International Journal for Parasitology 49 (2019) 501–513

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Transcriptome and histone epigenome of *Plasmodium vivax* salivary-gland sporozoites point to tight regulatory control and mechanisms for liver-stage differentiation in relapsing malaria

Vivax Sporozoite Consortium[1] Ivo Muller[a,b,c], Aaron R. Jex[a,c,d,*], Stefan H.I. Kappe[e], Sebastian A. Mikolajczak[e], Jetsumon Sattabongkot[g], Rapatbhorn Patrapuvich[f], Scott Lindner[h], Erika L. Flannery[e], Cristian Koepfli[a], Brendan Ansell[d], Anita Lerch[a], Samantha J. Emery-Corbin[a], Sarah Charnaud[a], Jeffrey Smith[a], Nicolas Merrienne[b], Kristian E. Swearingen[i], Robert L. Moritz[i], Michaela Petter[j,k], Michael F. Duffy[j], Vorada Chuenchob[e]

[a] *Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia*
[b] *Malaria: Parasites & Hosts Unit, Institut Pasteur, 28 Rue de Dr. Roux, 75015 Paris, France*
[c] *Department of Medical Biology, The University of Melbourne, Victoria 3010, Australia*
[d] *Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Corner of Park and Flemington Road, Parkville, Victoria 3010, Australia*
[e] *Seattle Children's Research Institute, 307 Westlake Avenue North, Suite 500, Seattle, WA 98109, USA*
[g] *Mahidol Vivax Research Center, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand*
[f] *Department of Global Health, University of Washington, Seattle, WA 98195, USA*
[h] *Department of Biochemistry and Molecular Biology, Center for Malaria Research, Pennsylvania State University, University Park, PA 16802, USA*
[i] *Institute for Systems Biology, Seattle, WA 98109, USA*
[j] *Department of Medicine Royal Melbourne Hospital, The Peter Doherty Institute, The University of Melbourne, 792 Elizabeth Street, Melbourne, Victoria 3000, Australia*
[k] *Institute of Microbiology, University Hospital Erlangen, Erlangen 91054, Germany*

ARTICLE INFO

Article history:
Received 9 December 2018
Received in revised form 1 February 2019
Accepted 7 February 2019
Available online 7 May 2019

Keywords:
Plasmodium vivax
Sporozoite
Transcriptome
RNA-seq
CHiP-Seq
Dormancy
Hypnozoite

ABSTRACT

Plasmodium vivax is the key obstacle to malaria elimination in Asia and Latin America, largely attributed to its ability to form resilient hypnozoites (sleeper cells) in the host liver that escape treatment and cause relapsing infections. The decision to form hypnozoites is made early in the liver infection and may already be set in sporozoites prior to invasion. To better understand these early stages of infection, we undertook a comprehensive transcriptomic and histone epigenetic characterization of *P. vivax* sporozoites. Through comparisons with recently published proteomic data for the *P. vivax* sporozoite, our study found that although highly transcribed, transcripts associated with functions needed for early infection of the vertebrate host are not detectable as proteins and may be regulated through translational repression. We identified differential transcription between the sporozoite and published transcriptomes of asexual blood stages and mixed versus hypnozoite-enriched liver stages. These comparisons point to multiple layers of transcriptional, post-transcriptional and post-translational control that appear active in sporozoites and to a lesser extent hypnozoites, but are largely absent in replicating liver schizonts or mixed blood stages. We also characterised histone epigenetic modifications in the *P. vivax* sporozoite and explored their role in regulating transcription. Collectively, these data support the hypothesis that the sporozoite is a tightly programmed stage to infect the human host and identify mechanisms for hypnozoite formation that may be further explored in liver stage models.

1. Introduction

Malaria is among the most significant infectious diseases impacting humans globally, with 3.3 billion people at risk of infection, 381 million suspected clinical cases and up to ~660,000 deaths attributed to malaria in 2014 (WHO, 2015). Two major par-

* Corresponding author at: Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia.
E-mail address: jex.a@wehi.edu.au (A.R. Jex).
[1] All authors contributed equally.

https://doi.org/10.1016/j.ijpara.2019.02.007
0020-7519/

asitic species contribute to the vast majority of human malaria, *Plasmodium falciparum* and *Plasmodium vivax*. Historically, *P. falciparum* has attracted the majority of global attention, due to its higher contribution to morbidity and mortality. However, *P. vivax* is broadly distributed, more pathogenic than previously thought, and is recognised as the key obstacle to malaria elimination in the Asia-Pacific and Americas (Feachem et al., 2010). Unlike *P. falciparum*, *P. vivax* can establish long-lasting 'sleeper cells' (=hypnozoites (HPZs)) in the host liver that emerge weeks, months or years after the primary infection (=relapsing malaria) (Price et al., 2009). Primaquine is the only approved drug that prevents relapse. However, the short half-life, long dosage regimens and incompatibility of primaquine with glucose-6-phosphate-dehydro genase deficiency (which requires pre-screening of recipients; Baird, 2013) makes it unsuitable for widespread use. As a consequence, *P. vivax* is overtaking *P. falciparum* as the primary cause of malaria in a number of co-endemic regions (Sattabongkot et al., 2004). Developing new tools to diagnose, treat and/or prevent HPZ infections is considered one of the highest priorities in the malaria elimination research agenda (Mueller et al., 2009).

When *Plasmodium* sporozoites (SPZs) are deposited by an infected mosquito, they likely traverse the skin cells, enter the blood-stream and are trafficked to the host liver, as has been shown in rodents (Lindner et al., 2012). The SPZs' journey from skin deposition to hepatocytes takes less than a few minutes (Shin et al., 1982). Upon reaching the liver, SPZs traverse Kupffer and endothelial cells to reach the parenchyma, moving through several hepatocytes before invading a final hepatocyte suitable for development (Mota et al., 2001; Lindner et al., 2012). With the hepatocyte, *P. vivax* SPZs either immediately continue development as replicating schizonts and establish a blood infection, or delay replication and persist as HPZs. Regulation of this major developmental fate decision is not understood and this represents a key gap in current knowledge of *P. vivax* biology and control.

It has been hypothesized that *P. vivax* SPZs exist within an inoculum as replicating 'tachysporozoites' and relapsing 'bradysporozoites' (Lysenko et al., 1977) and that these subpopulations may have distinct developmental fates as schizonts or HPZs, thus contributing to their relapse phenotype (Lysenko et al., 1977; Price et al., 2007; White, 2011). This observation is supported by the stability of different HPZ phenotypes (ratios of HPZ to schizont formation) in *P. vivax* infections of liver-chimeric mouse models (Mikolajczak et al., 2015). It is well documented that *P. vivax* HPZ activation patterns stratify with climate and geography (White, 2011) and recent modelling suggests transmission potential selects for HPZ phenotype (White et al., 2014). Clearly the ability for *P. vivax* to dynamically regulate HPZ formation and relapse phenotypes in response to high or low transmission periods under different climate conditions would confer a significant evolutionary advantage. Epigenetic programming of the SPZ is intriguing as a potential mechanism to regulate the liver-stage developmental fate.

