PNAS
www.pnas.org

Supplementary Information for

Paste manuscript title here

5-methylcytosine modification by *Plasmodium* NSUN2 stabilizes mRNA and mediates the development of gametocytes

Paste the full author list here

Meng Liu, Gangqiang Guo, Pengge Qian, Jianbing Mu, Binbin Lu, Xiaoqin He, Yanting Fan, Xiaomin Shang, Guang Yang, Shijun Shen, Wenju Liu, Liping Wang, Liang Gu , Quankai Mu, Xinyu Yu, Yuemeng Zhao, Richard Culleton, Jun Cao, Lubin Jiang, Thomas E. Wellems, Jing Yuan, Cizhong Jiang, Qingfeng Zhang

Paste corresponding author name here

Qingfeng Zhang, Thomas E. Wellems, Cizhong Jiang, or Jing Yuan
Email: qfzhang@tongji.edu.cn, twellems@niaid.nih.gov, czjiang@tongji.edu.cn, or yuanjing@xmu.edu.cn

This PDF file includes:

Supplementary Text with Detailed Methods
Figures S1 to S8
Captions for Datasets S1 to S8
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S8

Detailed Methods

Dot blot detection of m^5C Levels. Dot blot detection was performed as described previously (1). Biotin-labeled RNA oligonucleotides synthesized with or without m^5C were used as positive and negative controls. mRNA was isolated from preparations of total RNA by the Dynabeads® mRNA Purification Kit (Thermo Fisher Scientific). mRNA was denatured at 95 °C in a heat block for three min and chilled on ice immediately to prevent the re-formation of secondary structures of mRNA. A sample containing 0.5µg mRNA was applied directly onto the Hybond-N+ membrane (Millipore) and secured to the membrane by two treatment rounds in a HL-2000 Hybrilinker™ Hybridization Oven/UV Crosslinker (Analytik Jena US) operating in Autocrosslink mode (setting of 1,200 microjoules [x100]; 25-50 sec). After UV crosslinking, the membrane was incubated in 10 ml of 5% of non-fat milk in PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4, pH=7.4 with 0.05% Tween-20) for 1 h at room temperature with gentle shaking. The primary mouse anti-m^5C antibody (1:1000; Abcam, ab10805) was then added in 10 ml of antibody dilution buffer (non-fat milk in PBST) and incubated overnight at 4 °C. After three times washing (5 min each) in 10 ml of PBST, the membrane was finally incubated with HRP-conjugated anti-mouse IgG (1:5000; GE Healthcare) overnight at 4 °C. Signals from m^5C were developed with enhanced chemiluminescence (ECL) western blotting kit (GE Healthcare).

mRNA modification assessments by LC-MS/MS. Total RNA was extracted from synchronized parasites samples using Trizol (ThermoFisher Scientific) followed by treatment with DNase I (ThermoFisher Scientific). mRNA was isolated from total RNA samples using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Purified mRNA was quantified using the Qubit RNA HS Assay kit (ThermoFisher Scientific). mRNA was hydrolyzed to single nucleotides and the pretreated nucleoside solution was deproteinized using a Sartorius 10,000-Da MWCO spin filter. Analysis of nucleoside mixtures was performed on an Agilent 6460 QQQ mass spectrometer with an Agilent 1260 HPLC system. Multi reaction monitoring (MRM) mode provided high selectivity and sensitivity from parent-to-product ion transitions. LC-MS data was acquired using Agilent Qualitative Analysis software. MRM peaks of each modified nucleoside were extracted and normalized to the amount of purified mRNA. Detailed analysis of mRNA modifications was performed as previously described (2, 3).

Transcriptome sequencing. Total RNA was isolated from parasite preparations by the Direct-zol RNA Kit (Zymo Research). The poly(A) RNA fraction was selected using KAPA mRNA Capture Beads (KAPA), and fragmented to about 300–400 nucleotides (nt) in length. A KAPA Stranded mRNA-Seq Kit (KAPA) was used for RNA-seq library preparation for strand-specific RNA-seq. The libraries were sequenced on an Illumina NovaSeq 6000 platform to generate 150 bp pair-end reads.

RNA sequence processing and analysis. Low-quality and adaptor sequences were trimmed using cutadapt (v1.18) (4) with parameters: -a AGATCGGAAGAGC -AAGATCGGAAGAGC –trim-n -m 50 -q 20,20. RNA sequencing reads were aligned using hisat2 (v2.1.0) (5) with strand specific mode (--rna-strandedness RF) to the *P. falciparum* 3D7 genome (Pf 3D7 v32, obtained from PlasmoDB) or to the *P. yoelii* 17X genome (Py17X v45, obtained from PlasmoDB). Mapped reads were subsequently assembled into transcripts guided by the PlasmoDB gff annotation files (Pf3D7 v32/ Py17X v45) using featureCounts (v1.6.1) (6) with parameters: -M -p -B -C -s 2. Read counts were obtained using featureCounts (v1.6.1). FPKM for genes were calculated according to the formula $\mathrm{FPKM} = \frac{\mathrm{readcount}*10^3}{\mathrm{length}*\mathrm{totalcount}*10^{-6}}$.

RNA-Seq for mRNA half-life. ICR mouse blood with 4–6% gametocytemia was collected in anticoagulant tubes with heparin and immediately added to gametocyte culture medium (RPMI 1640 medium) in a blood/medium volume ratio of 1:10. Samples for mRNA half-life analysis were collected at timepoints of 0, 1, 3, and 5 h after the addition of actinomycin D (Sigma-Aldrich, 20 μg/ml) to inhibit the synthesis of new mRNAs as described (7, 8). Total RNA was extracted by Direct-zol RNA Kit (Zymo Research) as per the manufacturer's instructions. Prior to construction of the library with KAPA Stranded mRNA-Seq Kit Illumina platform (KK8421), ERCC RNA Spike-In Control Mixes (Ambion) was added proportionally to the total RNA of each sample (0.1 μL per sample) (9). Experimental results were obtained from two biological replicates. Raw mRNA sequence data from each sample were filtered to remove low-quality reads and adapter sequences were excised with cutadapt (v1.18) (4). Reads with average quality score ≥ 20 and length ≥ 50 bp were retained. Clean reads were mapped to the *P. yoelii* 17X genome (Py17X v45, obtained from PlasmoDB) with ERCC spike in sequence using hisat2 (v2.1.0) with strand specific mode (--rna-strandness RF). Uniquely mapped reads were used to calculate read count for each transcript, using software featureCounts. To remove low abundance transcripts, only transcripts with at least five read counts at start-point 0 h were kept for downstream

half-life analysis. The read count of each transcript in a time point was normalized against the sum of reads mapping to all ERCC spike-in transcripts. Half-lives of transcripts were calculated using R package MINPACK (10). Transcripts were assumed to exponentially decay according to the equation: $C = C_0 * e^{-kt}$, where C represents transcript abundance as a function of time, C_0 is the transcript abundance at 0 h, and k is the first order rate constant. Half-life values were calculated by the equation: $t_{½} = \ln(2)/k$.

***In vitro* transcription of mouse *dhfr* mRNA for control use in RNA-BisSeq experiments.** Institute of Cancer Research (ICR) mouse heart tissue (0.05g) was obtained and resuspended in 1 mL TRIzol (Invitrogen) for RNA extraction. The full-length *Mus musculus dhfr* CDS was amplified from cDNA and cloned between the Xba I and Xho I sites of the pBSKS vector (Stratagene, La Jolla, CA) containing a T7 promoter (forward primer: 5'-CACCGCGGTGGCGGCCGCTCTAGAATGGTTC GACCATTGAACTGC-3'; reverse primer: 5'-GTACCGGGCCCCCCCTCGAGTTAGTCTTTCTTCTCGTAGACTTC-3'). The pBSKS-dhfr vector was linearized by FastDigest Xho I (Thermo Scientific™). Five hundred ng of linearized pBSKS-dhfr vector was used as a DNA template for *in vitro* transcription with T7 RNA Polymerase (New England Biolabs) at 37 °C for 4 h in a 20 µl reaction mixture, according to the manufacturer's instructions. The DNA template was removed from the product mRNA by DNase I (Invitrogen™) digestion for 15 min at room temperature.

RNA-BisSeq data collection from transcripts and analysis. Parasites were isolated by treatment of 200–300 µl of loosely packed red blood cells from blood or culture at 3% parasitemia with 0.15% saponin in 1 × PBS on ice for 10 min. The parasite pellets were washed twice with precooling 1 × PBS and then resuspended in 1 mL TRIzol (Invitrogen). Total RNA was extracted using the Direct-zol RNA Kit (Zymo Research) according to the manufacturer's instructions. mRNA enrichment from the total RNA was performed using the Dynabeads™ mRNA Purification Kit (Invitrogen) according to the manufacturer's instructions.

RNA bisulfite treatment and purification of converted RNA were performed using the Methylamp™ RNA Bisulfite Conversion Kit (Epigentek Group Inc.) according to the manufacturer's instructions. One µg of mRNAs along with 5 ng of dihydrofolate reductase (*dhfr*) control RNA (1:2000) were used as an input RNA and unmethylated control, respectively. The resulting RNA was subsequently used for fragmentation and library construction using the

KAPA Stranded mRNA-Seq Kit (KAPA) according to the instructions provided by the manufacturer. Sequencing was performed on an Illumina NovaSeq 6000 instrument with paired end 150-bp read length.

