A Tale of Two Parasites: A Glimpse into the RNA Methylome of Patient-derived *P. falciparum* and *P. vivax* isolates

Priyanka Roy[1A], Sukriti Gujarati[1A], Pallavi Gupta[2,6,7], Ishaan Gupta[2], Tanmaya Mahapatra[3] Dinesh Gupta[4], Sanjay Kumar Kochar[5], Dhanpat Kumar Kochar[5], Ashis Das[1*]

[A]Joint co-first authors

[1] Molecular Parasitology and Systems Biology Laboratory, Department of Biological Sciences, Birla Institute of Technology and Science (BITS), Pilani campus, Rajasthan, India

[2]Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology (IIT), New Delhi, India

[3]Department of Computer Science & Information Systems, Birla Institute of Technology and Science (BITS), Pilani campus, Rajasthan, India

[4]Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

[5] Department of Medicine, S.P. Medical College, Bikaner, Rajasthan, India

[6] Australian Institute of Bioengineering and Nanotechnology (AIBN), The University of Queensland (UQ), St Lucia, Queensland, Australia

[7]University of Queensland – IIT Delhi Research Academy, New Delhi, India

*Correspondence: Ashis Das adas@pilani.bits-pilani.ac.in, ashisd28@gmail.com

ABSTRACT

Understanding the molecular mechanisms of the malarial parasites in hosts is crucial for developing effective treatments. Epitranscriptomic research on pathogens has unveiled the significance of RNA methylation in gene regulation and pathogenesis. We present the first report investigating methylation patterns determined using Nanopore Direct RNA Sequencing, to single-base resolution across RNA of P. falciparum and P. vivax from Malaria patients with hepatic dysfunction. Distinct RNA methylation profiles of N6-methyladenosine and 5-methylcytosine in the parasites suggest species-specific regulatory mechanisms. These modifications are unevenly present in the annotated regions of the mRNA, potentially influencing mRNA export and translation. Methylated antisense transcripts, splicing events and modified transcripts originating from apicoplast and mitochondrial genomes have been detected. We observe a striking overlap of differential methylation of isoforms with either of the modifications in both species. The correlation of these regulatory layers will decipher the post-transcriptional environment of malaria parasites in vivo and elucidate their inherent proteome plasticity.

KEYWORDS: RNA methylation, *Plasmodium falciparum*, *Plasmodium vivax*, Oxford Nanopore, Alternative splicing, m6A, m5C.

INTRODUCTION

The World Malaria Report states the global disease burden from malaria, caused by the protozoan parasites of the *Plasmodium* species, was 249 million cases in 2022.[1] Deaths from malaria remain at an unacceptably high rate post-pandemic. Clinical manifestations of malaria include fever, chills, headache, myalgias, and fatigue. However, if not treated promptly, it may also lead to a severe disease state manifesting as jaundice, seizures, renal failure, acute respiratory distress syndrome, and coma. *Plasmodium falciparum* and

Plasmodium vivax are especially responsible for most of the disease burden. While the rapid progression of case morbidity in *P. falciparum* malaria had already been well-documented, an increasing number of reports of severe clinical complications caused by *P. vivax* infections have also been reported in the past two decades.[2-7] The recent reports of *P. vivax* resurgence in the USA[8], which had been certified malaria-free since the 1970s, remain a cause for concern.

In the human host, the parasite undergoes schizogony by invading RBCs, host cell remodeling, genome replication, and RBC egress. Each step of this cascade involves a myriad of gene expressions to help the parasite survive in the host cell. While the parasite has been reported to be in a transcriptionally silent state during RBC invasion[9], transcriptomic reports from field and laboratory strains have shown differential expression of various invasion genes in parasites to survive through schizogony. Several studies have shown the involvement of long non-coding RNAs and natural antisense transcripts (NATs) as crucial in mediating gene expression, thus influencing parasite development within human hosts.[10,11,12] In addition to mRNA abundance and differential gene expression, epitranscriptomic studies in recent years have highlighted certain parasite-specific post-transcriptional modifications in *P. falciparum* as an additional layer of gene regulation.[13,14]

Alternative splicing is another potential mechanism for adding an additional layer of complexity to the parasite epitranscriptomic landscape. Nanopore sequencing of *Plasmodium* strains has recognized intron retention as the major alternative splicing event. Unlike short-read sequencing, long-read sequencing offers the advantage of assembling the complete transcriptome, further offering the opportunity to explore novel isoforms and un-cataloged splice junctions.[15,16]

This study elaborates on the RNA-specific methylation signatures in malaria parasites showing hepatic dysfunction complication by mono-infections of *P. vivax* or *P. falciparum.* The genome of *P. falciparum* is abundant in adenine and thymine bases, with the average genome-wide AT content being 81%, and that of *Plasmodium vivax* is GC-rich to >50%.[17,18] Here, we elaborate on the m5C and m6A chemical modifications of RNA in both sense and antisense transcripts to single base resolution and also present a comparative of the differentially methylated spliced isoforms in the two parasites.

METHODS

Sample Collection and processing

Venous blood samples (5ml) from patients infected with malaria were collected from S.P. Medical College, Bikaner, with the informed consent of the patients according to the guidelines of the relevant ethics committee. Malaria infection with P. vivax and P. falciparum was confirmed with RDTs (OptiMAL, Flow Inc., Portland, Oreg.) and microscopic investigation of peripheral blood films. The blood samples were separated on a Histopaque-gradient-based density separation (Histopaque 1077, Sigma Aldrich) to separate the erythrocytes from plasma and peripheral blood mononuclear cells (PBMCs). The erythrocytes were washed with sterile PBS, lysed with TRI Reagent (T9424, Sigma-Aldrich), and transported in a cold chain to BITS Pilani for further processing.

Total RNA was isolated from infected erythrocytes according to the manufacturer's protocol (Tri-Reagent, Sigma Aldrich) and resuspended in nuclease-free water. Molecular confirmation of infection with P. vivax or P. falciparum was done using 18S rRNA gene-based multiplex PCR and 28S rRNA gene-based nested PCR[19,20], and any possibility of mixed infection was negated. Viral or other causes of hepatic dysfunction were ruled out at the point of collection. This study used one sample each of P. vivax and P. falciparum. The patient's diagnosis was confirmed as hepatic dysfunction as per WHO guidelines[21] and upon our team of clinicians' consultation. The clinical characteristics of the sample are tabulated in Table S1.

Library Preparation

The concentration and purity of the isolated total RNA were measured by the NanoDrop UV/Vis spectrophotometer (SimpliNano™ spectrophotometer, GEHealthcare) and Qubit 4 Fluorometer (Qubit RNA HS Assay kit (Q32855), Invitrogen). The quality of the isolated total RNA was analyzed by denaturing agarose gel. One µg of total RNA was used for library preparation as per the manufacturer's protocol (SQK-RNA002, Oxford Nanopore Technologies), including the optional reverse transcription step recommended by ONT. ~20 ng of reverse transcribed and adapter-ligated RNA was used for loading on Spot-ON Flow Cell (R9) and sequenced on MinION sequencer (FLO-MN106D, Oxford Nanopore Technologies). Each of the two samples was sequenced on individual flow cells with repeated loading of 20 ng of the final library until exhaustion of pores, which gave a run time >=24 hours.