To determine fates in the SPZ stage, control of protein expression must take place. Studies using rodent malaria parasites have identified genes (Mueller et al., 2005) that are transcribed in SPZs but translationally repressed (i.e., present as transcript but un- or under-represented as protein), via RNA-binding proteins (Silvie et al., 2014a), and ready for immediate translation after the parasites' infection of the mammalian host cell (Mackellar et al., 2010; Mikolajczak et al., 2015). Epigenetic control of PfAP2-G through chromatin structural remodelling regulates gametocyte (dimorphic sexual stages) development in blood stages (Josling and Llinas, 2015). Studies of *P. falciparum* blood stages have identified the importance of histone modifications as a primary epigenetic regulator (Lopez-Rubio et al., 2009; Duffy et al., 2014) and characterized key markers of heterochromatin (H3K9me3) and euchromatin/transcriptional activation (H3K4me3 and H3K9ac). In *P. falciparum* SPZs, these marks are significantly reconfigured during development in the mosquito (Gomez-Diaz et al., 2017) and play a role in the silencing of genes expressed during vertebrate infection (Zanghi et al., 2018). Histone methyltransferase inhibitors stimulate *Plasmodium cynomolgi* HPZs to become schizonts in macaque hepatocytes (Malmquist et al., 2012; Dembele et al., 2014). Further, histone methyltransferases have been implicated in HPZ formation in studies of differential transcription in *P. cynomolgi* liver stages (Dembele et al., 2014; Cubi et al., 2017). It is therefore possible that translational repression and other mechanisms of epigenetic control contribute to the *P. vivax* SPZ fate decision and HPZ formation, persistence and activation.

Despite recent advances (Roobsoong et al., 2015), current approaches for in vitro *P. vivax* culture do not support routine maintenance in the laboratory and tools to directly perturb gene function are not established. Although in vitro liver stage assays and humanised mouse models are being developed (Mikolajczak et al., 2015), 'omics' analysis of *P. vivax* liver stage dormancy has, until recently (Gural et al., 2018), been impossible and even now is in its early stages. Recent characterization (Cubi et al., 2017) of liver stages (HPZs and schizonts) of *P. cynomolgi* (a related and relapsing parasite in macaques) provides valuable insight, but investigations in *P. vivax* directly are needed. The systems analysis of *P. vivax* SPZs that reside in the mosquito salivary glands and are poised for transmission and liver infection offers a key opportunity to gain insight into *P. vivax* infection. *Plasmodium vivax* SPZs have been explored previously by microarray (Westenberger et al., 2010) and most recently, in a single RNA-seq replicate (Kim et al., 2017) and a study on SPZ activation (Roth et al., 2018). No epigenetic data are currently available for any *P. vivax* life-cycle stage. Here, we present a detailed characterization of the *P. vivax* SPZ transcriptome and histone epigenome and use these data to better understand this key infective stage and the role of SPZ programming in invasion and infection of the human host, and development within the host liver.

2. Materials and methods

2.1. Ethics statement

Collection of venous blood from human patients with naturally acquired *P. vivax* infection was approved by the Ethical Review Committee of the Faculty of Tropical Medicine, Mahidol University, Thailand (Human Subjects Protocol number TMEC 11-033) with the informed written consent of each donor individual. All mouse tissue used in the current study was from preserved infected tissues generated previously (Mikolajczak et al., 2015) at the Seattle Children's Research Institute (SCRI: formerly Centre for Infectious Diseases Research) in Seattle, USA, under direct approval of the SCRI Institutional Animal Care and Use Committee (IACUC) and performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, USA. The SCRI has an Assurance from the Public Health Service (PHS Assurance number is A3640-01) through the Office of Laboratory Animal Welfare (OLAW) for work approved by its IACUC. All *P. vivax* isolates used in this prior study (Mikolajczak et al., 2015) were sourced from Tak and Ubon Ratchatani provinces in Thailand, as per the current study.

2.2. Material collection, isolation and preparation

Fifteen field isolates (PvSpz-Thai 1 to 15), representing symptomatic blood-stage malaria infections, were collected as venous blood (20 mL) from patients presenting at malaria clinics in Tak

and Ubon Ratchatani provinces in Thailand. Each isolate was used to establish infections in *Anopheles dirus* colonized at Mahidol University (Bangkok) by membrane feeding (Mikolajczak et al., 2015), and at 14–16 days post blood feeding, ~3–15 million SPZs were harvested per field isolate from the salivary glands of up to 1000 of these mosquitoes as per Kennedy et al. (2012) and shipped in preservative (Trizol (Thermo Scientific, USA) for RNA/DNA or 1% paraformaldehyde for DNA for ChiP-seq) to the Walter and Eliza Hall Institute, Australia (WEHI).

2.3. Transcriptomics sequencing and differential analysis

Messenger RNAs were purified from an aliquot (~0.5–1 million SPZs) of each of nine *P. vivax* field isolates (PvSPZ-Thai1 to Thai9) as per Zhu et al. (2016). Extracted RNA was DNase-treated (Sigma D5307, USA) as per the manufacturer's recommendations. RNA was quantified using the TapeStation High Sensitivity RNA kit (Agilent Technologies, USA) and subjected to RNA-seq on Illumina NextSeq using TruSeq library construction chemistry as per the manufacturer's instructions. Raw reads for each RNA-seq replicate are available through the Sequence Read Archive (Bioproject PRJNA376620). Sequencing adaptors were removed and low quality reads trimmed and filtered using Trimmomatic v. 0.36 (Bolger et al., 2014). To remove host contaminants, processed reads were aligned, as single-end reads, to the *Anopholes dirus* wrari2 genome (VectorBase version W1) using Bowtie2 (Langmead and Salzberg, 2012) (–very-sensitive preset). All non-host reads were aligned to the manually curated transcripts of the *P. vivax* P01 genome (Auburn et al., 2016) using RSEM (Grabherr et al., 2011) (pertinent settings: –bowtie2 –bowtie2-sensitivity-level very_sensitive –calc-ci –ci-memory 10,240 –estimate-rspd –paired-end). Transcript abundance for each gene in each replicate was calculated by RSEM as expected counts (EC) and transcripts per million (TPM) (Supplementary Fig. S1; Supplementary Tables S1 and S2). All replicate data was then assessed for mapping metrics, transcript saturation and other standard quality control metrics using QualiMap v 2.1.3 (Okonechnikov et al., 2016).