Raw sequencing data were filtered to remove low-quality reads and adapter contaminations with cutadapt (4) as detailed in Methods. Reads with average quality score ≥ 20 and length ≥ 50bp were kept. Clean reads were mapped to the *P. falciparum* 3D7 genome (Pf 3D7 v32, obtained from PlasmoDB) or to the *P. yoelii* 17X genome (Py17X v45, obtained from PlasmoDB) using meRanGh, and only unambiguously aligned reads were used to call m^5C sites using the meRanCall tool from the meRanTK toolkit (11) (FDR < 0.01). Conversion rates were calculated as previously described (12) using the mouse DHFR mRNA as the methylation conversion control for quality assessments (13). Analysis of the *dhfr* spike-in revealed C to T conversions were approximately 98% (Dataset S1). mRNA-BisSeq data collection and analysis were performed in two replicate experiments. At least 20 reads were obtained for the candidate cytosine positions, and these were recorded if methylated cytosine depth was greater than 5 and m^5C level was greater than 0.1. Detection of an m^5C site in both replicates was required for its inclusion in further analysis. Mean m^5C levels for each transcript were calculated from the m^5C levels of all sites detected in the transcript. The m^5C sites were mapped to the regions of introns, exons, 5'-UTRs, and 3'-UTRs as previously described (12). The meRanCall tool from meRanTK was used to localize m^5C sites preferences with an input parameter of -sc 10. Logo plots were generated with R package ggseqlogo (14). PyNSUN2 targeted genes were defined as transcripts with m^5C modifications in WT schizont parasites that were lost in *Pynsun2*-KO parasites.

Plasmid construction and selection of transfected parasites. To disrupt the *PfNSUN2* opening reading frame (ORF) CDS in WT 3D7 and NF54 strains, we used CRISPR/Cas9 to generate knock-out vector constructs. Briefly, the pL6 plasmid (15) was modified to contain a complementary sequence of sgRNA between the Xho I and Avr II sites. For each knock-out assay, at least two sgRNAs were individually incorporated into plasmids to target the CDS of their target gene (Dataset S6). Sequences (~ 1kb) homologous to the target genes were designed with the introduction of several base pair deletions and were inserted into the pL6 plasmid between Asc I and Afl II sites by the In-Fusion PCR Cloning System (primer pairs are listed in Dataset S6). Recombinant plasmids were transfected simultaneously with pUF-Cas9 vectors into ring-stage parasites of 3D7-G7 and NF54-F5 by electroporation as described previously (16). Drug-resistant parasites were obtained after three or four weeks of continuous culture under 2.5 nM WR99210

pressure (2). Selected knock-out or knock-in parasites were confirmed by PCR detection and cloned by limiting dilution.

P. yoelii genetic modifications were performed using the CRISPR/Cas9 plasmid pYCm as previously described (17). To obtain gene knock-outs, 5' and 3'-genomic fragments (400 to 700 bp) of target genes were amplified using the corresponding primers (Dataset S6). The PCR products were restriction-digested and cloned into matched sites of the pYCm vector. The sgRNA oligonucleotides were annealed and inserted into the pYCm vector. At least two sgRNAs were designed to disrupt the CDS of a target gene for each deletion modification using the online program EuPaGDT (18). For PyNSUN2 tagging, a 400 to 800 bp fragment from C-terminal of the CDS and a 400 to 800 bp segment from the 3'-UTR were amplified and fused with a DNA sequence 6HA in frame at the C-terminal of the gene. There were at least two sgRNAs designed to target sites close to the C-terminal of the gene CDS region for incorporation of the insert. Five μg of circular recombinant plasmid DNA were electroporated into infected RBCs using a Lonza Nucleofector, and the transfected parasites were intravenously injected into a naïve ICR mouse. Infected mice were treated with pyrimethamine (6 μg/ml) in drinking water; drug-resistant parasites were obtained following 5 to 7 days of drug selection.

Proteomics analysis. TMT-based quantitative proteomics analysis was performed by Novogene Bioinformatics Technology Co Ltd. (Beijing, China), as described previously (19). Briefly, gametocytes of *P. yoelii* were collected and ground individually with liquid nitrogen and lysed in protein lysis buffer (100 mM NH_4HCO_3, 8 M Urea and 0.2% SDS, pH = 8.0) followed by ultrasonication on ice for 5 min. Desalted peptides were labeled with TMT 6-plex reagents (TMT6plex™ Isobaric Label Reagent Set, Thermo Fisher) according to the manufacturer's instructions. Labeled peptides of all samples were mixed equally and desalted by peptide desalting spin columns (Thermo Fisher, 89,852). A C18 column (Waters BEH C18, 4.6 × 250 mm, 5 μm) on a Rigol L3000 HPLC was used to fractionate for TMT-labeled peptide mix. An EASY-nLC™ 1200 UHPLC system (Thermo Fisher) coupled with an Orbitrap Q Exactive™ HF-X mass spectrometer (Thermo Fisher) operated in the data-dependent acquisition (DDA) mode was used for shotgun proteomics analyses. The resulting spectra from each fraction were searched separately against *P. yoelii* 17X proteome (Py17X v45, obtained from PlasmoDB) by the search engines: Proteome Discoverer 2.2 (PD 2.2, thermo). At least one unique peptide with false discovery rates (FDR) < 1.0 % on the peptide and protein level,

respectively, was considered as a confident peptide spectrum match for further analysis. The intensities of the TMT reporter ions against those ones provided in the databases were compared for protein quantification.

Western blotting. Protein fractionation was performed with modifications according to Till *et al* (20) and Baumgarten *et al* (2). Synchronous middle- to late-stage parasites were released from 1 mL infected red blood cells (3% parasitemia) by 0.15% saponin lysis and washed twice in 1× pre-cooled PBS. The parasite pellet was first resuspended in 0.5 mL cytoplasmic lysis buffer (20mM Hepes pH7.9, 10mM KCl, 1.5mM MgCl2, 1mM EDTA, 1mM EGTA, 0.65% NP-40, 1mM DTT with 1×protease inhibitor cocktail (Thermo Scientific)) and incubated for 30 min on ice. Parasites were further homogenized for 100 strokes in a glass dounce homogenizer, and the supernatant containing cytoplasmic proteins was isolated via centrifugation (10,000 rpm, 10 min, 4 °C). Protein samples were separated on SDS-PAGE and transferred onto PVDF membrane and visualized using ECL Western Blotting Detection Kit (GE Healthcare). The primary antibodies used in this study were mouse anti-HA (1:2000, Roche) and anti-GAPDH (1:1000, Servicebio, cat# GB12002), and secondary antibody was HRP-conjugated goat anti-mouse IgG (1:5000, Abcam, cat# ab6789) .

Immunofluorescence assays. Immunofluorescence assays (IFAs) were performed on synchronous parasites fixed with 4% paraformaldehyde in 1× PBS as previously described (17, 21). PyNSUN2-6HA and PfNSUN2-3HA-3Ty1 proteins were detected with primary antibody mouse anti-HA (1:200; Santa Cruz) and mouse anti-Ty1 (1:300; Sigma), respectively. The secondary antibody, Alexa-Fluor-488-conjugated goat anti-Mouse IgG (H+L) (ThermoFisher Scientific) was diluted by 1:2000.

Statistical analyses. We adopted the non-parametric Wilcoxon test for statistical comparisons of m^5C levels, mRNA half-lives, transcript abundances and protein abundances after the datasets were subjected to the Kolmogorov Smirnov test (p-values <0.05) and found to have non-normal distributions. Non-parametric Wilcoxon testing was also used for gametocytemia and ookinete conversion rate comparisons after departures from normality (p-values <0.05) were demonstrated by Shapiro Wilk analysis.