Downstream bioinformatics analysis

Reads generated at the end of the sequencing procedure were basecalled using the base caller integrated into the MinKNOW software (v23.07.15). After base calling, the reads categorized in the "pass" folder were concatenated and used for downstream analysis. ONT sequencing data was checked for quality using FastQCv0.12.1. The reads were indexed with Nanopolishv0.14.0 (https://github.com/jts/nanopolish) using preset commands. The reads were aligned with Minimap2v2.24 using the reference transcriptome P. falciparum 3D7 (PlasmoDBv64) and P. vivax Sal 1 (PlasmoDBv63) for the samples, PFC and PVC, respectively.[22] For the alignment of antisense transcripts, the aforementioned reference transcriptomes were reverse complemented and used for read alignment. The mapped reads for sense and antisense transcripts were visualized in Integrative Genomics Viewer (IGV, https://www.igv.org/). Methylation calling in both sense and antisense transcripts was done by m6anet (https://github.com/GoekeLab/m6anet) for m6A modification calls and CHEUI: Methylation (CH3) Estimation Using Ionic current (https://github.com/comprna/CHEUI) for both m5C and m6A modification calls.[23,24] The concordant modification calls from the two tools were considered for further analysis. The analysis considered only the modified calls with a read_coverage of ≥20 and modified_probability >=0.9 to maintain data stringency. The gene that satisfied the above two criteria for modified bases was used for GO enrichment analysis by DAVID (https://david.ncifcrf.gov/) with a p-value cut-off (EASE score) <0.05. Transcripts showing methylation of a base with the above-mentioned

confidence were mapped to their position on the individual chromosome, and the modified base was mapped to genomic features using the annotation file of the reference genomes.

Transcriptome reconstruction and classification

The first step towards reconstructing the entire transcriptome from long reads is mapping them to the genome, for which Minimap2 was used.[22] This was followed by the FLAIR (Full-Length Alternative Isoform analysis of RNA) modules - correct, collapse, and quantify - to correct splice site misalignments, identify unique isoforms, and quantify each isoform, respectively.[25] These isoforms were then classified into structural categories such as FSM, ISM, NIC, NNC, etc., based on similarity to the existing annotations (Fig. S1) using SQANTI3.[26]

Differential methylation of isoforms

The splicing pattern of two alternative exons of a gene in a transcript is not always random. Pairs of exons may either be spliced together with a higher or lower frequency than by chance, leading to preferential mutual inclusion or exclusion, respectively. Such an inference is possible only from long-read RNA-sequencing data as it offers insights into patterns of inclusion of exons far apart in transcripts. To determine if some isoforms of a gene undergo methylation preferentially over others, the original Minimap2 + FLAIR workflow was performed twice - the original with all the sequencing reads or the 'total' workflow, and once with only those reads for which either type of methylation mark (m6A or 5mC) was identified, named 'meth m6A' and 'meth 5mC' respectively. Differential methylation of isoforms was judged by classifying isoforms of the gene into two categories - 'major' and 'others'. This classification was done using the 'total' workflow results where the isoform with the highest expression was labeled as the major. If two (or more) isoforms of a gene have the same and highest counts, anyone was randomly chosen. The minimum requirement to identify differential methylation of isoforms is the presence of two or more unique gene isoforms. Four sets of values are calculated for either methylation mark - counts of molecules of major isoform detected from methylated (a for m6A; A for 5mC) or total reads (c), and counts of all other isoforms from methylated (b for m6A; B for 5mC) or total reads (d) - which form the 2x2 contingency table for Fisher's test and LOR calculation (Figure S2). This was followed by multiple testing corrections of the obtained p-values using the Yekutieli-Benjamini method. A hypergeometric test was performed to test the significance of the overlap of genes displaying preferential methylation marks of both types.

RESULTS

Read statistics

Total RNA from two enriched, PCR-validated severe malaria samples was extracted as detailed (one each for *P. vivax* and *P. falciparum*). After a quality and integrity check, the RNA was used for library preparation for MinION sequencing. In the case of *P. vivax*, we obtained 672k reads, while for *P. falciparum*, we obtained 207k reads (Fig S3). Only the basecalled reads above Q>7 (as recommended by Guppy) were taken for further analysis. The longest mapped read length for PFC was ~5.2 kb, while that for PVC was ~3.5 kb. This

shows the utility of long-read sequencing technology, performed for the first time in malaria patient-derived RNA samples. A detailed summary of the read statistics is presented in Table S2.
Methylation calling for mapped reads revealed modifications in both sense and antisense transcripts (NATs). All the modified transcripts for m6A detected by m6anet were covered in the modification calls predicted by CHEUI. Here, we report both the tools' combined and concordant modification calls for m6A methylation. The total number of transcripts reported to be modified in both samples is summarized in Table S3.

Methylation modifications, m5C, and m6A have been reported in all RNA types, namely mRNA, rRNA, ncRNA, snoRNA, and pseudogenic transcripts. No common transcript with m5C and/or m6A modification was seen in sense and antisense transcript pairs in the PVC dataset. However, in the PFC dataset, common transcripts showing a particular modification in both its sense and antisense counterpart are shown in Table 1. Three of these genes have both the m5C and 6mA modifications in their sense-antisense transcript pairs, as depicted in Table 1.

An overlap of the common transcripts considering both the m5C and m6A events is depicted in Fig 1. Mapping of the genes showing RNA modifications to individual chromosomes for PFC (Supplementary sheet 2) and PVC (Figure S4 a-d) data has revealed no specific distribution of methylation modifications favoring any particular chromosome. Some m6A and m5C modifications also lie in the sense transcripts encoded by the mitochondrial genome in both PFC and PVC and by the apicoplast genome in PFC. These transcripts are listed in Table S4, along with their function.

Further mapping of methylated bases in both sense and antisense transcripts to the genomic features revealed the positions of the modified bases lying in 5’UTR, CDS, and 3’UTR annotated regions of these transcripts. The methylated bases mapped to the exons for the genes with the modification lying in the rRNA transcripts, possibly due to a lack of annotated UTRs for these transcripts. Methylation points in antisense transcripts are hypothetically correlated with the sense mRNA regions, considering antisense to be a reverse complement of sense mRNA. Most of the methylation events of mRNA transcripts are mapped to CDS, and very few are in 3’UTR or 5’UTR, as depicted in Figure S5(a-d).

The Venn diagram in (Fig 2 a-d) shows the total distribution of each of the methylated sense transcripts across 5’ UTR, CDS, and 3’ UTR regions of the mRNA. In the PFC dataset, 16 transcripts have m5C methylated bases lying in all the three - 5’UTR, CDS, and 3’UTR regions of the mRNA, whereas 40 transcripts have shown m6A methylated bases spread across all three annotated regions of the mRNA (Fig 2a-b). In the PVC dataset, nine of the m6A methylated transcripts have methylation marks in more than one of the annotated mRNA regions (Fig 2c), whereas 14 transcripts have shown the distribution of m5C methylation marks in more than one annotated region of mRNA (Figure 2d).