Transcriptional abundance in *P. vivax* SPZs was compared qualitatively (by ranked abundance) with previously published microarray data for *P. vivax* salivary gland SPZs (Westenberger et al., 2010) (Supplementary Fig. S2; Supplementary Table S3). As a further quality control, these RNA-seq data were compared also with previously published microarray data for *P. falciparum* salivary gland SPZs (Le Roch et al., 2004), as well as RNA-seq data from salivary gland SPZs generated here for *P. falciparum* (single replicate generated from *P. falciparum* 3D7 laboratory cultures isolated from *Anopholes stephensi*, processed as per Section 2.2, and previously published for *Plasmodium yoelii* (Lindner et al., 2013a). RNA-seq data from these additional *Plasmodium* spp. were (re)-analysed from raw reads and transcriptional abundance for each species was determined (raw counts and pme-EC and TPM data) as described above using gene models current as of 04 October 2016 (PlasmoDB release v29). Interspecific transcriptional behaviour was qualitatively compared by relative ranked abundance in each species using TPM data for single copy orthologs (SCOs; defined in PlasmoDB) only, shared between *P. vivax* and *P. falciparum* or shared among *P. vivax*, *P. falciparum* and *P. yoelii*.

2.3.1. Quantitative differential transcriptional analyses among P. vivax stages

Recently completed studies of the transcriptome of *P. vivax* for SPZ activation (Roth et al., 2018), as well as liver (Gural et al., 2018) and asexual blood stages (Zhu et al., 2016), support comparative transcriptomic study of SPZs, their biology and transcriptional regulation over the *P. vivax* life-cycle. Data for activated SPZs from Roth et al. (2018) had significantly lower depth coverage, with ~0.03 to 0.6 M reads mapping to the *P. vivax* P01 mRNA transcripts; compared with 0.7 to 15.3 M, 2.4 to 10.6 M and 18.7 to 57.6 M mapped reads for salivary SPZ, liver stages (Gural et al., 2018) and asexual blood stages (Zhu et al., 2016), respectively (Supplementary Figs. S3 and S4). This lower coverage could not be compensated for through data normalization and therefore data from Roth et al. (2018) was not included in our quantitative analyses, although qualitatively, many of the highly transcribed genes in Roth et al. (2018) SPZs were among the highly transcribed genes in salivary SPZs from the present study. The remaining RNAseq data presents an analytical challenge in that each (SPZ, liver stage and blood stage) is produced in a separate study and may be influenced by technical batch effects. To address this, we first examined *P. vivax* transcripts in a previous microarray study of multiple *P. vivax* life-cycle stages (Westenberger et al., 2010), including SPZs and several blood stages, to identify genes that were the most transcriptionally stable across the life-cycle. We identified ~160 genes with low transcriptional variability between SPZs and blood stages that covered the breadth of transcript abundance levels in Westenberger et al. (2010). These include ribosomal proteins, histones, translation initiation complex proteins and various chaperones (see Supplementary Fig. S5). We assessed transcription of these 160 genes among the current and recently published RNA-seq data for *P. vivax* and all were of similarly low variability (Supplementary Fig. S6). This suggests that any batch effect between the studies is sufficiently lower than the biological differences between each life-cycle stage, allowing informative comparisons.

To define transcripts that were up-regulated in SPZs relative to blood stages, we remapped raw reads representing early (18–24 h p.i.), mid (30–40 h p.i.) and late (42–46 h p.i.) *P. vivax* blood stage infections from Zhu et al. (2016) to the *P. vivax* P01 transcripts using RSEM and quality controlled using Qualimap, as per Section 2.3. Differential transcription between *P. vivax* salivary gland SPZs and mixed blood stages (Zhu et al., 2016) was assessed using EC data in EdgeR (Nikolayeva and Robinson, 2014) and limma (Ritchie et al., 2015) (differential transcription cut-off: $\geq$ 2-fold change in counts per million (CPM) and a False Discovery Rate (FDR) ≤ 0.05). Pearson Chi squared tests were used to detect over-represented Pfam domains and Gene Ontology (GO) terms among differentially transcribed genes in SPZs (Bonferroni-corrected $P < 0.05$), based on gene annotations in PlasmoDB (release v29).

We also compared transcription of the SPZ stages with recently published liver stage data from Gural et al. (2018) as per the SPZ to blood stage comparisons above, with the following modifications: (i) EC values were normalized using the 'upper quartile' method instead of Trimmed mean of M-value (TMM) normalization, (ii) differential transcription was assessed using a quasi-likelihood generalize linear model (instead of a linear model) and (iii) an FDR threshold for significance of $\leq$0.01 was used instead of $\leq$0.05. These differences related to specific attributes of the liver stage dataset, particularly the small number of replicates ($n = 2$) per experimental condition. Data visualization and interactive R-shiny plots were produced in R using the ggplot2 (Wickham and Chang, 2008; https://cran.r-project.org/web/packages/ggplot2/index.html), gplots (heatmap.2) (Warnes et al., 2009; https://github.com/cran/gplots) and Glimma (Law et al., 2016) packages.

2.4. Assessment of SPZ RNA-seq transcriptome by selective reverse Transcription-quantitative PCR (RT-qPCR)

Two intron-spanning primer pairs were designed per gene of interest using Primer3 and BLAST. Primer pairs were tested in two concentrations (0.75 ng and 2.83 ng per reaction) to determine efficiency and specificity. Product was run on a 1% agarose gel with ethidium bromide. Primer pairs indicating non-specific priming

were removed. The resulting 11 primer pairs (Table 1) were used on four additional SPZ samples; PvSPZ-Thai12 to 15, with RNA extraction, DNAse treatment and quantitation as per Section 2.2. RNA was reverse transcribed (Sensifast, Bioline, USA) and used at 0.75 ng per reaction, run on a Roche LightCycler 480 II. Melt curves were assessed and products were run on a gel to ensure specificity again. The Crossing-point PCR cycle (Cp) threshold was set automatically. ΔCp value was calculated as target gene – comparator gene (SERA and CelTOS were used). Data were log transformed and the fold change calculated.

2.5. Salivary gland SPZ and liver-stage immunofluorescence assays (IFAs)

IFAs were performed as per Mikolajczak et al. (2015) using preserved, vivax-infected mouse liver tissue generated previously for that study. In Mikolajczak et al. (2015), female FRG (fumarylacetoacetate hydrolase (F), recombination activation gene 2 (R), interleukin-2 receptor subunit gamma (G)) triple knock-out (KO) mice engrafted with human hepatocytes (FRG KO huHep) were purchased from Yecuris Corporation (Oregon, USA). Mice were infected through i.v. injection into the tail with 3.5×10^5–1×10^6 SPZs isolated from the salivary glands of infected mosquitoes in 100 μl of RPMI media. Liver stages for the current study were obtained from 10 μm formalin-fixed paraffin-embedded day 7 liver stages generated previously (Mikolajczak et al., 2015) from FRG KO huHep mice (Mikolajczak et al., 2015); these were deparaffinized prior to staining. Fresh salivary gland SPZs were fixed in acetone per Mikolajczak et al. (2015). All cells were incubated twice for 3 min in Xylene, then 100% Ethanol, and finally once for 3 min each in 95%, 70%, and 50% Ethanol. The cells were rinsed in deionized water and permeabilized immediately in 1X Tris Buffered Saline (TBS), containing Triton X-100 and 30% hydrogen peroxide. The cells were blocked in 5% milk in 1X TBS. The hepatocytes were stained overnight with a rabbit polyclonal LISP1 antibody (A), a rabbit polyclonal UIS4 antibody (B), and a rabbit polyclonal BIP antibody (C) in blocking buffer. The cells were washed with 1X TBS and the primary antibodies were detected with goat anti-rabbit Alexa Fluor 488 antibody (Life Technologies). The cells were washed in 1X TBS. The hepatocytes were rinsed in Potassium Permanganate ($KMnO_4$) and washed in 1X TBS. The cells were incubated in DAPI for 5 min.