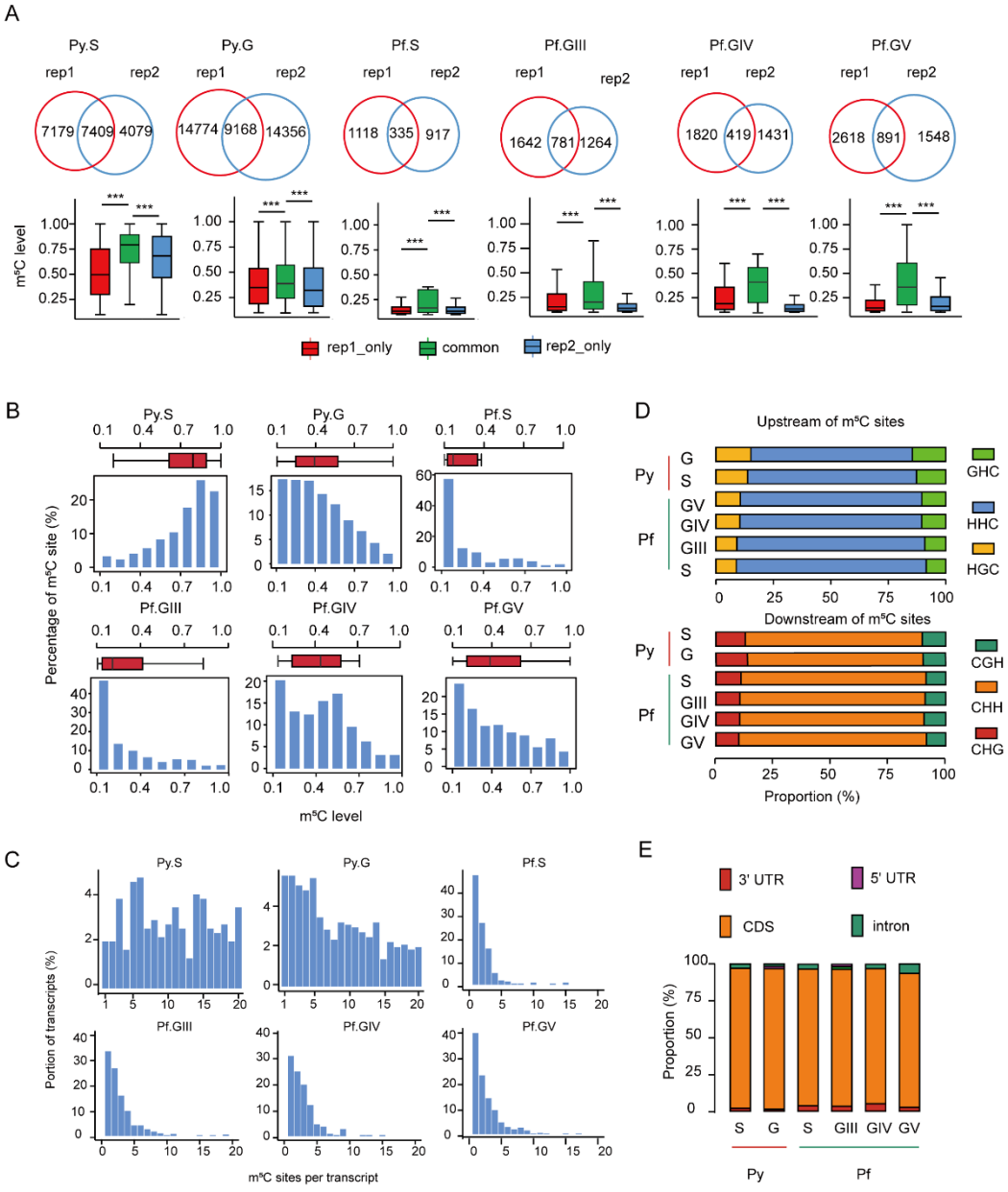


Fig. S1. Detection and transcriptome-wide mapping of m^{5}C mRNA modifications in *Plasmodium*. A, Venn diagrams display the numbers of mRNA m^{5}C sites that are shared or unique in two independent replicate determinations (rep1, rep2) from schizont and gametocyte stages of *P. yoelii* (Py.S, Py.G) and *P. falciparum* (Pf.S, Pf.GIII, Pf.GIV, Pf.GV). Box plots indicate m^{5}C levels at the shared and unique sites in each replicate pair. (***: p-value < 0.001, Wilcoxon test). Maximum, minimum, and median values are indicated in the box plots. **B,** Histograms and box plots indicate the distributions of modification levels at m^{5}C sites in the schizont and gametocyte stages of *P. yoelii* and *P. falciparum*. **C,** Frequency histograms indicate the proportion (%) of transcripts carrying 1–20 m^{5}C sites in schizonts or gametocytes of *P. yoelii* (Py.S, Py.G) and *P. falciparum* (Pf.S, Pf.GIII, Pf.GIV, Pf.GV). See Dataset S1 for the proportions of transcripts carrying more than 20 m^{5}C sites. **D**, Proportions of nucleotide triplets extending from m^{5}C sites upstream or downstream in the parasite stages of *P. yoelii* (Py) or *P. falciparum* (Pf). H represents A, C or U. **E**, Transcriptome-wide distributions of mRNA m^{5}C sites in the predicted CDS, introns, and untranslated

flanking regions in schizonts and gametocytes of *P. yoelii* (Py.S, Py.G) and *P. falciparum* (Pf.S, Pf.GIII, Pf.GIV, Pf.GV). The presence of m^5C in predicted intron regions suggests the methylation of incompletely or alternatively spliced transcripts.

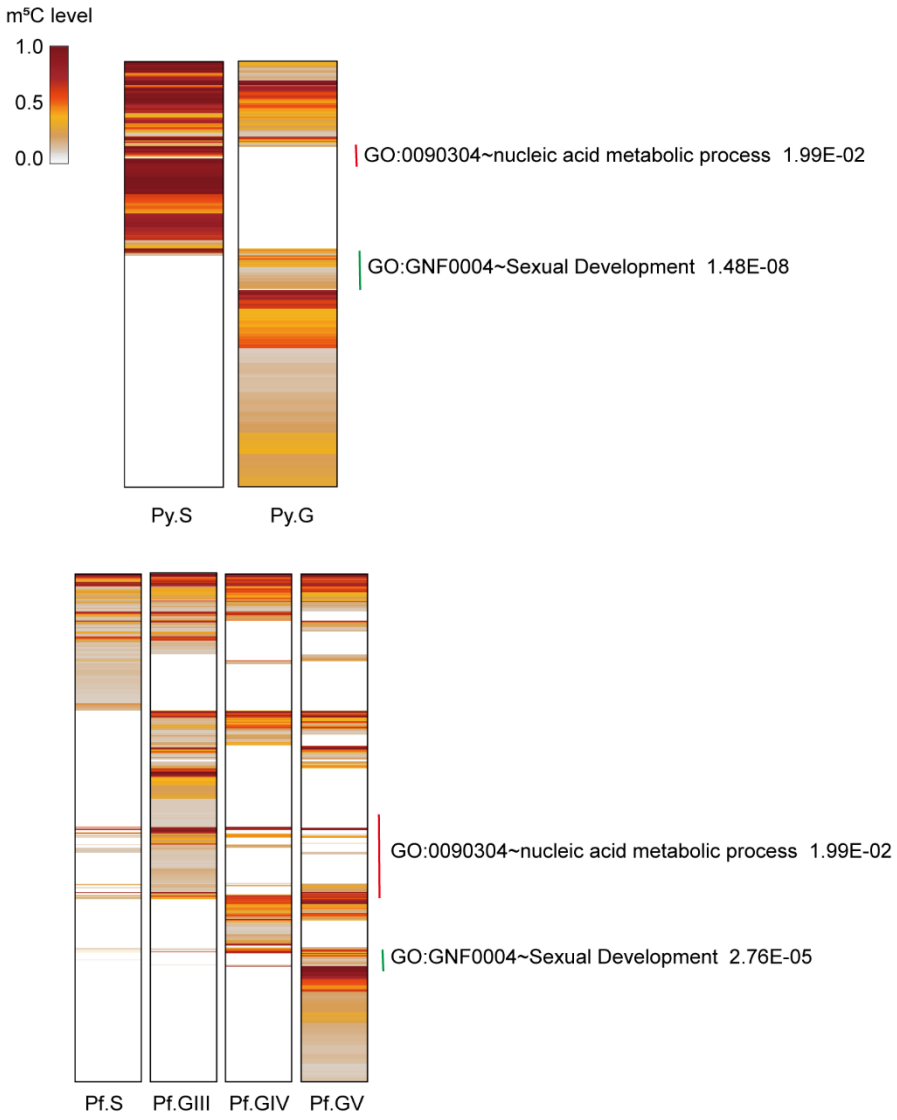


Fig. S2. **Heatmaps representing transcript m^{5}C levels in schizont and gametocyte stages of *P. yoelii* (Py.S, Py.G) and *P. falciparum* (Pf.S, Pf.GIII, Pf.GIV, Pf.GV).** Vertical bars indicate regions with sequences that match with GO terms for sexual development and nucleic acid metabolic processes.