<u>Gene Ontology Enrichment Analysis</u>

The Gene Ontology Enrichment analysis for the methylated genes for m5C and m6A modifications mapped the biological processes corresponding to cellular protein metabolic process, translation, and organonitrogen compound biosynthetic process, detailed in Fig 3a-b. Apart from this, some individual GO terms are identified separately in PVC (Figure S6a-b) and PFC datasets (Figure S6c-d and Supplementary sheet 3). The Gene Ontology Enrichment analysis for the methylated genes for the m5C and m6A modifications in *P. vivax* was primarily mapped into the molecular functions and cellular components that are expected to have an impact on several metabolic and biosynthetic processes, impacting parasite development.

Further, in both the datasets, the enrichment of genes in the KEGG pathway revealed – 'Ribosome (pfa03010/pvx03010),' along with the genes associated with the biological process of translation. Among the significant KEGG pathways identified in the PFC (m6A) dataset are - pfa00010:Glycolysis/Gluconeogenesis and pfa01230:Biosynthesis of amino acids, associated with biological processes like – biosynthesis of amino acid, nucleotide phosphorylation, and glycolytic process, which may suggest activation of pathways involved in cellular energy production and nucleotide metabolism. This suggests that m6A modifications in *P. falciparum* might alter the metabolic adaptations in the parasite in response to it's host cell environment.

Apart from the house-keeping metabolic processes like glycolysis and ribosome biogenesis, the gene ontology enrichment analysis of methylated m6A transcripts in the PFC dataset has also identified highly specific biological processes like GO:0020027~hemoglobin metabolic process, GO:0006458~'de novo' protein folding, and GO:0051262~protein tetramerization. These processes, when coupled with cellular locations of GO:0020020~food vacuole or GO:0005775~vacuolar lumen, signify the process of methylation important for the parasite's food vacuole operations optimized for amino acid utilization and hemoglobin catabolism.

The gene enrichment analysis of methylated natural antisense transcripts of the PFC dataset identified specific biological processes like - GO:0051383~kinetochore organization and GO:0000278~mitotic cell cycle. The cellular location enriched for GO:0000775~chromosome suggests the involvement of the methylated antisense transcripts in cellular proliferation dynamics in the nucleus, regulating underlying cell division in *P. falciparum*. On the other hand, a similar investigation of the GO terms for both m6A and m5C methylated NATs enriched in PVC shows numerous functions indicative of modulation of nucleic acid synthesis and biomolecule binding, although, a specific process has not been identified. In this enrichment process, two of the important genes are PVX_098040 and PVX_081285 which encode asparagine—tRNA ligase and selenocysteine-specific elongation factor selB homologue (both putative), which is suggestive of events leading to protein modifications and functional changes.

<u>Orthologs in the two datasets</u>

We further identified orthologs of methylated transcripts across *P. falciparum* and *P. vivax* data, revealing conserved epigenetic patterns (Supplementary sheet 4). This cross-species analysis hints at the potential evolutionary significance of methylated genes in *Plasmodium*, which may impact translational repression or activation, influencing the pathogenesis of severe malaria.[27,28] One class of significant molecules exhibiting both m5C and m6A methylation are some members of the plasmepsin family. Another pair of orthologous genes PVX_090230.1 and PF3D7_0202500.1 encodes for early transcribed membrane protein 2 (ETRAMP2) and displays m6A modification on sense transcripts of both the datasets with a high confidence of 0.999.

Differential methylation of isoforms

As long reads ideally capture the entire transcript, reads with RNA modification must correspond to transcript molecules with that modification. Therefore, we tested the isoforms detected for specific methylation, i.e., some isoforms of a gene preferentially undergoing methylation (m6A and/or 5mC) versus others. Applying Fishers' test, of the genes with m6A and/or, 5mC RNA modifications detected in either species, those with two or more isoforms were tested for differential RNA methylation of isoforms. This yielded 24 and 19 genes in *P. vivax* and 205 and 182 genes in *P. falciparum* for m6A and 5mC, respectively. There was a significant overlap of genes with each time of modification in either species (Fig 4).

DISCUSSION

More than 170 chemical modifications of RNA are known to date.[29] m6A is the most abundant internal modification on mRNA, altering RNAs' splicing, translation, stability, and nuclear export. The m6A deposition levels in *P. falciparum* are relatively more than those of any other known eukaryotic species. They are strictly regulated by methyltransferase (MTC) writer complex (PfMT-A70) identified by the YTH reader homolog.[30,31] Over 99% of m6A sequencing reads in these cited studies matched to regions that code for proteins. This study also supports the previous findings for m6A since most modifications are present in the coding segments of the mRNA, followed by the 5' and 3' UTRs. In addition, we also find a similar phenomenon for m5C.

In the PFC dataset, the presence of both of these modifications on some key mRNA transcripts and their NATs (Table 1) reveals a dual methylation pattern that may influence the fate of these two transcripts, thereby affecting the parasite's survival in the host. Of these, the Signal recognition particle subunit SRP68 (PF3D7_0621900.1) is shown to function as a transcription factor.[32] This is consistent with the enrichment of the GO term 'gene expression' in the gene ontology of m6A methylated transcripts for PFC (Supplementary sheet 3). Another transcript, ETRAMP 11.2 (PF3D7_1102800.1) from the PFC dataset, shows predominately both adenine and cytosine methylations in the 3'UTR, suggesting enhanced instability of the mRNA in severe disease conditions. This may suggest that some of these molecules need to be further investigated for their use as serological markers, especially in cases of disease severity.

Transcripts of the ETRAMP family are orthologous in both datasets (Supplementary sheet 4) with high confidence of methylation (>=0.99) dispersed at distinct annotated regions of mRNA. Proteins of the ETRAMP family are known to be widely expressed in the early asexual stages of all *Plasmodium* species at the host-cell interface.[33,34,35] Upon early intra-erythrocytic development, the parasites need to export much of their proteins into the host cell to sustain their survival within the host. This molecule has an important role in transporting proteins outward from the parasite, and this role could be enhanced in cases of severe disease. Bulk protein breakdown, secretory protein maturation, egress, invasion, and endothelial adhesion are all impacted by the aspartyl proteases of the plasmepsin family.[36] Some transcripts encoding this family have shown methylation in the 3'UTR regions, whereas others have shown methylation in the CDS. This could hint at the key family members of plasmepsin proteins, which need to be enhanced in severe disease conditions for the parasite's survival.

In addition to base modifications in sense and antisense transcripts, we have detected the phenomenon of splicing of various transcripts in both the above categories. The PF3D7_0210100 gene was detected with a novel structural isoform with a new splicing site (Novel not in-catalog). Previously, this gene has been reported with 5' and 3' alternative splicing events.[37] We have also found at least 212 genes in the PFC dataset associated with novel isoforms with a new combination of known splice sites. Of these novel isoform categories, PF3D7_1358900 is found to exist with as many as 14 exons with a diverse combination of known splice sites. Despite *P. falciparum* being a well-annotated genome, we report new transcripts and splicing events not present in the existing annotations of the PF3D7 genome(v64). Integration of this isoform data with other proteomics/metabolomics data can contribute to a comprehensive understanding of molecular networks underlying malaria caused by *P. falciparum*.

