. Of the targets of SUB1, we find that SERA5 (PF3D7_0207600, 8.6-fold, $p = 0.003773$), MSP1 (PF3D7_0930300, 6.3-fold, $p = 0.002094$), RAP1 (PF3D7_1410400, 6.5-fold, $p = 0.001688$), and RAP3 (PF3D7_0501500, 7.2-fold, $p = 0.01376$) are also upregulated in cases of severe malaria (Suppl. Table S1, Fig. 4). Other known targets of SUB1, such as SERA6 (PF3D7_0207500), MSP6 (PF3D7_1035500), and MSP7 (PF3D7_1335100) are not differentially expressed between our samples (Suppl. Table S1). SERA5 is involved with erythrocyte egress during the *Plasmodium* asexual blood stage. The mechanism of SERA5 function is unclear, however studies have shown its activity in merozoite egress (19,23).

MSP1 is associated with erythrocyte egress in *P. falciparum* (21). After post-translational processing by SUB1 to yield its active form, MSP1 is attached to the cell membrane using a glycosylphosphatidylinositol (GPI) anchor P1 (19). The GPI anchor P1 (PF3D7_1216500) is also upregulated in severe malaria 3.8-fold ($p = 0.02439$). Another surface protein, MSP8, attaches to the cell surface via the GPI anchor in addition to MSP1. MSP8 (PF3D7_0502400) is upregulated 1.7-fold ($p = 0.02565$) (Suppl. Table S1, Fig. 4).

Ca^{2+} pathways and other erythrocyte egress regulatory proteins are upregulated in severe malaria. Several regulatory proteins of SERA5 are also upregulated (Suppl. Table S1, Fig. 4, yellow boxes). One such gene is PF3D7_1013500, which encodes phosphoinositide-specific phospholipase C (PLC), which we found to be upregulated 2.6-fold ($p = 0.00137$). PLC is responsible for hydrolyzing diacylglycerol to form inositol triphosphate (IP_3). IP_3 stimulates the release of Ca^{2+} ions from the endoplasmic reticulum (ER), through load activated Ca^{2+} channels (23). The expression of the gene coding for Ca^{2+} channels (PF3D7_1362300) is also upregulated 3.0-fold in cases of severe malaria ($p= 0.000281$).

In addition to the Ca^{2+} regulatory pathway of SERA5, Raf kinase inhibitor protein (RKIP), encoded by PF3D7_1219700, is up regulated 5.3-fold ($p = 0.008315$) (Suppl. Table S1). Another gene related to SERA5 is tyrosine kinase like protein (TKL, PF3D7_1349300), which is upregulated 2.3-fold ($p = 0.0001718$) in severe malaria (Suppl. Table S1, Fig. 4).

DISCUSSION

We found significantly more differentially expressed genes than Tonkin-Hill et al did by 1124 (3). Reciprocally, Tonkin-Hill et al found 61 differentially expressed genes more than our analysis did. There was substantial overlap in the findings, especially with regard to the

sign (+/-) of the fold change. In other words, upregulated genes found by Tonkin-Hill et al were also found to be upregulated in severe malaria in our reanalysis. However, Tonkin-Hill et al used a cutoff of $p < 0.1$ whereas we use the cutoff of $p < 0.05$. Of the 358 differentially expressed genes that Tonkin-Hill et al found, most were down regulated and about half of their significant hits were also found by our analysis. On the other hand, the majority of our differentially expressed genes were upregulated in severe malaria and were not found by Tonkin-Hill et al. Therefore, our analysis is more comprehensive and rigorous. Tonkin-Hill et al. used the limma/Voom alignment software and found 358 differentially expressed genes (3). By contrast, we used the HISAT2 alignment method and found that there were 595 differentially expressed genes. The reason that DESeq2 was used is because it is designed for analysis of high dimensional count data. Since we are observing the entire transcriptome rather than only the var gene, we felt as if DESeq2 would yield our best and most accurate results. The large difference in differentially expressed genes found may have resulted from the difference of analysis using DEeq2 instead of limma/voom. In comparison to gene analysis, DESeq2 is known to find more genes than limma/Voom. DESeq2 uses the non-linear model while limma/Voom uses the linear model (24). Upon looking deeper into the results of a study performed by Tong, both computational tools share the gene results in upregulated and downregulated genes. However, it is noticeable that DESeq2 has found more unique genes (24). This can explain our finding of a significantly higher amount of differentially expressed genes found. The finding could also be explained by how we chose to filter the reads, including non-var genes and excluding *P. vivax* aligned reads. This analysis of entire transcriptomes and use of DESeq2 has revealed an additional 1124 differentially-expressed genes compared to Tonkin-Hill et al's study (3).