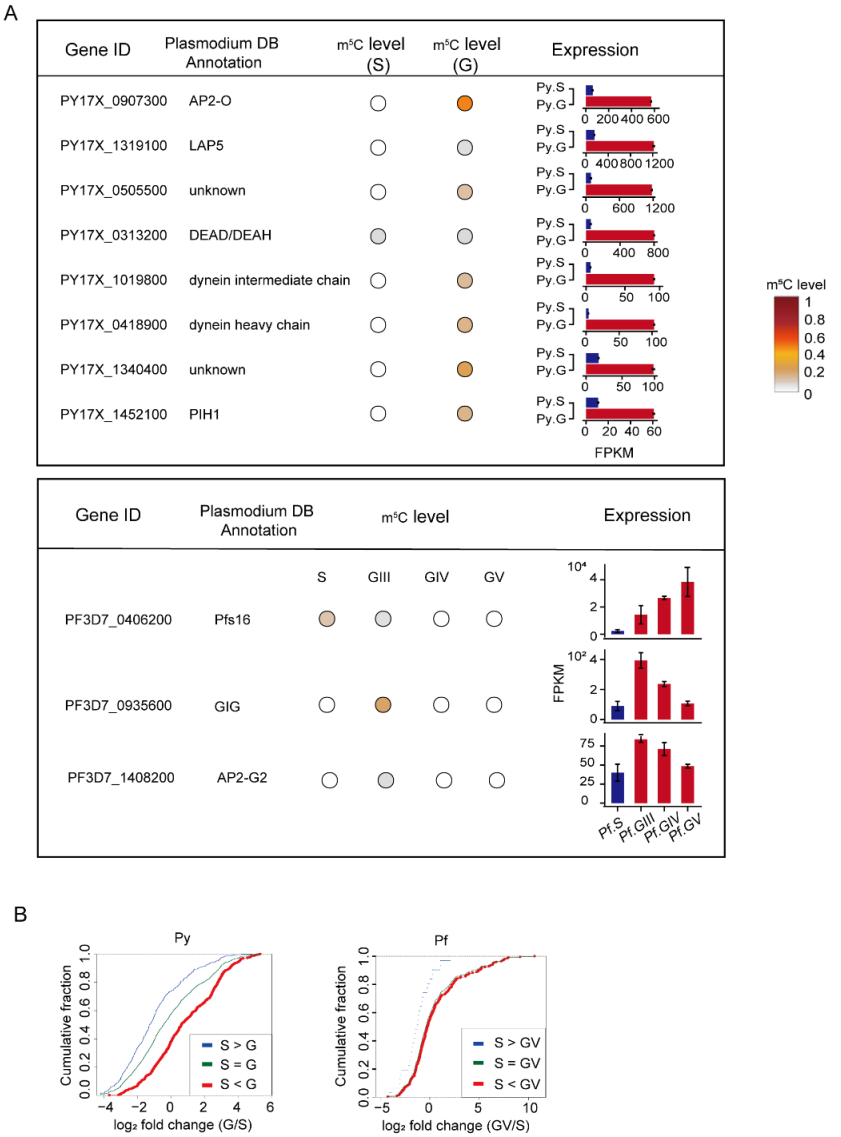


Fig. S3. Comparatively high m^{5}C methylation levels in the transcripts of gametocytogenesis-associated genes. **A**, Relative m^{5}C levels are indicated for selected gametocytogenesis-associated transcripts alongside heatmaps of their relative expression levels in *P. yoelii* and *P. falciparum* schizonts and gametocytes. Data on the m^{5}C and expression levels were each obtained from two biological replicates using RNA-BisSeq and RNA-Seq, respectively. **B**, Abundance of transcripts in gametocytes relative to schizonts correlates with their stage-specific m^{5}C methylation levels. Levels of individual transcripts in gametocytes of *P. yoelii* (G) or *P. falciparum* (GV) relative to their levels in schizonts (S) are plotted by $\log_2$ ratio. Results from transcripts with lesser or greater m^{5}C levels in gametocytes than schizonts are shown in blue (S > G) and red (S < G), respectively, while those with little or no difference of m^{5}C levels are shown in green (S = G). The m^{5}C levels were normalized to mRNA transcript abundance levels before comparison.

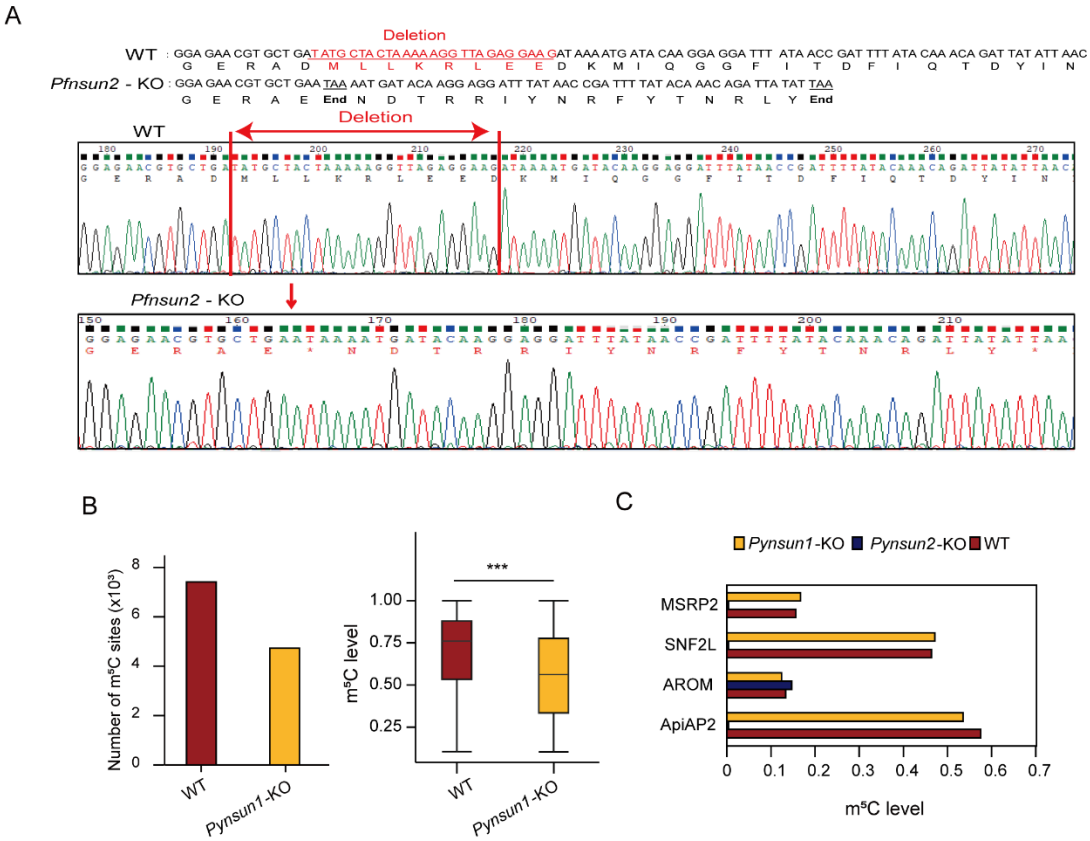


Fig. S4 Disruption of the *Pfnsun2* gene in *P. falciparum* and of *Pynsun2* gene in *P. yoelii*. A, Predicted protein sequences and electropherograms from the *P. falciparum* WT and *Pfnsun2*-KO parasites. An early stop codon (followed by additional stop codons downstream) is introduced by the KO deletion. Bioinformatic searches found no predicted functions for the truncated nub of peptide encoded upstream of the first stop codon. The red arrow indicates the deletion site. **B**, Histogram shows numbers of mRNA m^5C sites detected in schizont stages of WT and *Pynsun1*-KO *P. yoelii* clones. Box plot shows the corresponding m^5C levels in WT and *Pynsun1*-KO parasites (right panel). **(*****: *P*-value < 0.001, Wilcoxon test.) **C**, Box plot shows the m^5C levels of transcripts from the ApiAP2, AROM, SNF2L and MSRO2 genes in WT, *Pynsun2*-KO, and *Pynsun1*-KO parasites.