P. vivax exhibits alternative splicing similar to *P. falciparum.*[37] The full-length transcriptome data from the PVC dataset has significantly expanded the isoform catalog (Table 2). It is noteworthy that 50 isoforms, which were previously unannotated in the catalog (Novel in-catalog) demonstrated similar quality to isoforms that were already known to be FSM. Furthermore, we determined the percentage of isoforms classified as antisense (n=215), genic (n=362), intergenic (n=897), and fusion (n=2; transcriptional read-through between two or more adjacent genes). For example, we found two fusion events PVX_002870; PVX_002867 & PVX_116525; PVX_116530, which may indicate a major role for splicing events in influencing proteins with potentially altered activity, which may have a significant role in parasite survival and adaptability under severe disease manifestations.

Investigating differentially methylated isoforms of individual genes in PVC showed 24 and 19 preferentially methylated transcripts, respectively, for m6A and m5C modifications (Figure 4). Of these, 13 transcripts showed a high degree of overlap with both methylation signatures being present; these genes include DMC1 and protein phosphatase inhibitor 2, suggesting a possible role in parasite adaptation. There were others that included conserved hypothetical proteins or were unannotated. Additionally, many genes mentioned above had

upregulated antisense transcripts (unpublished data from RNAseq studies from multiple samples for the same manifestation) that were not methylated. This is in contrast to the sense counterparts, which showed extensive differential methylation and could suggest a critical role for the antisense transcripts in regulating the translation of this methylated sense.

In both PVC & PFC datasets, the differentially methylated sense isoforms show methylation overlap for more than 40% of genes for both methylation events. Additionally, antisense transcripts[10] (unpublished) and full splice matches of transcript isoforms encoded by the mitochondrial genome are also present. Taken together, this could imply dynamic regulation of translation in the mitochondria. It is thought that the mitochondria of *Plasmodium* sp., while in the asexual stages within the human host, are comparatively in a minimally functional state.[38,39] It is intriguing that despite this fact, we have detected a multi-layered gene regulatory phenomenon in this organelle.

The methylation events in sense transcripts are primarily enriched in the coding regions, followed by 5'UTR and 3'UTR. However, in the case of antisense, the methylated transcripts in CDS, 5'UTR, and 3'UTR are sporadically dispersed. Additionally, we have seen that in *P. vivax*, there are 48 transcripts that contain both m6A and m5C in the sense transcripts, and seven transcripts contain both of these modifications amongst the NATs (Figure S7A). In contrast, in the PFC dataset, 415 sense transcripts show both m6A and m5C modifications and 16 transcripts show both these modifications in the NATs (Figure S7B).

It has been earlier reported[30] that m6A mRNA modification reduces the stability of the transcripts. Other studies have shown that modification at the 3'UTR regions enhances RNA degradation.[40] Our data shows the presence of significant amounts of m6A modifications predominately in the CDS of the mRNA transcripts (Figure S5a-d). This could suggest enhanced RNA stability, potentially increasing translational efficiency in case of severe disease.

This is the first report to comprehensively profile the m5C and m6A modification distribution in sense and natural antisense transcripts from the clinical transcriptome of *P. falciparum* and *P. vivax*. These modifications may potentially modify mRNA stability and translation. Further, we have identified methylated transcripts encoded from Plasmodium's mitochondria and apicoplast genomes for the first time. A comprehensive analysis of the quantification of splicing events in methylated sense and antisense transcripts is presented, along with the preferential methylation of the major spliced isoforms in the two datasets. This study reports several conserved hypothetical genes and unannotated transcripts from the genomes of the two *Plasmodium* species, opening future avenues for investigating and characterizing these novel molecules, and also stresses the need to sequence the parasites in their natural host environment to identify molecules expressed *in vivo* for individual disease phenotypes.

ACKNOWLEDGEMENTS

This work is funded by the Indian Council of Medical Research (ICMR) under the Extramural Ad-hoc Scheme (PID: 2019-1121). We thank Birla Institute of Technology (BITS), Pilani (Pilani campus), International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, Indian Institute of Technology (IIT), New Delhi, India and SP Medical College, Bikaner, for providing the necessary facilities required for this work.

Ethics Approval - Sample collection was earlier approved by SP Medical College Hospital's Ethics Committee (No.F. (Acad)SPMC/2003/2395). Permission to use these samples for further studies was given through IERB approval No. F29(Acad)SPMC/2020/3151 dated 05.09.2020. We thank all the patients for their voluntary consent and participation.

AUTHOR CONTRIBUTIONS

AD conceived the study, organized the material, and networked it to facilitate the outcome. SG and PR were involved in performing the sequencing experimentation and subsequent analysis. AD, SG, and PR wrote the manuscript. PG and IG analyzed the alternative splicing events from the sequenced reads. IG has been closely involved in discussions since the study's inception. TM supervised the application of various *in-silico* tools and assisted in troubleshooting when required. DG was involved in consultations on data analysis and related discussions. DKK and SKK supervised sample collection and clinical characterization. All authors read and approved the manuscript.

REFERENCES

1. World Health Organization. *World Malaria Report 2022*. (2023).
2. Kochar, D. K., Saxena, V., Singh, N., Kochar, S. K., Kumar, S. V., & Das, A. *Plasmodium vivax* Malaria, *Emerging Infectious Diseases*, 11-1 (2005). www.cdc.gov/eid
3. Kochar DK, Das A, Kochar SK, Saxena V, Sirohi P, Garg S, Kochar A, Khatri M, & Gupta V. Severe *Plasmodium vivax* Malaria: A Report on Serial Cases from Bikaner in Northwestern India. *American Journal of Tropical Medicine and Hygiene*. Vol. 80(2), pp. 194-198. (2009).
4. Matlani M, Kojom L, Mishra N, Dogra V, & Singh V. Severe vivax malaria trends in the last two years: a study from a tertiary care center, Delhi, India. *Annals of Clinical Microbiology and Antimicrobials*, 19:49. (2020).
5. Lacerda, M, Mourão M, Alexandre A, Siqueira A, Magalhães B, Martinez-Espinosa E, Santana S, Brasil P, Ventura M, Tada M, Couto V, Silva A, Silva R, & Alecrim M. Understanding the clinical spectrum of complicated Plasmodium vivax malaria: a systematic review on the contributions of the Brazilian literature. *Malaria Journal*, 11(1). (2012).
6. Siqueira A, Lacerda M, Magalhães B, Mourão M, Melo G, Alexandre M, Alecrim M, Kochar D, Kochar S, Kochar A, Nayak K, Portillo H, Guinovart C, Alonso P, & Bassat Q. Characterization of Plasmodium vivax-associated admissions to reference hospitals in Brazil and India. *BMC Medicine*, 13:57. (2015).