2.6. Histone chromatin immunoprecipitation sequencing (ChIP-seq) and analysis

Aliquots of 2–6 million freshly isolated SPZs (PvSPZ-Thai7 to Thai11) were fixed with 1% paraformaldehyde for 10 min at 37 °C and the reaction subsequently quenched by adding glycine to a final concentration of 125 mM. After three washes with PBS, SPZ pellets were stored at −80 °C and shipped to Australia. Nuclei were released from the SPZs by dounce homogenization in lysis buffer (10 mM Hepes pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EDTA, 1 mM DTT, 1x EDTA-free protease inhibitor cocktail (Roche, USA), 0.25% NP40). Nuclei were pelleted by centrifugation at 21,000*g* for 10 min at 4 °C and resuspended in SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8.1, 1x EDTA-free protease inhibitor cocktail). Chromatin was sheared into 200–1000 bp fragments by sonication for 16 cycles in 30 s intervals (on/off, high setting) using a Bioruptor (Diagenode, USA) and diluted 1:10 in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris pH 8.1, 150 mM NaCl). Chromatin was precleared for 1 h with protein A/G sepharose (4FastFlow, GE Healthcare, USA) equilibrated in 0.1% BSA (Sigma-Aldrich, USA) in ChIP dilution buffer. Chromatin from 3×10^5 nuclei was taken aside as input material. Chromatin from approximately 3×10^6 SPZ nuclei was used for each ChIP. ChIP was carried out over night at 4 °C with 5 μg of antibody (H3K9me3 (Active Motif), H3K4me3 (Abcam), H3K9ac (Upstate), H4K16ac (Abcam)) and 10 μl each of equilibrated protein A and G sepharose beads (4FastFlow, GE Healthcare). After washes in low salt, high salt, LiCl, and Tris-EDTA (TE) buffers (EZ-ChIP Kit, Millipore, USA), precipitated complexes were eluted in 1% SDS, 0.1 M $NaHCO_3$. Cross-linking of the immune complexes and input material was reversed for 6 h at 45 °C after addition of 500 mM NaCl and 20 μg/ml of proteinase K (NEB). DNA was purified using the a MinElute® PCR purification kit (Qiagen, USA) and paired-end sequenced on Illumina NextSeq using TruSeq library construction chemistry as per the manufacturer's instructions. Raw reads for each ChIP-seq replicate are available through the Sequence Read Archive (Bioproject PRJNA376620).

Fastq files were checked for quality using fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapter sequences were trimmed using cutadapt (Martin, 2011). Paired-end reads were mapped to the *P. vivax* P01 strain genome annotation using Bowtie2 (Langmead and Salzberg, 2012). The alignment files were converted to Bam format, sorted and indexed using Samtools (Li et al., 2009). ChIP peaks were called relative to input using MACS2 (Zhang et al., 2008) in paired-end mode with a q value less than or equal to 0.01. Peaks and peak summits were converted to sorted BED files. Bedtools intersect (Quinlan and Hall, 2010) was used to identify genes that intersected H3K9me3 peaks and Bedtools closest was used to identify genes that were closest to and downstream of H3K9ac and H3K4me3 peak summits.

2.7. Sequence motif analysis

Conserved sequence motifs were identified using the program DREME (Bailey, 2011). Only genes in the top decile of transcription

Table 1
Oligonucleotide primer pairs used for reverse transcription – quantitative PCR based comparison of representative *Plasmodium vivax* transcripts detected by RNAseq here relative to their transcriptional abundance inferred from prior microarray data published by Westenberger et al. (2010).

Name	Gene	Forward Primer	Reverse Primer
RPS27	PVX_122245	ACCACCTTGTTTAGCCATGC	TAATTTGCACTTTCCACCCGTT
D13	PVX_089510	CTGTACACGCACGAGCTGGC	CAGCTCCTTGACGCCACTG
G10	PVX_080110	ACGAGCTGTACTACAAGCGGA	TTTCTCCTGCACCAGGTAGTC
AP2	PVX_086995[a]	GCCCCACTGGAAGTTTTGGA	CGTTCAGCCGCTGGTAGTAT
SERA	PVX_003790	CTGAAGACCTCCAGGGACAAG	TTTCTGCCTCTCCAGTGATATCTTT
CelTOS	PVX_123510*	CCCCCAAAGGCAAAATGAACAA	CGCTCTTTCCCCTCAAGGAC
GEST	PVX_118040	GACATATCAAGCAGTGAGGGA	CATGTTGTGGCCTTTATATGCTG
ALBA4	PVX_083270	TATCAACGGAGCCTTTGCCG	GGACTTGATTTCCTCGTCGG
PUF2	PVX_089945	ATCATAGAGAACGTCGACAAGCTTA	CTACGTTTCCAGGTTGCTGATC
14–3-3	PVX_089505	GACAACTTGACCTTGTGGACGTC	TACTCGAGGCCTTCATCCTTCGATT
ZIPCO	PVX_001980	TTAGCTCAATTGCTTGTGGCTTTTT	TGCCACTAACTCCAAGGAAATAACT

[a] Denotes single exon gene.

showing no evidence of protein expression in multiple salivary gland SPZ replicates were considered as putatively translationally repressed. We queried coding regions and regions upstream of the transcriptional start site (TSS) for each gene, defined by Zhu et al. (2016) and/or predicted here from all RNA-seq data using the Tuxedo suite (Trapnell et al., 2012), for enriched sequence motifs in comparison to 170 genes found to be in the top decile of both transcriptional and expressional abundance in the same SPZ replicates. In searching for motifs associated with highly transcribed genes with stable H3K9ac marks within 1 kb of the TSS (or up to the 3′ end of the next gene upstream), we compared H3K9ac marked genes in the top decile of transcription with the same number of H3K9ac marked genes in the bottom decile of transcription. In both instances, an e-value threshold of 0.05 was considered the minimum threshold for statistical significance.