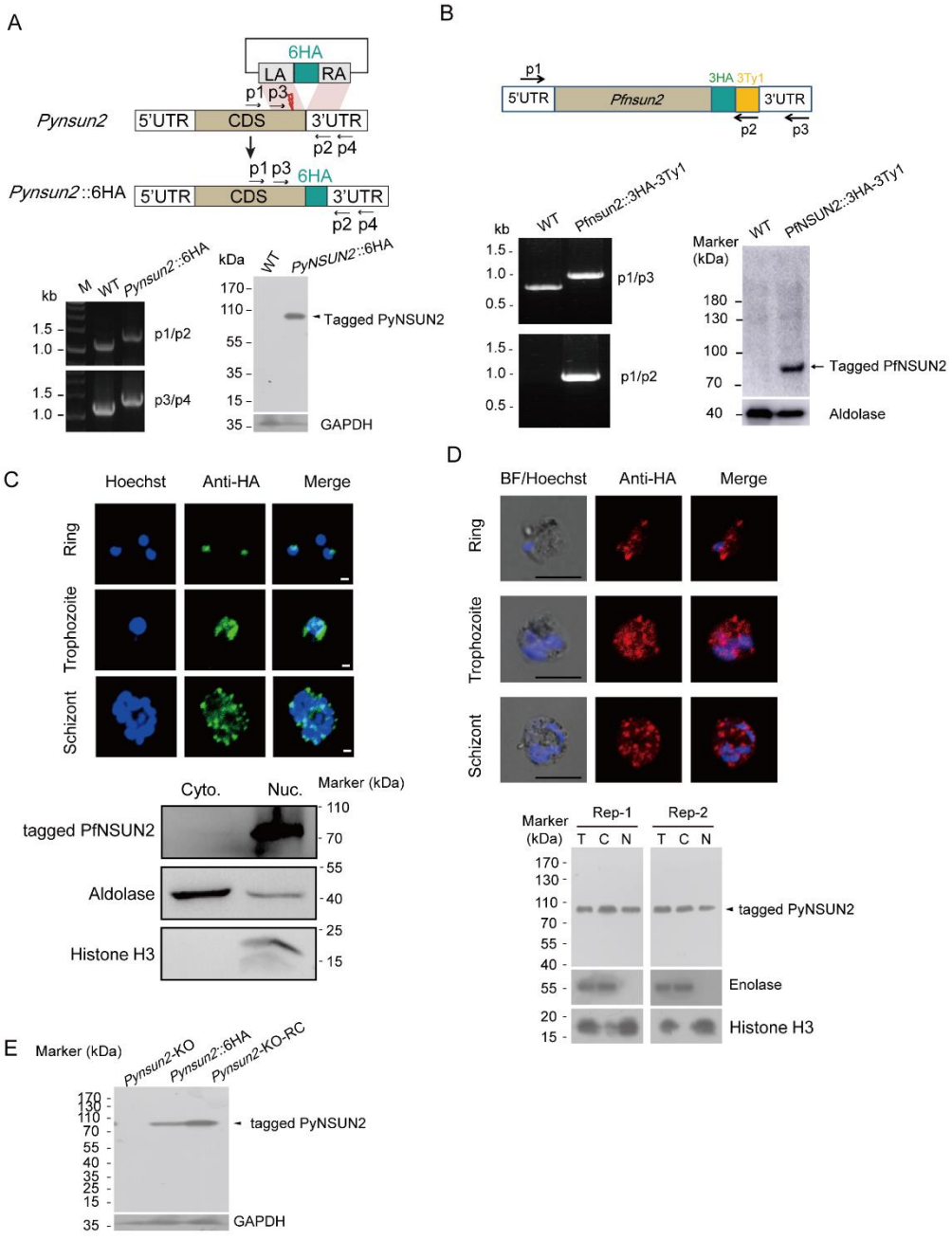


Fig. S5. Cellular localization of *Pfnsun2* and *Pynsun2* in trophozoites and schizonts. A, Schematic representation of the CRISPR/Cas9-mediated tagging of the endogenous *Pynsun2* gene. LA: left arm; RA: right arm. 6HA: hexameric hemagglutinin epitope tag for monoclonal antibody detection. Red thunderbolt indicates the site for sgRNA targeting. The lower left panel shows the PCR products from the modified parasite clone *Pynsun2::6HA* that were amplified by the indicated primer pairs. Lower right panel shows Western blot detection of PyNSUN2:6HA protein in parasite blood stages. GAPDH detection was used for the loading control. **B**, Upper panel shows the construct for the generation of the epitope-tagged *P. falciparum* line by CRISPR-Cas9 modification. The lower left panel shows the PCR products in the modified PyNSUN2::3HA-3Ty1 parasite clone that were obtained by the indicated primer sets.

Lower, right panel shows Western blot detection of the tagged protein in asexual blood stages of PfNSUN2::3HA-3Ty1 parasites with anti-HA monoclonal antibodies. *P. falciparum* aldolase served as the loading control. **C**, Upper panel: IFA detection of tagged protein in PfNSUN2::3HA-3Ty1 ring, trophozoite, and schizont stage parasites. Nuclei were stained by DAPI; the tagged protein was detected by anti-HA monoclonal antibody. The bar represents 1 μm. Lower panel shows Western blot analysis of nuclear and cytoplasmic extracts from the PfNSUN2::3HA-3Ty1 blood stages using anti-HA monoclonal antibody. Aldolase and Histone 3 detection served for confirmation of the nuclear and cytoplasmic extracts and loading controls. Cyto., cytoplasmic extract. Nuc., nuclear extract. **D**, Upper panel shows IFA detection of tagged protein in PyNSUN2::6HA ring, trophozoite and schizont stage parasites. Scale bar = 5 μm. Lower panel shows Western blot analysis demonstrating HA-tagged PyNSUN2 protein in the cytoplasmic (C) and nuclear (N) cell fractions from schizont stage parasites. Enolase and histone H3 served for confirmation of the cytoplasmic and nuclear extracts and loading controls. **E**, Western blot detection of 6×HA tagged PyNSUN2 protein after its restoration in the *Pynsun2*-KO-RC line. *Pynsun2*-KO and a separate line of *Pynsun2*::6HA parasites serve as negative and positive controls, respectively. GAPDH detection served for loading controls.

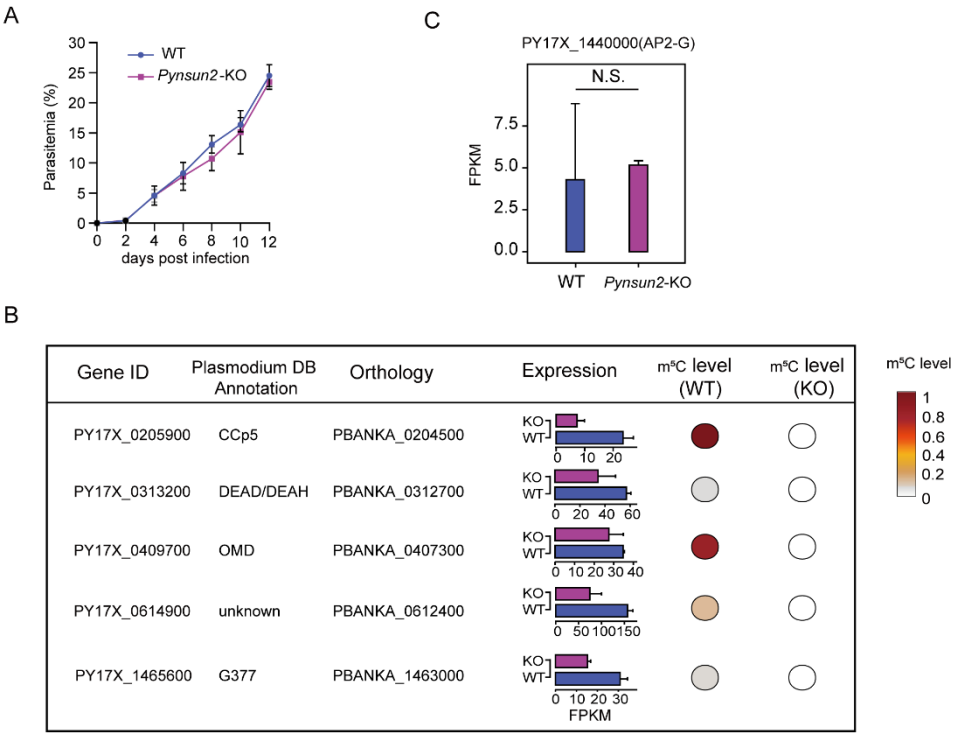


Fig. S6. **Comparable development of blood stage parasitemias but marked reduction of AP2-G induced genes and their m^{5}C levels in *Pynsun2*-KO relative to WT *P. yoelii* parasites. A**, Results demonstrating comparable development of blood stage parasitemias in WT- and *Pynsun2*-KO-infected mice. Data points present the means ± SEM from three independent biological experiments. **B**, Decreased expression of AP2-G induced genes associated with reduced m^{5}C levels in the *Pynsun2*-KO (KO) parasites. Error bars represent SEM calculated from two biological replicates. **C**, Histogram of AP2-G (PY17X_1440000) FPKM indicating transcript levels in WT and *Pynsun2*-KO parasites. Medians and 95% CIs from two biological replicate experiments are indicated. **(**N.S.: no significant difference, Wilcoxon test)