7. Chu, C. S., Stolbrink, M, Stolady D, Saito M, Beau C., Choun, K., Wah, T. G., Mu, N., Htoo, K., Nu, B., Keereevijit, A., Wiladpaingern, J., Carrara, V., Phyo, A. P., Lwin, K. M., Luxemburger, C., Proux, S., Charunwatthana, P., McGready, R., . . . Nosten, F. Severe Falciparum and Vivax Malaria on the Thailand-Myanmar Border: A Review of 1503 Cases. *Clinical Infectious Diseases*, 77(5), 721–728. (2023)
8. Health Alert Network (HAN) – 00494, Locally Acquired Malaria Cases Identified in the United States. *CDC Health Alert Network*, (2023).
9. Weiner, A., Dahan-Pasternak, N., Shimoni, E., Shinder, V., von Huth, P., Elbaum, M., & Dzikowski, R. 3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite *Plasmodium falciparum*. *Cellular Microbiology*, *13*(7), 967–977. (2011).
10. Boopathi, P. A., Subudhi, A. K., Garg, S., Middha, S., Acharya, J., Pakalapati, D., Saxena, V., Aiyaz, M., Chand, B., Mugasimangalam, R. C., Kochar, S. K., Sirohi, P., Kochar, D. K., & Das, A. Revealing natural antisense transcripts from Plasmodium vivax isolates: Evidence of genome regulation in complicated malaria. *Infection, Genetics, and Evolution*, *20*, 428–443. (2013).
11. López-Barragán, M. J., Lemieux, J., Quiñones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., & Su, X. zhuan. Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics*, *12*:587. (2011).
12. Thompson, T. A., Chahine, Z., & Le Roch, K. G. The role of long noncoding RNAs in malaria parasites. *Trends in Parasitology*. Vol. 39 (7), pp. 517–531. (2023).
13. Hollin, T., Chahine, Z., & Le Roch, K. G. Epigenetic Regulation and Chromatin Remodeling in Malaria Parasites. *Annual Review of Microbiology*, 77:255-76 (2023).
14. Catacalos, C., Krohannon, A., Somalraju, S., Meyer, K. D., Janga, S. C., & Chakrabarti, K. Epitranscriptomics in parasitic protists: Role of RNA chemical modifications in posttranscriptional gene regulation. In *PLoS Pathogens*, Vol. 18(12). (2022).
15. Lee V, Judd L, Jex A, Holt K, Tonkin C & Ralph S. Direct Nanopore Sequencing of mRNA Reveals Landscape of Transcript Isoforms in Apicomplexan Parasites. *mSystems* 6:e01081-20. (2021).
16. Sorber, K., Dimon, M., & DeRisi, L.J. RNA-Seq splicing analysis in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Research*, 39(9), 3820-3835. (2011).
17. Antinori, S., Galimberti, L., Milazzo, L., & Corbellino, M. Biology of Human Malaria Plasmodia Including *Plasmodium knowlesi*. *Mediterranean Journal of Hematology and Infectious Diseases*, Vol. 4. (2012).
18. Carlton, J. M., Adams, J. H., Silva, J. C., Bidwell, S. L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S. V., Merino, E. F., Amedeo, P., Cheng, Q., Coulson, R. M. R., Crabb, B. S., Del Portillo, H. A., Essien, K., Feldblyum, T. V., Fernandez-Becerra, C., Gilson, P. R., Gueye, A. H., Fraser-Liggett, C. M. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, *455*(7214), 757–763. (2008).
19. Pakalapati, D., Garg, S., Middha, S., Acharya, J., Subudhi, A. K., Boopathi, A. P., Saxena, V., Kochar, S. K., Kochar, D. K., & Das, A. Development and evaluation of a

28S rRNA gene-based nested PCR assay for *P. falciparum* and *P. vivax*. *Pathogens and Global Health*, *107*(4), 180–188. (2013).

20. Pakalapati, D., Garg, S., Middha, S., Kochar, A., Subudhi, A. K., Arunachalam, B. P., Kumar Kochar, S., Saxena, V., Rp, P., Acharya, J., Kumar Kochar, D., & Das, A. Comparative evaluation of microscopy, OptiMAL禪and 18S rRNA gene-based multiplex PCR for detection of *Plasmodium falciparum* & *Plasmodium vivax* from field isolates of Bikaner, India. *Asian Pacific Journal of Tropical Medicine*, 346-351. (2013).
21. Guidelines for the treatment of malaria – 3rd Edition. *World Health Organization* (2015).
22. Li, Heng. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics*, 34 (18): 3094–3100
23. Acera Mateos, P., Sethi, A. J., Ravindran, A., Guarnacci, M., Srivastava, A., Xu, J., Woodward, K., Yuen, Z. W. S., Mahmud, S., Kanchi, M., Zhou, Y., Sneddon, A., Hamilton, W., Gao, J., Starrs, L. M., Hayashi, R., Wickramasinghe, V., Zarnack, K., Preiss, T., Eyras, E. Simultaneous identification of m6A and m5C reveals coordinated RNA modification at single-molecule resolution. Preprint at https://doi.org/10.1101/2022.03.14.484124 (2023).
24. Hendra, C., Pratanwanich, P. N., Wan, Y. K., Goh, W. S. S., Thiery, A., & Göke, J. Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nature Methods*, *19*(12), 1590–1598. (2022).
25. Tang, Alison D., Cameron M. Soulette, Marijke J. van Baren, Kevyn Hart, Eva Hrabeta-Robinson, Catherine J. Wu, and Angela N. Brooks. 2020. "Full-Length Transcript Characterization of SF3B1 Mutation in Chronic Lymphocytic Leukemia Reveals Downregulation of Retained Introns." *Nature Communications* 11 (1): 1438.
26. Tardaguila, Manuel, Lorena de la Fuente, Cristina Marti, Cécile Pereira, Francisco Jose Pardo-Palacios, Hector Del Risco, Marc Ferrell, et al. 2018. "SQANTI: Extensive Characterization of Long-Read Transcript Sequences for Quality Control in Full-Length Transcriptome Identification and Quantification." *Genome Research* 28 (3): 396–411.
27. Bennink S., & Pradel G. The molecular machinery of translational control in malaria parasites. *Molecular Microbiology*. Vol. 112(6), 1658-73. (2019).
28. Higgins S., Xing K., Kim H., Kain D., Wang F., Dhabangi A., Musoke C., Cserti G., Tracey K., Kain K., Liles W. Systemic release of high mobility group box 1(HMGB1) protein is associated with severe and fatal *Plasmodium falciparum* malaria. *Malaria Journal*. 12:105. (2013).
29. Schaefer, M., Kapoor, U., & Jantsch, M. F. Understanding RNA modifications: The promises and technological bottlenecks of the 'epitranscriptome'. *Open Biology* 7: 170077. (2017).
30. Baumgarten, S., Bryant, J. M., Sinha, A., Reyser, T., Preiser, P. R., Dedon, P. C., & Scherf, A. Transcriptome-wide dynamics of extensive m6A mRNA methylation during Plasmodium falciparum blood-stage development. *Nature Microbiology*, *4*(12), 2246–2259. (2019).
31. Govindaraju, G., Kadumuri, R. V., Sethumadhavan, D. V., Jabeena, C. A., Chavali, S., & Rajavelu, A. N6-Adenosine methylation on mRNA is recognized by YTH2 domain

protein of human malaria parasite *Plasmodium falciparum*. *Epigenetics and Chromatin*, *13*(1). (2020).