3. Results and discussion

Mosquito infections were generated by membrane feeding of blood samples taken from *P. vivax*-infected patients in western Thailand. Approximately 3–15 million *P. vivax* SPZs were harvested per isolate from *Anopheles dirus* salivary glands. Using RNA-seq (PvSPZ-Thai1 to 9), we detected transcription for 5714 *P. vivax* genes (Auburn et al., 2016) at high coverage (4930 with a mean CPM $\geq$ 1.0; Supplementary Fig. S1 and Supplementary Table S1 and S2). Among the most highly transcribed genes detected are *csp* (circumsporozoite protein), five *etramps* (early transcribed membrane proteins), including *uis3* (up-regulated in infective SPZs), *uis4* and *lsap-1* (liver stage associated protein 1), a variety of genes involved in cell transversal and initiation of invasion, including *celtos* (cell traversal protein for ookinetes and SPZs), *gest* (gamete egress and SPZ traversal protein), *spect1* (SPZ protein essential for cell traversal) and *siap-1* (SPZ invasion associated protein), and genes associated with translational repression (*alba1*, *alba4* and *Puf2*). Collectively, these genes account for >1/3rd of all transcripts in the SPZ. Although we found only moderate agreement (R^2 = 0.35; Supplementary Fig. S2) between our RNA-seq data and previous microarray data for *P. vivax* SPZs and blood stages (Westenberger et al., 2010), improved transcript detection and quantitation is expected with the increased technical resolution of RNA-seq over microarray. Supporting this, we find higher correlation between RNA-seq data from *P. vivax* and *P. falciparum* (single replicate sequenced herein for comparative purposes) SPZ datasets (R^2 = 0.42), compared with either species relative to published microarray data (Supplementary Fig. S2 and Supplementary Table S3).

Although microarray supports the high transcription in SPZs of genes such as *uis4*, *csp*, *celtos* and several other *etramps*, 27% and 16% of the most abundant 1% of transcribed genes in our SPZ RNA-seq data are absent from the top decile or quartile, respectively, in the existing *P. vivax* SPZ microarray data (Westenberger et al., 2010). Among these are genes involved in early invasion/hepatocyte development, such as *lsap-1*, *celtos*, *gest* and *siap-1*, or translational repression (e.g., *alba-1* and *alba-4*); orthologs of these genes are also in the top percentile of transcripts in RNA-seq (see Lindner et al., 2013a; Zanghi et al., 2018) and previous microarray data (Le Roch et al., 2004; Mikolajczak et al., 2008) for human-infecting *P. falciparum* and murine-infecting *P. yoelii* SPZs, suggesting many are indeed more abundant than previously characterized. A subset of representative transcripts including *Pv_AP2-X* (PVP01_0733100), *d13*, *gest*, *g10* (PVP01_1011100), 40S ribosomal protein S27 (PVP01_1409300), *puf-2*, *zipco* and 14-3-3 were tested by RT-qPCR for their transcript abundance relative to *celtos* and *sera* (Fig. 1A and Supplementary Table S4). This representative set differed markedly in their relative abundance between our RNAseq and previous microarray data (Westenberger et al., 2010). To control for batch effects introduced by collection of the SPZs used here for RNAseq, this testing was conducted in an additional six sample replicates representing four additional clinical *P. vivax* isolates (PvSPZ-Thai12-15; with PvSPZ-Thai15 tested in technical triplicate). The RT-qPCR results agreed with the RNAseq data for these transcripts (Fig. 1A and Supplementary Table S4).

Our RNA-seq data showed strong qualitative agreement also with available data for *P. falciparum* (Zanghi et al., 2018) and *P. yoelii* SPZs (Lindner et al., 2013a: Supplementary Table S5). Genes highly transcribed in salivary gland SPZs of all three species include *celtos*, *gest*, *trap*, *siap1*, *spect1* and *puf2*. There are 696 *P. vivax* genes shared as orthologs between *P. vivax* P01 and *P. vivax* Sal1 lacking a defined ortholog in *P. falciparum* or *P. yoelli* transcribed at a mean of $\geq$1 TPM in *P. vivax* salivary gland SPZs (Supplementary Table S6). Prominent among these are *vir* (n = 25) and *Pv-fam* (41 fam-e, 16 fam-b, 14 fam-a, 8 fam-d and 3 fam-h) genes and, interestingly, a number of 'merozoite surface protein' 3 and 7 homologs (n = 5 of each). Both *msp3* and *msp7* have undergone significant expansion in *P. vivax* relative to *P. falciparum* and *P. yoelii* (Carlton et al., 2008) and may have repurposed functions in SPZs.

3.1. Plasmodium vivax SPZs transcriptome compared with proteome

We compared relative protein abundance presented in a recently published *P. vivax* SPZ proteome (Swearingen et al., 2017) with relative transcript abundance from the current study (Fig. 1B and Supplementary Table S7). The proteomic study incorporated data from the same PvSPZ-Thai1 and PvSPZ-Thai5 isolates tested by RNA-seq here. We identified 2402 *P. vivax* genes transcribed in the SPZ (CPM > 1) for which no protein expression was detected. Although many of these are lowly transcribed and likely below the detection sensitivity of LC–MS proteomics, others are among the most highly transcribed genes in the SPZ, indicating these may be under translational repression.

Translational repression, the mechanism through which transcripts are held in stasis by RNA binding proteins, has been demonstrated to have important functional roles in the transition of *Plasmodium* spp. between the vertebrate and invertebrate host. More than 700 genes have been identified as translationally repressed in *Plasmodium berghei* ('rodent malaria') gametocytes based on DOZI (DEAD box RNA helicase "development of zygote inhibited") pulldowns (Guerreiro et al., 2014). Translational repression mechanisms mediated through Puf-2 have been explored in SPZs of several *Plasmodium* spp. and regulate some of the most abundant transcripts in the SPZ such as *uis-3* and *uis-4*. UIS3 and UIS4 are the best characterized proteins under translational repression by Puf-2 in SPZs (Silvie et al., 2014b) and are essential for liver stage development (Mueller et al., 2005).

In considering genes that may be translationally repressed (i.e., transcribed but not translated) in the *P. vivax* SPZ, we confine our observations to those transcripts representing the top decile of transcript abundance to ensure their lack of detection as proteins was not due to limitations in the detection sensitivity of the proteomic dataset. Approximately one-third of transcripts in the top decile of transcriptional abundance (n = 170 of 558) in *P. vivax* SPZs were not detectable as peptides in either proteomic replicate (Fig. 1B and Supplementary Table S7). Of these 170 putatively repressed transcripts, 156 and 154 have orthologs in *P. falciparum* and *P. yoelii* respectively, with 89 and 118 of these also not detected as proteins in *P. falciparum* and *P. yoelii* salivary gland SPZs (Lindner et al., 2013b) either, despite being highly transcribed in these stages (see Lindner et al., 2013a; Gomez-Diaz et al., 2017; Supplementary Tables S3–S5), and 133 (78.2%) having no detect-