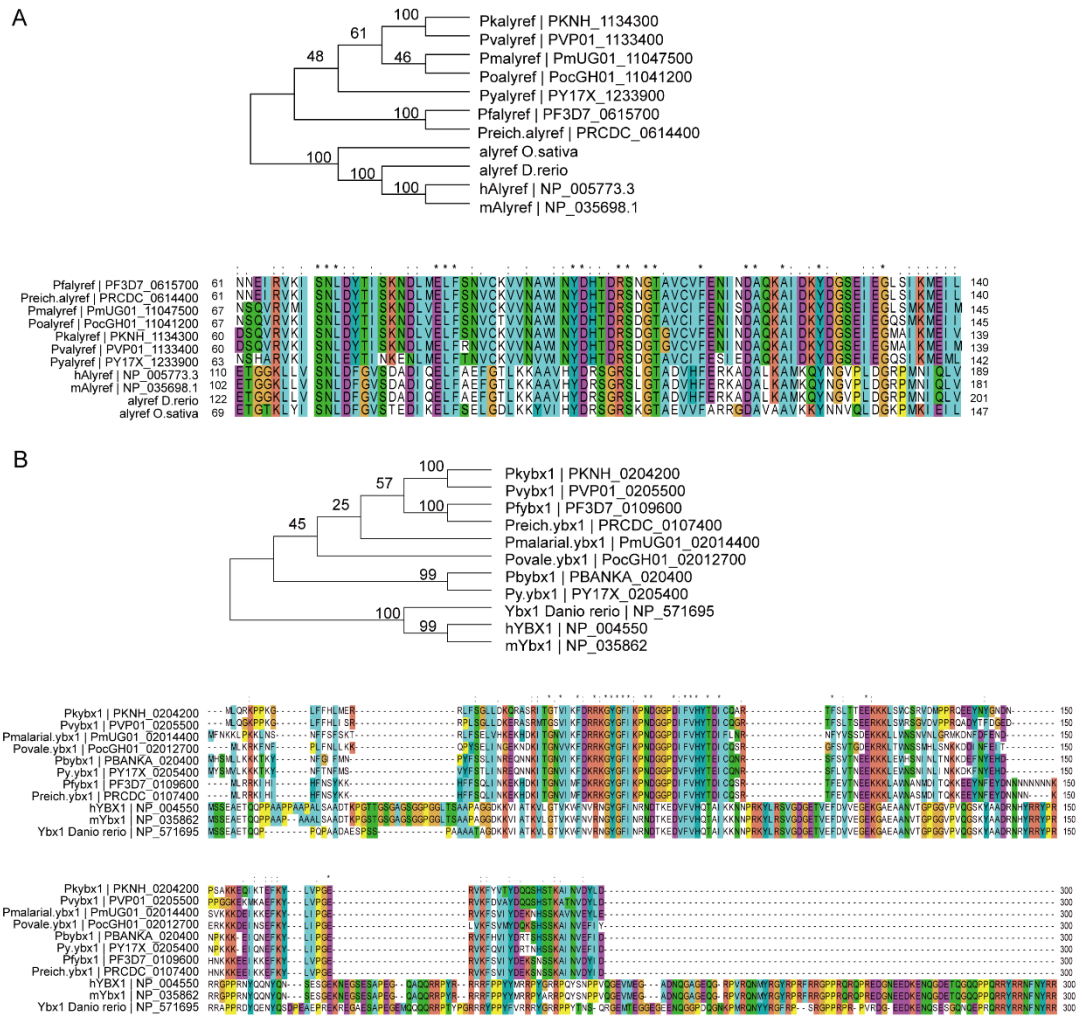
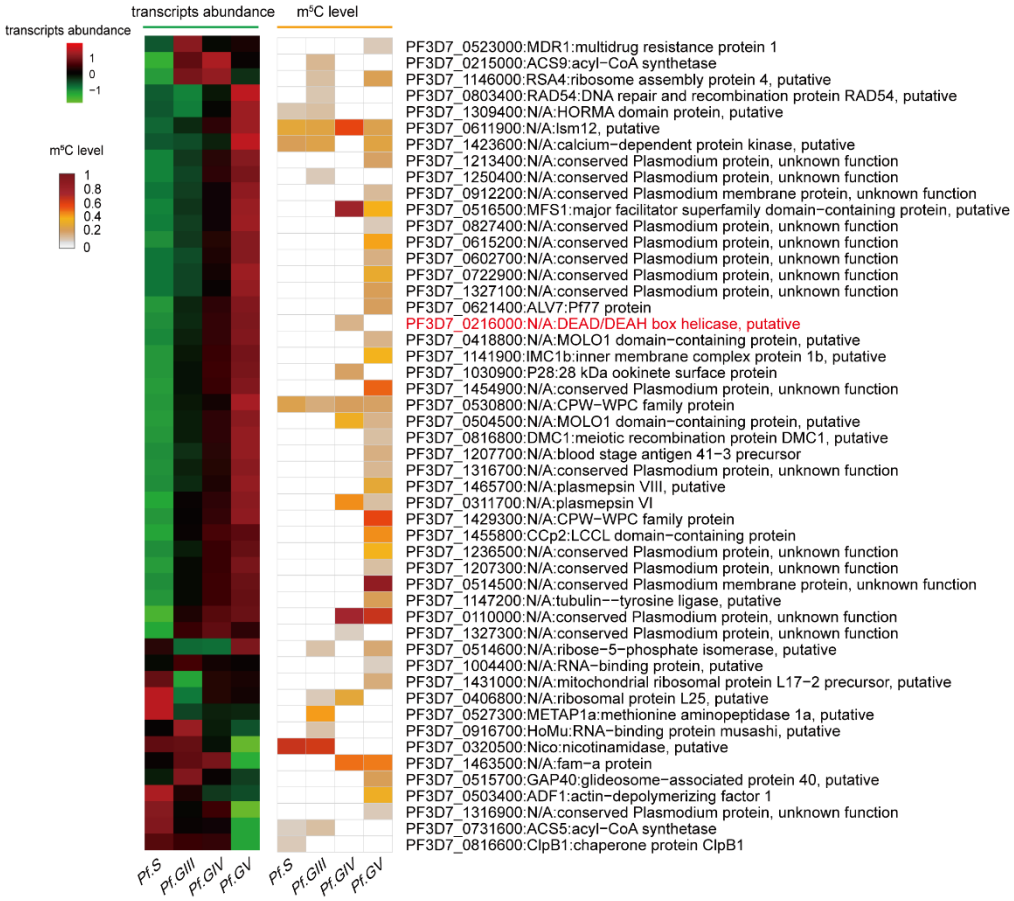
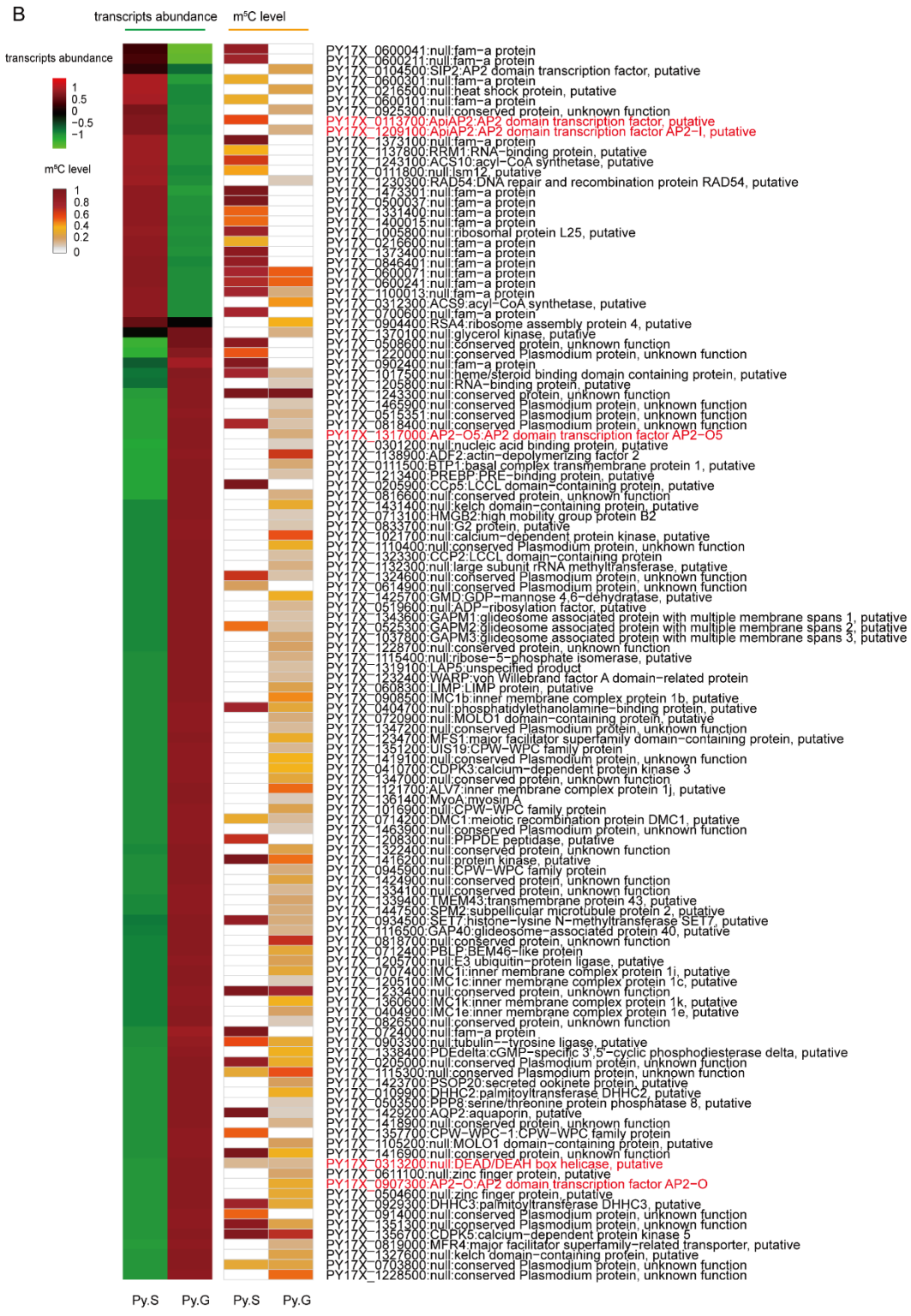


Fig. S7. **Phylogenetic analysis of *P. yoelii* and *P. falciparum* genes with potential homology to RNA m^{5}C readers ALYREF (A) and Ybx1 (B).** The neighbor-joining (NJ) phylogeny was performed using MEGA 5.2.2 with 1000 replicates. Sequences were aligned using ClustalX 2.1.