Overall, we find that there are 10-fold more upregulated than downregulated genes in *P. falciparum* in severe malaria in Indonesia (Suppl. Table S1, S2, Fig. 3). Our findings are dissimilar to what was found in previous gene expression studies of malaria virulence in which there were not significantly more upregulated genes in severe cases (5, 8, 25). These previous studies used DNA microarrays to conduct differential expression analysis, which may have limited the number and variety of transcripts detected. Furthermore, there are many differences in patient demographics in our study versus in others, such as geographic region, age, and parasite species, which can account for these differences. These caveats are further discussed in the Study Limitations section below.

Previous studies on the biochemical function of the gene products contribute to our hypothesis that malaria virulence is mediated by erythrocyte egress through SERA5 (20). DiCre-mediated excision of PF3D7_0207600 resulted in a 50% reduction in intracellular *Plasmodium* population and invasion rate, which links erythrocyte egress with increased virulence (20). MSP8 has also been implicated in erythrocyte egress, which provides further evidence for the mechanisms of *P. falciparum* virulence by erythrocyte egress and invasion (26). Although MSP1 forms a complex with MSP6 and MSP7 *in vivo*, we only found that MSP1 transcripts were upregulated in severe malaria. This suggests that the increased erythrocyte egress function of the complex is specifically regulated by MSP1 transcript over-expression. The upregulation of merozoite surface proteins in severe malaria suggests mechanisms of erythrocyte invasion and merozoite egress. Malaria virulence through erythrocyte egress and invading new host cells may also be regulated through SERA5 by crosstalk with other cell biological mechanisms. The upregulation of PLC, Ca^{2+} channels, and RKIP further support additional mechanisms of SERA5 regulation. A previous study has shown that TKL interacts with SERA5 *in vitro,* although the exact function of TKL is not known (27). Upregulation of TKL along with other regulators of SERA5 provides evidence that TKL is a regulator of SERA5 and plays a role in merozoite egress. A possible hypothesis is that TKL plays a direct role in the function of SERA5 during merozoite egress (27).

Our investigation of upregulated genes relating to surface antigen regulation by differential expression analysis has identified other possible parallel mechanisms of malaria virulence. Previous investigation has found that the role of Rh5 in erythrocyte invasion is to draw the erythrocyte closer to the surface of the *Plasmodium* with the aid of p113 protein (28). Rh5 is expressed during erythrocyte invasion and first binds to the p113 on the surface of the plasmodium (28). Afterwards, Rh5 binds to the basigin receptors on the surface of the targeted erythrocytes (28). Rh5 and p113 are both not upregulated in severe malaria (Suppl.

Table S1). The interaction between Rh5 and p113 works to draw the erythrocyte closer to the *Plasmodium* in preparation for the formation of the moving junction (28).