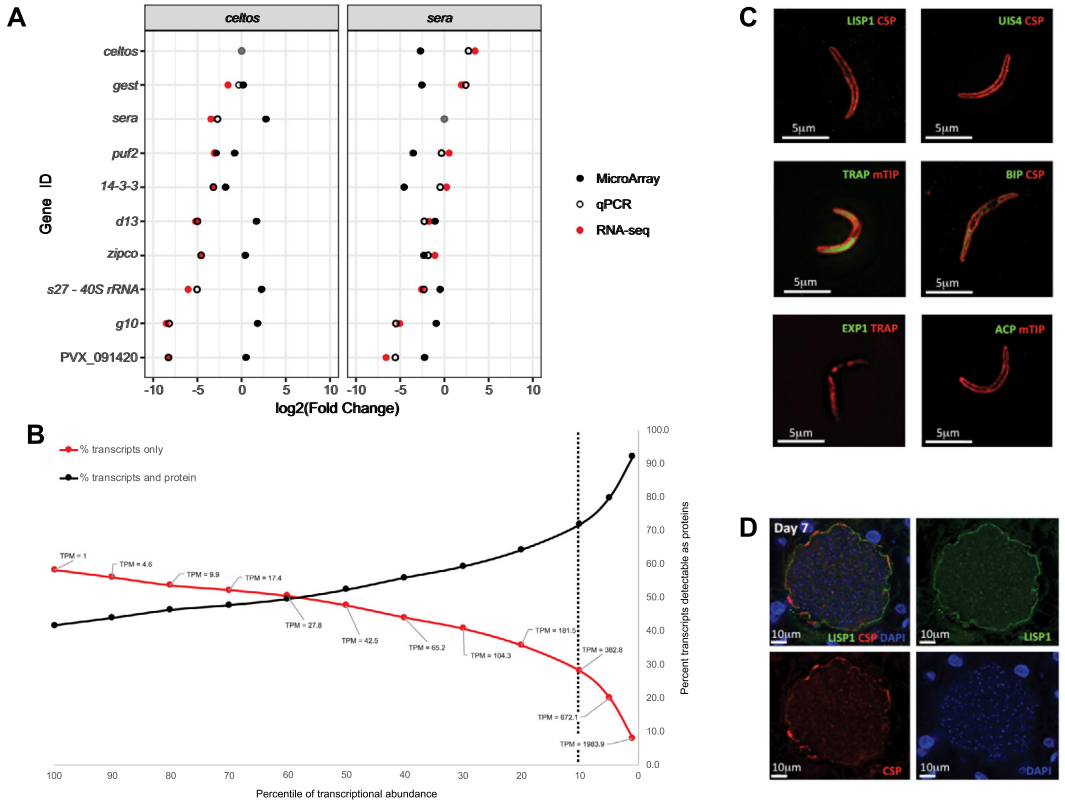

Fig. 1. Transcriptional activity of the *Plasmodium vivax* sporozoite and evidence for translational repression. (A) Relative transcript abundance of key marker genes for sporozoites inferred by RNA-seq and Reverse Transcription – quantitative PCR (RT-qPCR)) relative to previously published microarray data (Westenberger et al., 2010). (B) Relative proportion of genes detectable as transcripts and proteins or transcripts only in RNA-seq and previously published proteomic data. Dashed line shows cut-off used in the current study for putatively repressed transcripts. Immunofluorescent staining of select proteins either known (UIS4) or predicted here (LISP1, EXP1 and ACP) to be translationally repressed in sporozoites in (C) sporozoite stages. CSP, mTIP as known positive controls and TRAP and BIP as experimental positive controls and (D) liver stage (schizonts) at 7 days post-infection in HuHep mice. Liver expression of EXP1 and ACP has been demonstrated by immunofluorescence assay in Mikolajczak et al. (2015), using the same antibodies as used here. ACP, acyl carrier protein; BIP, Immunoglobin-Binding Protein; CSP, circumsporozoite protein; EXP1, Exported Protein 1; LISP1, Liver Stage Protein 1; mTIP, Myosin A Tail Domain Interacting Protein; TRAP, Thrombospondin-reglated Anonymous Protein; UIS4, Up-In-Sporozoite Protein 4.

able SPZ expression (>1 unique peptide count) in LC–MS data deposited for any species in PlasmoDB (Supplementary Table S8). In contrast, 106 of these putatively repressed transcripts with orthologs in other *Plasmodium* spp. (Supplementary Table S8) for which proteomic data are available in PlasmoDB, are detectable (>1 unique peptide count) by LC–MS methods in at least one other life-cycle stage, indicating against a technical issue (e.g., inability to be trypsin-digested) preventing their detection in the *P. vivax* SPZ proteome (Swearingen et al., 2017). In addition to *uis3* and *uis4*, genes involved in liver stage development and detectable as transcripts but not proteins in the *P. vivax* SPZs include *lsap1* (liver stage associated protein 1), *zipco* (ZIP domain-containing protein), several other *etramps* (PVP01_1271000, PVP01_0422600, PVP01_0504800 and PVP01_0734800), *pv1* (parasitophorous vacuole protein 1) and *lisp1* and *lisp2* (PVP01_1330800 and PVP01_0304700). Also notable among genes detectable as transcripts but not proteins in SPZs is a putative 'Yippee' homolog (PVP01_0724100). Yippee is a DNA-binding protein that, in humans (YPEL3), suppresses cell growth (Kelley et al., 2010) and is regulated through histone acetylation (Tuttle et al., 2011), making it noteworthy in the context of *P. vivax* HPZ developmental arrest.

Although verifying each putatively repressed transcript will require further empirical data, our system level approach is supported by immunofluorescent microscopy (Fig. 1C and 1D) of UIS4, LISP1, EXP1 and ACP (PVP01_0416300). These represent one known and three putative (i.e., newly proposed here) translationally repressed genes in *P. vivax* SPZs, and are compared with TRAP and BiP (which are both transcribed and expressed as protein in the *P. vivax* SPZ; Supplementary Table S8). The *lisp1* gene is an interesting find. In *P. berghei*, *lisp1* is essential for rupture of the parasitophorus vacuolar membrane (PVM) during liver stage development, allowing release of the merozoite into the host blood stream. *Pv-lisp1* is ~350-fold and ~1350-fold more highly transcribed in *P. vivax* SPZs compared with SPZs of either *P. falciparum* or *P. yoelii* (see Supplementary Table S5). IFAs using LISP1-specific monoclonal antibodies (Fig. 1C) show that this protein is undetectable in SPZs but clearly expressed at 7 days p.i. in liver schizonts (Fig. 1D).

3.2. Transcription in *P. vivax* SPZs relative to other life-cycle stages

3.2.1. *Plasmodium vivax* SPZ relative to blood-stage transcriptome

To identify transcripts up-regulated in SPZs, we first compared the *P. vivax* SPZ transcriptome with RNA-seq data for *P. vivax* blood stages (Zhu et al., 2016) (Fig. 2A and 2B and Supplementary Figs. S7-S10). We identified 1672 up- and 1958 down-regulated transcripts (FDR $\leq$ 0.05; minimum two-fold change in CPM; Supplementary Table S9; Interactive Glimma Plot – Supplementary Data S1) and explored patterns among these differentially transcribed genes (DTGs) by protein family (Fig. 2C) and GO classifications (Supplementary Table S10).