32. Li J, Zhou F, Zhan D, Gao Q, Cui N, Li J, Iakhiaeva E, Zwieb C, Lin B, Wong J. A Novel Histone H4 Arginine 3 Methylation-sensitive HistoneH4 Binding Activity and Transcriptional Regulatory Function for Signal Recognition Particle Subunits SRP68 and SRP72. *The Journal of Biological Chemistry*. Vol. 287(48), pp. 40641-40651. (2012).
33. Spielmann T, David Fergusen J, Beck H. etramps, a New *Plasmodium falciparum* Gene Family Coding for Developmentally Regulated and Highly Charged Membrane Proteins Located at the Parasite–Host Cell Interface. *Molecular Biology of the Cell*. Vol. 14, 1529-1544. (2003). https://10.1091/mbc.E02– 04 – 02
34. Sindhe, M. K., Ray, A., Kumar, S., & Surolia, N. Interactions of Plasmodium falciparum ETRAMP14.1 with PfEMP1, translocon components and other ETRAMP members at the interface of host-parasite, Preprint at https://doi.org/10.1101/775650. (2019).
35. Lee, K. S., Han, J. H., Park, J., Ha, K., Park, W., Hong S., Na, S., Cheng, Y., & Han, E. Evaluation of antibody responses to the early transcribed membrane protein family in *Plasmodium vivax*. *Parasites & Vectors*, 12:594. (2019).
36. Nasamu, A. S., Polino, A. J., Istvan, E. S., & Goldberg, D. E. Malaria parasite plasmepsins: More than just plain old degradative pepsins. *The Journal of Biological Chemistry*, 295(25), 8425–8441. (2020).
37. Otto D. T., Willinksi, D., Assefa, S., Keane, T., Sarry, L. R., Bohme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C. & Llinas, M. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Molecular Microbiology*, 76(1), 12-24. (2010).
38. Vaidya, B. A. & Mather, W. M. Mitochondrial Evolution and Functions in Malaria Parasites. *Annual Review of Microbiology*, 63:249-67. (2009).
39. Krungkrai R. S., Leangaramgul, P., Kudan, S., Prapunwattana, P. & Krungkrai J. Mitochondrial Heterogeneity in Human Malarial Parasite Plasmodium falciparum, *ScienceAsia*, 25:77-83. (1999).
40. Gagliardi, D., & Dziembowski, A. 5′ and 3′ modifications controlling RNA degradation: from safeguards to executioners. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1762), 20180160. (2018).

FIGURES

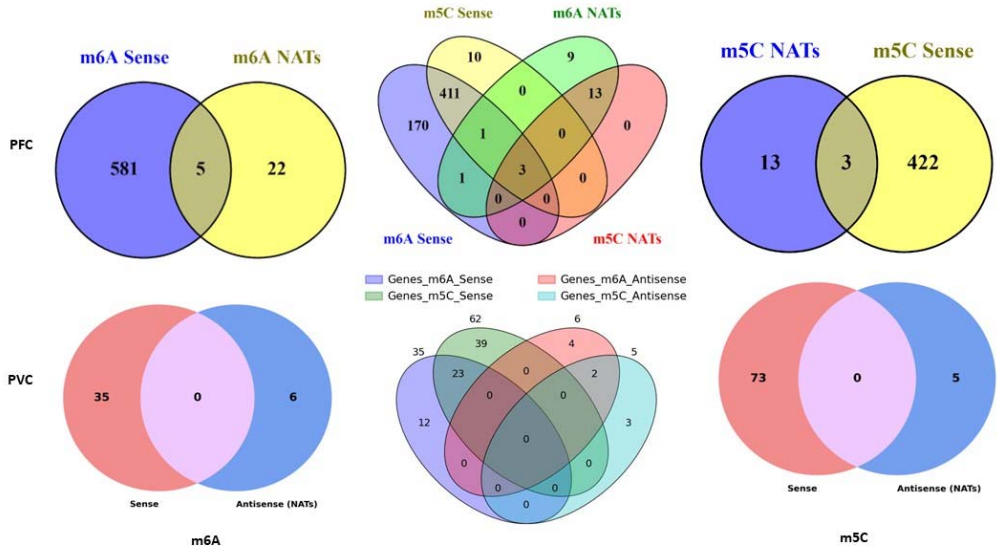
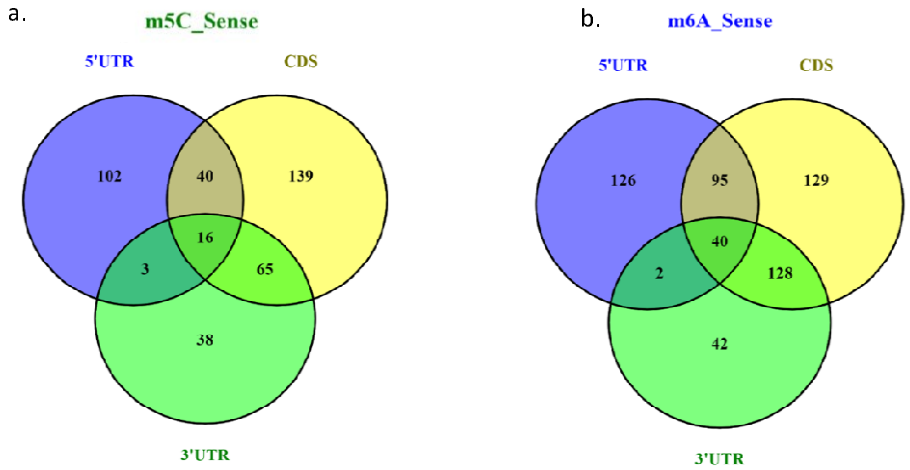
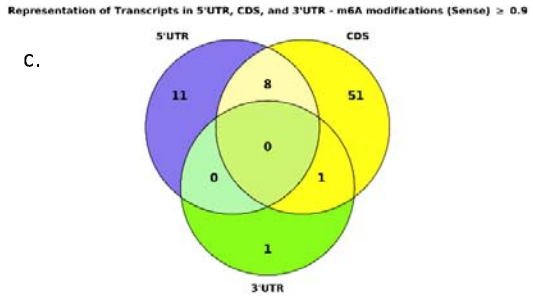
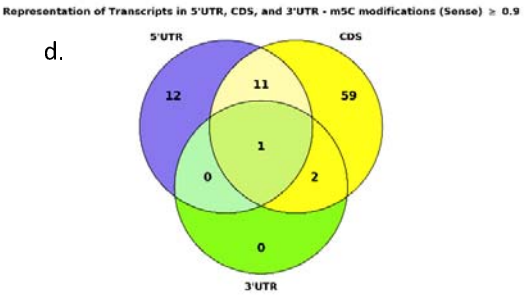
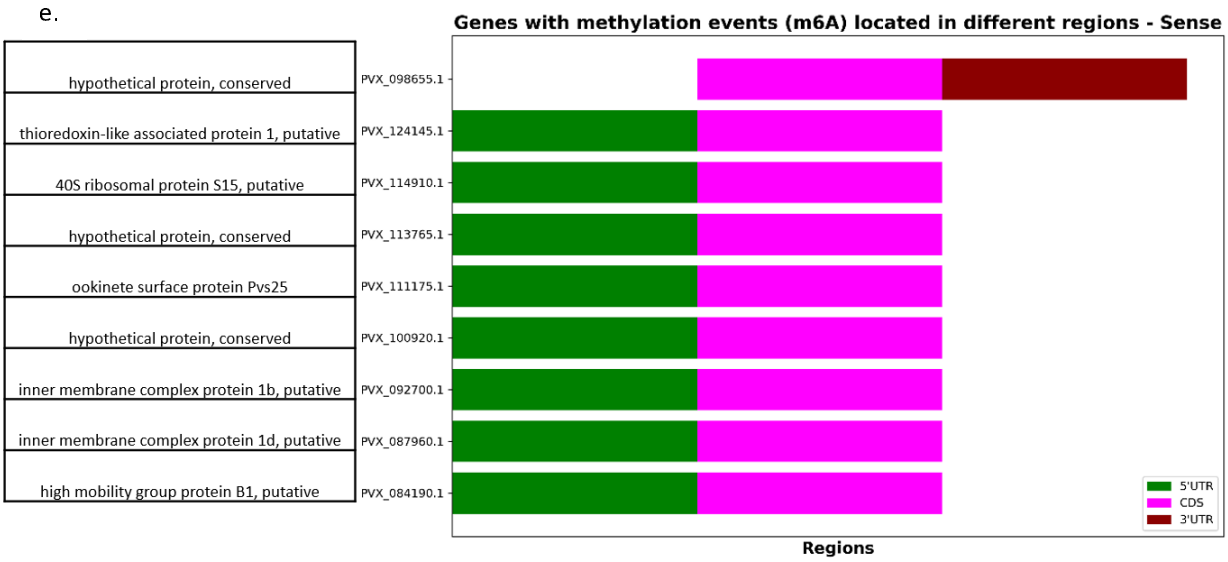