Limitations There are several caveats that limit our ability to apply our findings to a generalized mechanism of malaria virulence caused by *P. falciparum* gene expression. First, our data comes from one study in Indonesia, which gives us sufficient numbers to achieve significant power in our differential expression analysis. However, the mechanism of malaria virulence may be dependent on environmental effects and interactions with the host's biology, which affects how wide we can extrapolate the findings. Furthermore, Bobowik et al. 2021's meta-analysis concludes that there is a unique gene expression profile among *Plasmodium* strains in Indonesia compared to other Asian and African strains (9). Second, the majority of *the P. falciparum* cells collected from whole blood in this study were in the erythrocytic ring and trophozoite stages, which might bias our data towards mechanisms involving such related life stages. Other gene expression studies using patient blood samples find similar proportions of parasitic life cycle stages (5, 9,29). Follow-up genetic analysis on the functions of these proteins involved in merozoite egress and erythrocyte invasion can corroborate our findings. Third, our filter step might have removed aligned reads that are too similar to distinguish between *P. vivax* and *P. falciparum*. Our filtration step removed about 4% of the aligned reads that matched to the *P. vivax* genome. Therefore, our analysis may not be a complete picture of all *P. falciparum* genes that are regulated in severe and uncomplicated malaria.

Conclusions Efficient egress from the infected erythrocyte, and subsequent invasion of new erythrocytes is an important component of the *Plasmodium* life cycle. Merozoite egress and erythrocyte invasion are especially important with regards to the severity of malaria. The blood stage of the *Plasmodium* is the cause of all clinical symptoms (1). Consistent with this hypothesis, we have observed increased transcription of four out of at least eight critical proteins associated with erythrocyte egress and subsequent erythrocyte invasion in severe cases of *P. falciparum*. It is important to note that we have not been able to confirm an increase in protein abundance or function *in vivo.* These upregulated genes are Rh5, MSP1, MSP8 and SERA5 (Suppl. Table S1, Fig. 4). In addition, upregulation of the regulatory proteins of SERA5 has been shown in severe cases of *P. falciparum* (Suppl. Table S1, Fig. 4). This data show that malaria severity is correlated with the expression of these genes. In addition, CDPK1, Ca^{2+} Channels, PLC, and RAF kinase inhibitor are also up-regulated (Suppl. Table S1, Fig. 4). Several other gene expression analyses also find similar differential expression among genes involved in erythrocyte egress and invasion (5, 7,8,29). Our findings are consistent with previous studies which have been done to characterize the function and importance of these genes in *P. falciparum* (19–21,23,26–30).

Future Directions This work provides many possible new directions for wet-lab studies of the mechanisms of malaria virulence by *Plasmodium* surface proteins. Further work needs to be done to study the parasite *in vitro* to first confirm our RNA-seq analysis findings. We suggest that other groups investigate the biochemical mechanisms of these proteins and genetic interactions to discover their exact role and possible pathways in erythrocyte invasion and merozoite egress. Specifically, further investigation of the effect of artificially increasing MSP1 transcript expression would confirm our hypothesis. Biochemical and genetic research characterizing the binding and functional interactions between TKL and SERA5 is needed to more fully understand how it affects the process of merozoite egress.

ACKNOWLEDGEMENTS

The work completed in this manuscript was done under the supervision of Dr. Lindsey M.F. Mao, Assistant Professor at Benedictine University Department of Biological Sciences. We thank the Greenville University Catalyst Fund for supporting all investigators involved in this study through the Summer Undergraduate Research Experience program. We also thank Drs. William B. Ahern, Eugene A. Dunkley, Scott A. Pattenaude and Tiara G. Pérez-Morales for reading and providing edits to the manuscript. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

CONTRIBUTIONS

E.C. and G.P. contributed equally to this work in conceptualization, data analysis, literature review, and writing the manuscript. L.S. made revisions according to reviewer comments, created schematic figures, and updated the analysis with the current build of the reference genomes. L.M.F.M. designed the analysis and edited the manuscript.