RNA recognition motifs (RRM-1 and RRM-6) and helicase domains (Helicase-C and DEAD box helicases) are over-represented ($P < 0.05$) among transcripts up-regulated in SPZs

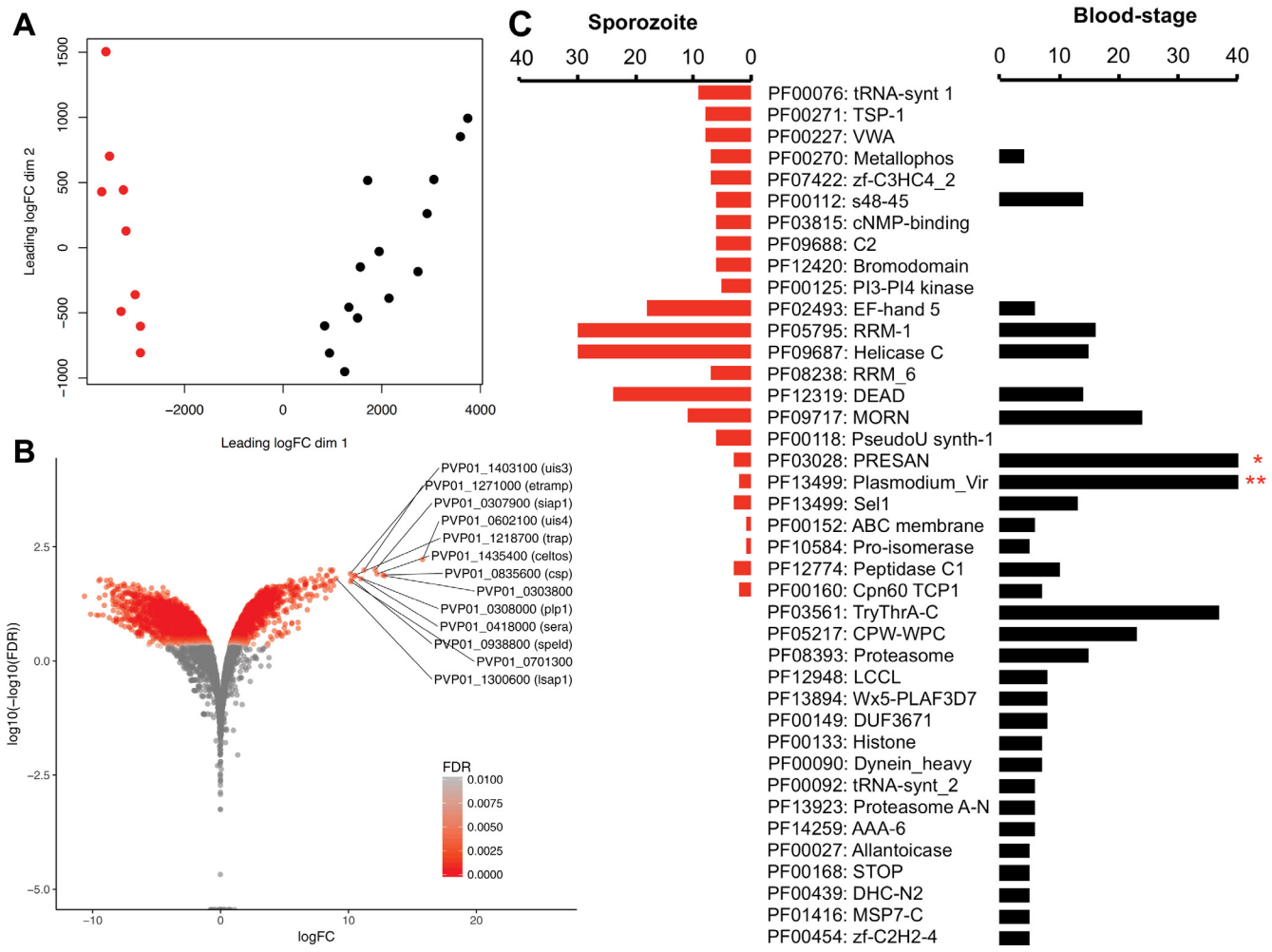

Fig. 2. Differential transcription between *Plasmodium vivax* salivary gland sporozoites and blood stages. (A) Biological Coefficient of Variation plot showing separation between blood-stage (black) and salivary gland sporozoite (red) biological replicates. (B) Volcano plot of distribution of fold-changes (FC) in transcription between blood stages and salivary gland sporozoites relative to statistical significance threshold (False Discovery Rate (FDR) ≤ 0.05). Positive FC represents up-regulated transcription in the sporozoite stage. (C) Mirror plot showing pFam domains statistically significantly (FDR ≤ 0.05) over-represented in salivary gland sporozoite up-regulated (red) or blood-stage up-regulated (black) transcripts. Presan and Vir data bars truncated in blood-stage data for presentation. * – 55 PRESAN domains are in this dataset. ** – 99 Vir domains are in this dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 2C), consistent with translational repression through ribonucleoprotein (RNP) granules (Kramer, 2014). In *Plasmodium*, translational repression regulates key life-cycle transitions coinciding with switching between the mosquito and the mammalian host (either as SPZs or gametocytes) (Kramer, 2014). For example, although *uis4* is the most abundant transcript in the infectious SPZ ((Mikolajczak et al., 2008; Westenberger et al., 2010); Supplementary Table S2), UIS4 is translationally repressed in this stage (Silvie et al., 2014a) and only expressed after hepatocyte invasion (Aly et al., 2011). In SPZs, PUF2 binds to specific mRNA transcripts and prevents their translation (Lindner et al., 2013a), and SAP1 stabilises the repressed transcripts and prevents their degradation (Aly et al., 2011). Consistent with this, *Puf2* and *sap1* (PVP01_0947600) are up-regulated in the SPZ relative to blood-stages. Indeed, *Puf2* (PVP01_0526500) is among the top percentile of transcripts in salivary SPZs and expressed at high levels in the proteome (Swearingen et al., 2017). Our data implicate other known translational repressors (Kramer, 2014) that may act in *P. vivax* SPZs. Among these are *alba-2* and *alba-4*, both of which are among the top 2% of genes transcribed in SPZs and ~14 to 20-fold more highly transcribed in SPZs relative to blood stages; ALBA-2 is in the top 100 most abundant proteins in the *P. vivax* SPZ proteome (Swearingen et al., 2017). In addition, *P. vivax* SPZs are enriched for HoMu (homolog of Musashi; top decile of SPZ proteins by abundance (Swearingen et al., 2017)) and ptbp (polypyrimidine tract binding protein). In *Drosphila*, Musashi regulates eukaryotic stem cell differentiation through translational repression (Okano et al., 2002) and HoMu localizes with DOZI and CITH in *Plasmodium* gametocytes (Cui et al., 2015). PTBP is linked to mRNA stability, splice regulation and translational initiation (Lasko, 2003) and may perform a complementary role to SAP1.