A


transcripts abundance
m5C level
transcripts abundance
1
0
−1
m5C level
1
0.8
0.6
0.4
0.2
0
Pf.S
Pf.GIII
Pf.GIV
Pf.GV
Pf.S
Pf.GIII
Pf.GIV
Pf.GV
PF3D7_0523000:MDR1:multidrug resistance protein 1
PF3D7_0215000:ACS9:acyl−CoA synthetase
PF3D7_1146000:RSA4:ribosome assembly protein 4, putative
PF3D7_0803400:RAD54:DNA repair and recombination protein RAD54, putative
PF3D7_1309400:N/A:HORMA domain protein, putative
PF3D7_0611900:N/A:lsm12, putative
PF3D7_1423600:N/A:calcium−dependent protein kinase, putative
PF3D7_1213400:N/A:conserved Plasmodium protein, unknown function
PF3D7_1250400:N/A:conserved Plasmodium protein, unknown function
PF3D7_0912200:N/A:conserved Plasmodium membrane protein, unknown function
PF3D7_0516500:MFS1:major facilitator superfamily domain−containing protein, putative
PF3D7_0827400:N/A:conserved Plasmodium protein, unknown function
PF3D7_0615200:N/A:conserved Plasmodium protein, unknown function
PF3D7_0602700:N/A:conserved Plasmodium protein, unknown function
PF3D7_0722900:N/A:conserved Plasmodium protein, unknown function
PF3D7_1327100:N/A:conserved Plasmodium protein, unknown function
PF3D7_0621400:ALV7:Pf77 protein
PF3D7_0216000:N/A:DEAD/DEAH box helicase, putative
PF3D7_0418800:N/A:MOLO1 domain−containing protein, putative
PF3D7_1141900:IMC1b:inner membrane complex protein 1b, putative
PF3D7_1030900:P28:28 kDa ookinete surface protein
PF3D7_1454900:N/A:conserved Plasmodium protein, unknown function
PF3D7_0530800:N/A:CPW−WPC family protein
PF3D7_0504500:N/A:MOLO1 domain−containing protein, putative
PF3D7_0816800:DMC1:meiotic recombination protein DMC1, putative
PF3D7_1207700:N/A:blood stage antigen 41−3 precursor
PF3D7_1316700:N/A:conserved Plasmodium protein, unknown function
PF3D7_1465700:N/A:plasmepsin VIII, putative
PF3D7_0311700:N/A:plasmepsin VI
PF3D7_1429300:N/A:CPW−WPC family protein
PF3D7_1455800:CCp2:LCCL domain−containing protein
PF3D7_1236500:N/A:conserved Plasmodium protein, unknown function
PF3D7_1207300:N/A:conserved Plasmodium protein, unknown function
PF3D7_0514500:N/A:conserved Plasmodium membrane protein, unknown function
PF3D7_1147200:N/A:tubulin−−tyrosine ligase, putative
PF3D7_0110000:N/A:conserved Plasmodium protein, unknown function
PF3D7_1327300:N/A:conserved Plasmodium protein, unknown function
PF3D7_0514600:N/A:ribose−5−phosphate isomerase, putative
PF3D7_1004400:N/A:RNA−binding protein, putative
PF3D7_1431000:N/A:mitochondrial ribosomal protein L17−2 precursor, putative
PF3D7_0406800:N/A:ribosomal protein L25, putative
PF3D7_0527300:METAP1a:methionine aminopeptidase 1a, putative
PF3D7_0916700:HoMu:RNA−binding protein musashi, putative
PF3D7_0320500:Nico:nicotinamidase, putative
PF3D7_1463500:N/A:fam−a protein
PF3D7_0515700:GAP40:glideosome−associated protein 40, putative
PF3D7_0503400:ADF1:actin−depolymerizing factor 1
PF3D7_1316900:N/A:conserved Plasmodium protein, unknown function
PF3D7_0731600:ACS5:acyl−CoA synthetase
PF3D7_0816600:ClpB1:chaperone protein ClpB1


B
transcripts abundance
m5C level
transcripts abundance
1
0.5
0
−0.5
−1
m5C level
1
0.8
0.6
0.4
0.2
0
PY17X_0600041:null:fam−a protein
PY17X_0600211:null:fam−a protein
PY17X_0104500:SIP2:AP2 domain transcription factor, putative
PY17X_0600301:null:fam−a protein
PY17X_0216500:null:heat shock protein, putative
PY17X_0600101:null:fam−a protein
PY17X_0925900:null:conserved protein, unknown function
PY17X_0113700:ApiAP2:AP2 domain transcription factor, putative
PY17X_1209100:ApiAP2:AP2 domain transcription factor AP2−I, putative
PY17X_1373100:null:fam−a protein
PY17X_1137800:RRM1:RNA−binding protein, putative
PY17X_1243100:ACS10:acyl−CoA synthetase, putative
PY17X_0111600:null:lsm12, putative
PY17X_1230300:RAD54:DNA repair and recombination protein RAD54, putative
PY17X_1473301:null:fam−a protein
PY17X_0500037:null:fam−a protein
PY17X_1331400:null:fam−a protein
PY17X_1400015:null:fam−a protein
PY17X_1005800:null:ribosomal protein L25, putative
PY17X_0216600:null:fam−a protein
PY17X_1373400:null:fam−a protein
PY17X_0846401:null:fam−a protein
PY17X_0600071:null:fam−a protein
PY17X_0600241:null:fam−a protein
PY17X_1100013:null:fam−a protein
PY17X_0312300:ACS9:acyl−CoA synthetase, putative
PY17X_0700600:null:fam−a protein
PY17X_0904400:RSA4:ribosome assembly protein 4, putative
PY17X_1370100:null:glycerol kinase, putative
PY17X_0508600:null:conserved protein, unknown function
PY17X_1220000:null:conserved Plasmodium protein, unknown function
PY17X_0902400:null:fam−a protein
PY17X_1017500:null:heme/steroid binding domain containing protein, putative
PY17X_1205800:null:RNA−binding protein, putative
PY17X_1243300:null:conserved protein, unknown function
PY17X_1465900:null:conserved Plasmodium protein, unknown function
PY17X_0515351:null:conserved Plasmodium protein, unknown function
PY17X_0818400:null:conserved Plasmodium protein, unknown function
PY17X_1317000:AP2−O5:AP2 domain transcription factor AP2−O5
PY17X_0301200:null:nucleic acid binding protein, putative
PY17X_1138900:ADF2:actin−depolymerizing factor 2
PY17X_0111500:BTP1:basal complex transmembrane protein 1, putative
PY17X_1213400:PREBP:PRE−binding protein, putative
PY17X_0205900:CCp5:LCCL domain−containing protein, putative
PY17X_0816600:null:conserved protein, unknown function
PY17X_1431400:null:kelch domain−containing protein, putative
PY17X_0713100:HMGB2:high mobility group protein B2
PY17X_0833700:null:G2 protein, putative
PY17X_1021700:null:calcium−dependent protein kinase, putative
PY17X_1110400:null:conserved Plasmodium protein, unknown function
PY17X_1323300:CCP2:LCCL domain−containing protein
PY17X_1132300:null:large subunit rRNA methyltransferase, putative
PY17X_1324600:null:conserved Plasmodium protein, unknown function
PY17X_0614900:null:conserved Plasmodium protein, unknown function
PY17X_1425700:GMD:GDP−mannose 4,6−dehydratase, putative
PY17X_0519600:null:ADP−ribosylation factor, putative
PY17X_1343600:GAPM1:glideosome associated protein with multiple membrane spans 1, putative
PY17X_0525300:GAPM2:glideosome associated protein with multiple membrane spans 2, putative
PY17X_1037800:GAPM3:glideosome associated protein with multiple membrane spans 3, putative
PY17X_1228700:null:conserved protein, unknown function
PY17X_1115400:null:ribose−5−phosphate isomerase, putative
PY17X_1319100:LAP5:unspecified product
PY17X_1232400:WARP:von Willebrand factor A domain−related protein
PY17X_0608300:LIMP:LIMP protein, putative
PY17X_0908500:IMC1b:inner membrane complex protein 1b, putative
PY17X_0404700:null:phosphatidylethanolamine−binding protein, putative
PY17X_0720900:null:MOLO1 domain−containing protein, putative
PY17X_1347200:null:conserved Plasmodium protein, unknown function
PY17X_1234700:MFS1:major facilitator superfamily domain−containing protein, putative
PY17X_1351200:UIS19:CPW−WPC family protein
PY17X_1419100:null:conserved Plasmodium protein, unknown function
PY17X_0410700:CDPK3:calcium−dependent protein kinase 3
PY17X_1347000:null:conserved protein, unknown function
PY17X_1121700:ALV7:inner membrane complex protein 1j, putative
PY17X_1361400:MyoA:myosin A
PY17X_1016900:null:CPW−WPC family protein
PY17X_0714200:DMC1:meiotic recombination protein DMC1, putative
PY17X_1463900:null:conserved Plasmodium protein, unknown function
PY17X_1208300:null:PPPDE peptidase, putative
PY17X_1322400:null:conserved protein, unknown function
PY17X_1416200:null:protein kinase, putative
PY17X_0945900:null:CPW−WPC family protein
PY17X_1424900:null:conserved protein, unknown function
PY17X_1344100:null:conserved protein, unknown function
PY17X_1339400:TMEM43:transmembrane protein 43, putative
PY17X_1447500:SPM2:subpellicular microtubule protein 2, putative
PY17X_0934500:SET7:histone−lysine N−methyltransferase SET7, putative
PY17X_1116500:GAP40:glideosome−associated protein 40, putative
PY17X_0818700:null:conserved protein, unknown function
PY17X_0712400:PBLP:BEM46−like protein
PY17X_1205700:null:E3 ubiquitin−protein ligase, putative
PY17X_0707400:IMC1i:inner membrane complex protein 1i, putative
PY17X_1205100:IMC1c:inner membrane complex protein 1c, putative
PY17X_1233400:null:conserved protein, unknown function
PY17X_1360600:IMC1k:inner membrane complex protein 1k, putative
PY17X_0404900:IMC1e:inner membrane complex protein 1e, putative
PY17X_0826500:null:conserved protein, unknown function
PY17X_0724000:null:fam−a protein
PY17X_0903300:null:tubulin−−tyrosine ligase, putative
PY17X_1338400:PDEdelta:cGMP−specific 3',5'−cyclic phosphodiesterase delta, putative
PY17X_0205000:null:conserved Plasmodium protein, unknown function
PY17X_1115300:null:conserved Plasmodium protein, unknown function
PY17X_1423700:PSOP20:secreted ookinete protein, putative
PY17X_0109900:DHHC2:palmitoyltransferase DHHC2, putative
PY17X_0503500:PPP8:serine/threonine protein phosphatase 8, putative
PY17X_1429200:AQP2:aquaporin, putative
PY17X_1418900:null:conserved protein, unknown function
PY17X_1357700:CPW−WPC−1:CPW−WPC family protein
PY17X_1105200:null:MOLO1 domain−containing protein, putative
PY17X_1416900:null:conserved protein, unknown function
PY17X_0313200:null:DEAD/DEAH box helicase, putative
PY17X_0611100:null:zinc finger protein, putative
PY17X_0907300:AP2−O:AP2 domain transcription factor AP2−O
PY17X_0504600:null:zinc finger protein, putative
PY17X_0929300:DHHC3:palmitoyltransferase DHHC3, putative
PY17X_0914000:null:conserved Plasmodium protein, unknown function
PY17X_1351300:null:conserved Plasmodium protein, unknown function
PY17X_1356700:CDPK5:calcium−dependent protein kinase 5
PY17X_0819000:MFR4:major facilitator superfamily−related transporter, putative
PY17X_1327600:null:kelch domain−containing protein, putative
PY17X_0703800:null:conserved Plasmodium protein, unknown function
PY17X_1228500:null:conserved Plasmodium protein, unknown function
Py.S
Py.G
Py.S
Py.G