REFERENCES

1. 2019. World malaria report 2019. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.
2. **Nadjm B, Behrens R.** 2012. Malaria: An Update for Physicians. Infectious Disease Clinics of North America, 26, 2: 243-259. https://doi.org/10.1016/j.idc.2012.03.010.
3. **Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, Marfurt J, Cobbold SA, Rambhatla JS, McConville MJ, Rogerson SJ, Brown GV, Price RN, Anstey NM, Papenfuss AT, Duffy MF**. 2018. The *Plasmodium falciparum* transcriptome in severe malaria reveals altered expression of genes involved in important processes including surface antigen–encoding var genes. PLOS Biology 16(3): e2004328. https://doi.org/10.1371/journal.pbio.2004328
4. **Barros Pinto MP, Marques G.** 2020. Severe malaria. Infection 48, 1: 143–146. https://doi.org/10.1007/s15010-019-01371-w
5. **Llinás M, Bozdeck Z, Wong ED, Adai AT, DeRisi JL**. 2006. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. Nucleic Acids Research, 34, 4: 1166–1173. https://doi.org/10.1093/nar/gkj517
6. **Milner DA, Pochet N, Krupka M, Williams C, Seydel K, Taylor TE, Regev A, Wirth D, Daily JP, Mesirov JP**. 2012. Transcriptional profiling of *Plasmodium falciparum* parasites from patients with severe malaria identifies distinct low vs. high parasitemic clusters. PLoS ONE 7, 7: 1–11. https://doi.org/10.1371/journal.pone.0040739
7. **Daily JP, Le Roch KG, Sarr O, Ndiaye D, Lukens A, Zhou Y, Ndir O, Mboup S, Sultan A, Winzeler EA, Wirth DF.** 2005. In Vivo Transcriptome of *Plasmodium faciparum* Reveals Overexpression of Transcripts That Encode Surface Proteins. Journal of Infectious Diseases 191, 7: 1196–1203. https://doi.org/10.1086/428289
8. **Almelli T, Nuel G, Bischoff E, Aubouy A, Elati M, Wang CW, Coppée J-Y, Ayissi GN, Basco, LK, Rogier C, Ndam NT, Tahar R**. 2014. Differences in Gene Transcriptomic Pattern of *Plasmodium faciparum* in Children with Cerebral Malaraia and Asymptomatic Carriers. PLoS ONE 9, 12: e114401.
9. **Bobowik K, Syafruddin D, Darusallam CC, Sudoyo H, Wells C, Romero IG.** 2021. Transcriptomic profiles of *Plasmodium falciparum* and *Plasmodium vivax*-infected individuals in Indonesia. bioRxiv: 2021.01.07.425684.
10. **Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čeck M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J.** 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic acids research 44, W1: W3–W10. https://doi.org/10.1093/nar/gkw343
11. **Krueger F.** 2012. Trim Galore. Babraham Institute. toolshed.g2.bx.psu.edu/repos/bgruening/trim_galore/trim_galore/0.4.3.1
12. **Kim D, Langmead B, Salzberg SL**. 2015. HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12, 4: 357–360. https://doi.org/10.1038/nmeth.3317
13. **Quinlan AR, Hall IM**. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 6: 841–842. https://doi.org/10.1093/bioinformatics/btq033
14. **Breese MR, Liu Y.** 2013. NGSUtils: A software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics 29, 4: 494–496. https://doi.org/10.1093/bioinformatics/bts731
15. **Liao Y, Smyth GK, Shi W**. 2013. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 7: 923–930. https://doi.org/10.1093/bioinformatics/btt656
16. **Love MI, Huber, W, Anders** S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 12: 1–21. https://doi.org/10.1186/s13059-014-0550-8
17. **Wickham H.** 2016. ggplot2. In Use R! Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4
18. **Autino B, Corbett Y, Castelli F**. 2012. Pathogenesis of malaria in tissues and blood. Mediterranean Journal of Hematology and Infectious Diseases 4, 1. https://doi.org/10.4084/MJHID.2012.061
19. **Withers-Martinez C, Suarez C, Fulle S, Kher S, Penzo M, Ebejer J-P, Koussis K, Hackett F, Jirgensons A, Finn P, Blackman MJ**. 2012. *Plasmodium subtilisin*-like protease 1 (SUB1): Insights into the active-site structure, specificity and function of a pan-malaria drug target.