Fig 1 Top Panel: Representation of genes in PFC data (from left to right) – m6A modification in sense and antisense, both m5C and m6A modification in sense and antisense transcripts, m5C modification in sense and antisense transcripts.
Bottom panel: Representation of genes in PVC data (from left to right) - m6A modification in sense and antisense, both m5C and m6A modification in sense and antisense transcripts, m5C modification in sense and antisense transcripts.









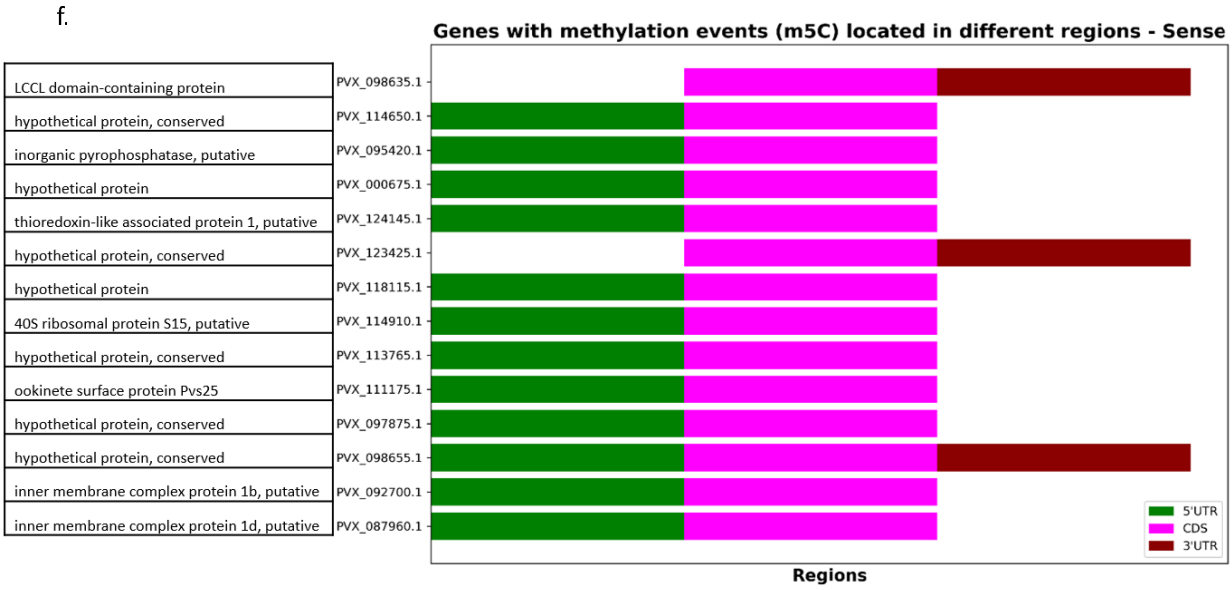
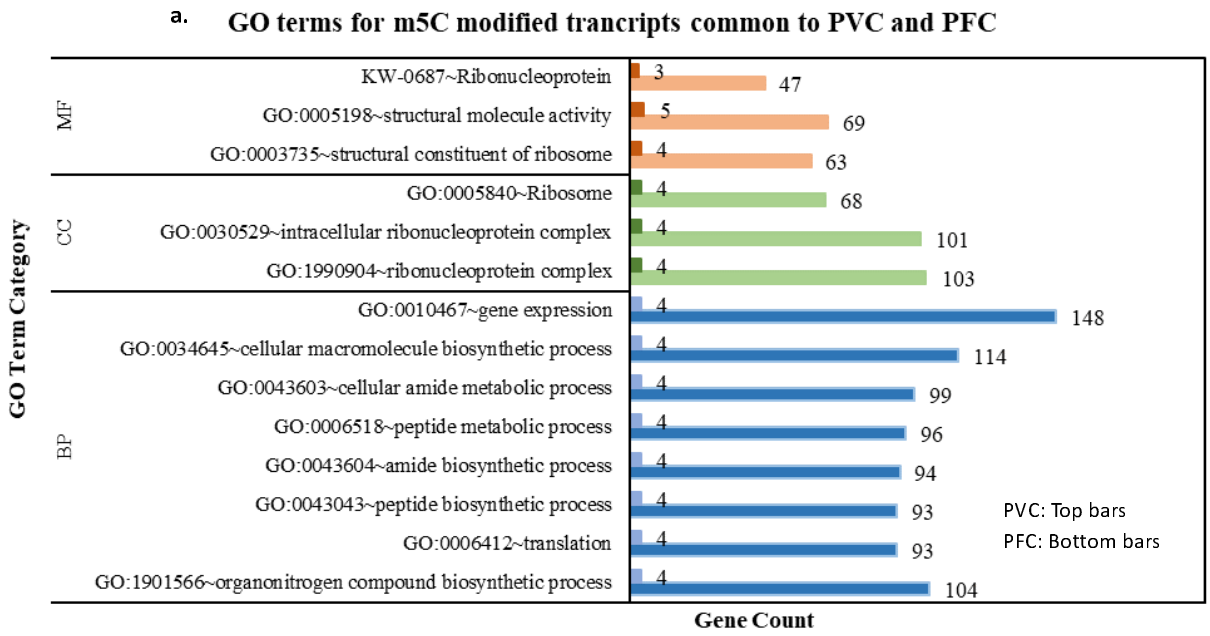