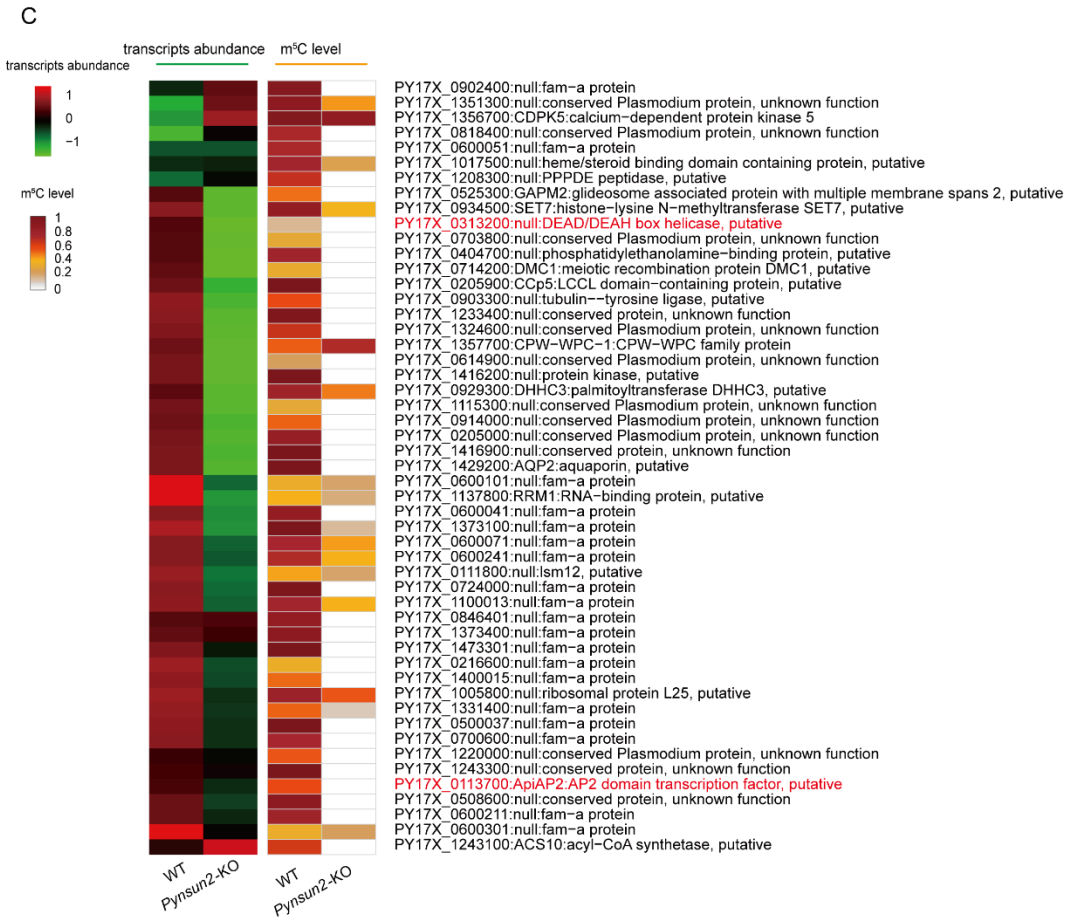


Fig. S8. **Transcript abundance and m^{5}C levels of mRNAs bound by DOZI in *P. yoelii* and *P. falciparum*. A,** Heatmaps of the transcript abundances (gene expression) and m^{5}C levels of mRNAs bound by DOZI in WT *P. falciparum*. **B,** Heat maps of the transcript abundances and m^{5}C levels of mRNAs bound by DOZI in stages of WT *P. yoelii*. **C,** Transcript abundances and m^{5}C levels of mRNAs bound by DOZI in schizont stages of the *P. yoelii* WT and *Pynsun2*-KO lines.

Dataset S1 (separate file). Summary of RNA-BisSeq mapping results and statistics of m^5C sites per transcript from schizonts and gametocytes of *P. yoelii* and *P. falciparum.*

Dataset S2 (separate file). m^5C levels of transcripts in schizont and/or gametocyte stages of *P. falciparum* and *P. yoelii* (WT, *Pynsun2*-KO, and *Pynsun2*-KO-RC lines).

Dataset S3 (separate file). Transcripts with greater than 3-fold upregulation in gametocytes relative to schizonts of the *P. falciparum* (NF54) and *P. yoelii* (17XNL) lines.

Dataset S4 (separate file). Individual half-lives and m^5C levels of transcripts in *P. yoelii* gametocyte stages.

Dataset S5 (separate file). Protein abundances and m^5C levels of transcripts in schizonts and gametocytes of *P. yoelii*.

Dataset S6 (separate file). sgRNA and primer sequences used in constructs and designs for gene knock-out experiments.

Dataset S7 (separate file). *P. yoelii* WT, *Pynsun2*-KO, and *Pynsun2*-KO-RC transcript levels for genes showing ≥ 3-fold expression change after *Pynsun2* knock-out.

Dataset S8 (separate file). mRNA and m^5C levels of DOZI-associated transcripts in *P. yoelii* and *P. falciparum.*

References

1. Shen LS, Liang Z, & Yu H (2017) Dot blot analysis of N6-methyladenosine RNA modification levels. *Bio-Protocol* 7(1):e2095.
2. Baumgarten S, *et al.* (2019) Transcriptome-wide dynamics of extensive m(6)A mRNA methylation during Plasmodium falciparum blood-stage development. *Nat Microbiol* 4(12):2246-2259.
3. Guo G, *et al.* (2020) Disease activity-associated alteration of mRNA m(5)C methylation in CD4(+) T cells of systemic lupus erythematosus. *Front Cell Dev Biol* 8:430.
4. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17(1):10-12.
5. Kim D, Langmead B, & Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12(4):357-360.
6. Liao Y, Smyth GK, & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923-930.
7. Shock JL, Fischer KF, & DeRisi JL (2007) Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol* 8(7):R134.
8. Chen X, *et al.* (2019) 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat Cell Biol* 21(8):978-990.
9. Wang X, *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505(7481):117-120.
10. Moré JJ (1978) The Levenberg-Marquardt algorithm: implementation and theory. *Proceedings of the Biennial Conference Held at Dundee, June 28–July 1, 1977*, Numerical Analysis. Lecture Notes in Mathematics, ed Watson (Springer, Berlin, Heidelberg), Vol 630, pp 105-116.
11. Rieder D, Amort T, Kugler E, Lusser A, & Trajanoski Z (2016) meRanTK: methylated RNA analysis ToolKit. *Bioinformatics* 32(5):782-785.
12. Yang X, *et al.* (2017) 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res* 27(5):606-625.
13. Schaefer M (2015) RNA 5-methylcytosine analysis by bisulfite sequencing. *Methods Enzymol* 560:297-329.
14. Wagih O (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33(22):3645-3647.
15. Fan Y, *et al.* (2020) Rrp6 regulates heterochromatic gene silencing via ncRNA RUF6 decay in malaria parasites. *mBio* 11(3):e01110-01120.
16. Jiang L, *et al.* (2013) PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. *Nature* 499(7457):223-227.
17. Jiang Y, *et al.* (2020) An intracellular membrane protein GEP1 regulates xanthurenic acid induced gametogenesis of malaria parasites. *Nat Commun* 11(1):1764.
18. Wright DA, *et al.* (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* 1(3):1637-1652.
19. Jia S, *et al.* (2020) TMT-based proteomic analysis of the fish-borne spoiler Pseudomonas psychrophila subjected to chitosan oligosaccharides in fish juice system. *Food Microbiol* 90:103494.

20. Voss TS, Mini T, Jenoe P, & Beck HP (2002) Plasmodium falciparum possesses a cell cycle-regulated short type replication protein A large subunit encoded by an unusual transcript. *J Biol Chem* 277(20):17493-17501.
21. Zhang Q, *et al.* (2014) Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. *Nature* 513(7518):431-435.