International Journal for Parasitology 42, 6: 597–612. https://doi.org/10.1016/j.ijpara.2012.04.005

20. **Collins CR, Hackett F, Atid J, Tan MSY, Blackman MJ.** 2017. The Plasmodium falciparum pseudoprotease SERA5 regulates the kinetics and efficiency of malaria parasite egress from host erythrocytes. PLOS Pathogens 13(7): e1006453 https://doi.org/10.1371/journal.ppat.1006453
21. **Das S, Hertrich N, Perrin AJ, Withers-Martinez C, Collins CR, Jones ML, Watermeyer JM, Fobes ET, Martin SR, Saibil HR, Wright GJ, Treeck M, Epp C, Blackman MJ.** 2015. Processing of *Plasmodium falciparum* Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs. Cell Host and Microbe 18, 4: 433–444. https://doi.org/10.1016/j.chom.2015.09.007
22. **Thomas JA, Tan MSY, Bisson C, Borg A, Umrekar TR, Hackett F, Hale VL, Vizcay-Barrena G, Fleck RA, Snijders AP, Saibil HR, Blackman MJ**. 2018. A protease cascade regulates release of the human malaria parasite *Plasmodium falciparum* from host red blood cells. Nature Microbiology 3, 4: 447–455. https://doi.org/10.1038/s41564-018-0111-0
23. **Yamagishi J, Natori A, Tolba MEM, Mongan AE, Sugimoto C, Katayama T, Kawashima S, Makalowski W, Maeda R, Eshita Y, Tuda J, Suzuki Y.** 2014. Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. Genome Research 24, 9: 1433–1444. https://doi.org/10.1101/gr.158980.113
24. **Tong Y.** 2021. The comparison of limma and DESeq2 in gene analysis. E3S Web of Conferences, 271, 03058. https://doi.org/10.1051/e3sconf/202127103058
25. **Iyer GR, Singh S, Kaur I, Agarwal S, Siddiqui MA, Bansal A, Kumar G, Saini E, Paul G, Mohmmed A, Chitnis CE, Malhotra P.** 2018. Calcium-dependent phosphorylation of *Plasmodium falciparum* serine repeat antigen 5 triggers merozoite egress. Journal of Biological Chemistry 293, 25: 9736–9746. https://doi.org/10.1074/jbc.RA117.001540
26. **Drew DR, Sanders PR, Crabb BS.** 2005. *Plasmodium falciparum* merozoite surface protein 8 is a ring-stage membrane protein that localizes to the parasitophorous vacuole of infected erythrocytes. Infection and Immunity 73, 7: 3912–3922. https://doi.org/10.1128/IAI.73.7.3912-3922.2005
27. **Gnangnon B, Fréville A, Cailliau K, Leroy C, De Witte C, Tulasne D, Martoriarti A, Jung V, Guerrera IC, Marion S, Khalife J, Pierrot C**. 2019. Plasmodium pseudo-Tyrosine Kinase-like binds PP1 and SERA5 and is exported to host erythrocytes. Scientific Reports 9, 1: 1–21. https://doi.org/10.1038/s41598-019-44542-3
28. **Galaway F, Drought LG, Fala M, Cross N, Kemp AC, Rayner JC, Wright GJ**. 2017. P113 is a merozoite surface protein that binds the N terminus of *Plasmodium falciparum* RH5. Nature Communications 8. https://doi.org/10.1038/ncomms14333
29. **Lemieux JE, Gomez-Escobar N, Feller A, Carret C, Amambua-Ngwa A, Pinches R, Day F, Kyes SA, Conway DJ, Holmes CC, Newbold CI**. 2009. 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 of the United States of America 106, 18: 7559–7564. https://doi.org/10.1073/pnas.0811829106
30. **Mukherjee P, Burgio G, Heitlinger E**. 2020. Publicly available transcriptomes provide the opportunity for dual RNA-Seq meta analysis in Plasmodium infection. bioRxiv: 576116. https://doi.org/10.1101/576116