Fig 2 a -Number of transcripts with methylation events (m5C) located in different annotated features of mRNA in PFC. Transcript IDs are further detailed in Supplementary Sheet 2.
b. Number of transcripts with methylation events (m6A) located in different annotated features of mRNA in PFC. Transcript IDs are further detailed in Supplementary Sheet 2.
c - Number of transcripts with methylation events (m6A) located in different annotated features of mRNA in PVC.
d. - Number of transcripts with methylation events (m5C) located in different annotated features of mRNA in PVC.
e. Modified transcripts (m6A) with overlapping annotated regions are mapped.



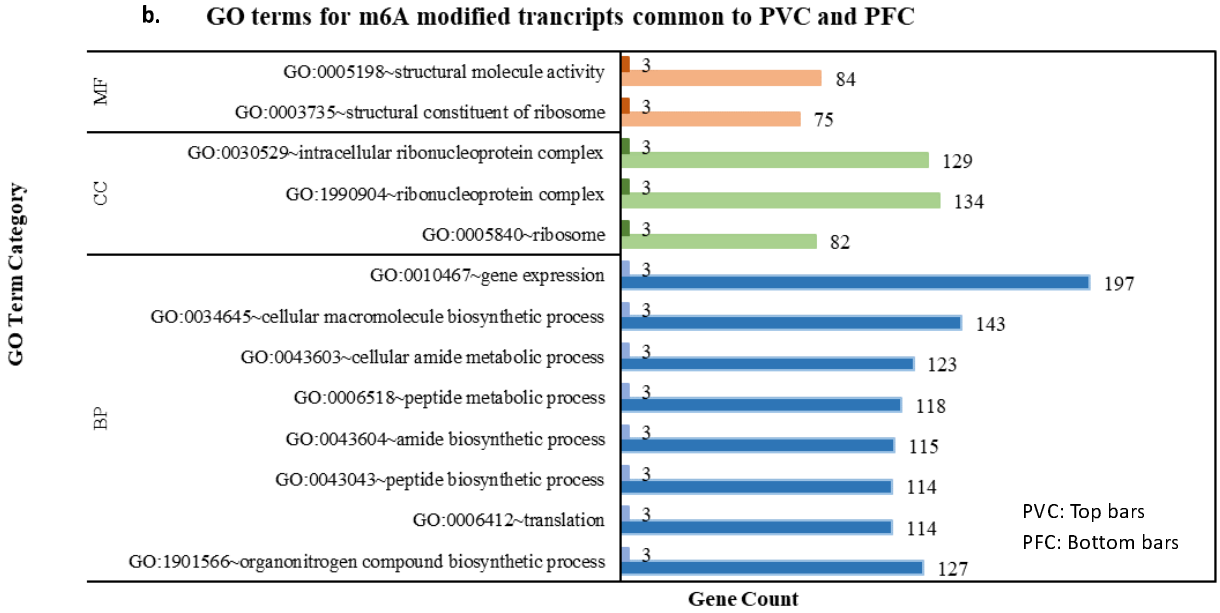


Fig 3a Common GO terms m5C modified transcripts in both PFC and PVC datasets.
Fig3b Common GO terms m6A modified transcripts in both PFC and PVC datasets
BP – Biological Process; CC – Cellular Component; MF – Molecular Function

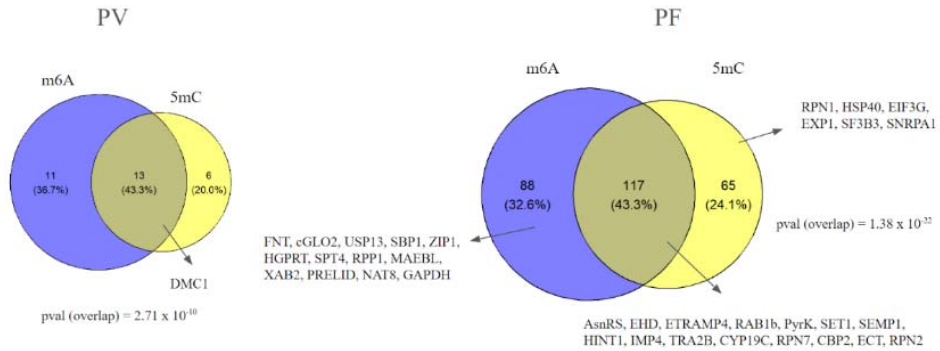


Figure 4: Overlap of m6A and 5mC methylation signatures in the two *Plasmodium* species datasets. Of the genes displaying differential methylation of isoforms, the genes for which gene names are available (others either have only gene IDs or are unannotated) have been marked here.

TABLES

Type of Modification	Transcript ID	Gene Description	Region of the methylated base in sense RNA	Region of the methylated base in NAT*
m5C	PF3D7_0904000.1	GTPase-activating protein, putative	5'UTR	3'UTR
	PF3D7_1102800.1	early transcribed membrane protein 11.2	3'UTR	3'UTR
	PF3D7_1329600.1	conserved Plasmodium protein, unknown function	CDS	3'UTR
m6A	PF3D7_0621900.1	signal recognition particle subunit SRP68, putative	5'UTR / CDS	3'UTR
	PF3D7_0904000.1	GTPase-activating protein, putative	5'UTR	3'UTR
	PF3D7_1102800.1	early transcribed membrane protein 11.2	3'UTR	3'UTR
	PF3D7_1329600.1	conserved Plasmodium protein, unknown function	5'UTR / CDS	3'UTR
	PF3D7_1408700.1	conserved protein, unknown function	5'UTR / CDS	3'UTR

Table 1 Gene description with m6A and m5C modification in both sense and natural antisense transcripts in the PFC dataset. Methylated base may occur more than once in the given annotated region of the transcript.
***Hypothetically oriented to sense RNA**

Structural category of isoforms	Number of isoforms	
	Plasmodium vivax	***Plasmodium falciparum***
Full-splice match (FSM)	1269	1981

Incomplete splice match (ISM)	335	443
Genic	362	279
Intergenic	897	162
Antisense	215	154
Novel in-catalog	50	231
Novel not-in-catalog	0	1
Fusion	2	2
Genic intron	0	1
Total	**3128 (2 missing?)**	**3258**

Table 2: Summary of SQANTI3 output