Transcripts encoding nucleic acid binding domains such as bromodomains (PF00439; which can also bind lysine-acetylated proteins), zinc fingers (PF13923) and EF hand domains (PF13499) are also enriched in SPZs. Included among these proteins are a putative ApiAP2 transcription factor (PVP01_1211900) and a homologue of the *Drosophila* zinc-binding protein 'Yippee' (PVP01_0724100). Thrombospondin-1 like repeats (TSR: PF00090) and von Willebrand factor type A domains (PF00092) are enriched in SPZs as well. In *P. falciparum* SPZs, genes enriched in TSR domains are important in invasion of the mosquito salivary gland (e.g., *trap*) and secretory vesicles released by SPZs upon entering the vertebrate host (e.g., *csp*) (Tucker, 2004). Notably, the *P. cynomolgi* ortholog of PVP01_0207300, PCYB_021650, is

transcriptionally up-regulated in HPZs relative to replicating schizonts (Cubi et al., 2017), indicating a target of significant interest when considering HPZ formation and/or biology, and suggesting that the list here may contain other genes important in HPZ biology.

3.2.2. *Plasmodium vivax* SPZ relative to *Plasmodium* spp. liver stage transcriptomes

New advances in *P. vivax* liver culture has allowed recent publication of mixed stage and HPZ-enriched transcriptomes (Gural et al., 2018). This is an early, yet highly valuable, study and, due no doubt to the difficulty in generating the material, is limited to biological duplicates. Noting this, although we undertake differential transcriptomic studies of this dataset here, we recognize that additional biological replication is needed and have used a higher burden of significance (FDR $\leq$ 0.01 and $\geq$2-fold change) than used with blood-stages. Nevertheless, these comparisons identified 1015 and 856 SPZ up-regulated transcripts relative to mixed liver-stages (mLS) and HPZ-enriched liver stages (HPZs), respectively, and 1007 and 1079 transcripts up-regulated in mLS and HPZs relative to SPZs, respectively (Fig. 3, Supplementary Figs. S11–S13 and Supplementary Tables S11 and S12 and Interactive Glimma Plot – Supplementary Data S1).

Compared with mLS transcriptomes, SPZs are enriched Pfam domains (Fig. 3A) associated with transcriptional regulation (PF00176, PF01096, PF01661 and PF08711), translational repression/regulation (PF00076, PF00279, PF01008, PF01873, PF01917, PF02847, PF02854, PF13893 and PF14259), DNA/RNA binding (PF0097, PF13445, PF13639, PF14570 and PF15247) and chromatin regulation (PF00271, PF00850 and PF13489) are over-represented in SPZ (SPZ) compared with mLS stages (FDR $\leq$ 0.05). These include many of the transcripts similarly up-regulated in comparison to blood stages (e.g., *uis4*, *celtos*, *puf2*, *siap1* and *plp-1*). In contrast, the mLS transcriptome is enriched for Pfam domains associated with cell division (PF00493), merozoite formation (PF07133 and PF12984), proteasome function (PF00227, PF00400, PF00656, PF01344, PF01398 and PF03981), protein export / vesicle function (PF00350 and PF00996), membrane proteins (PF01105, PF03011,

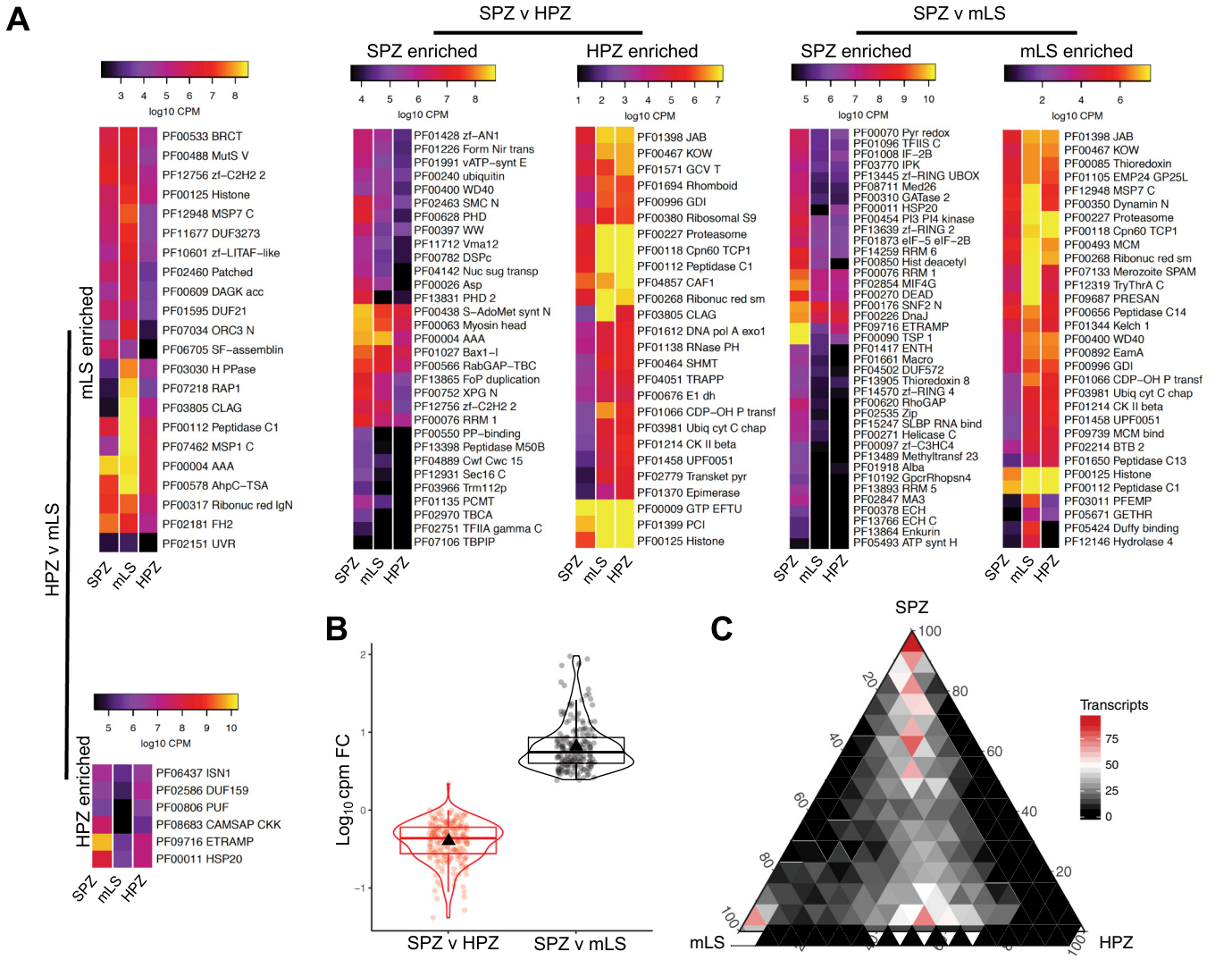

Fig. 3. Differential transcription between *Plasmodium vivax* sporozoites (SPZ), mixed (mLS) and hypnozoite (HPZ) enriched liver stages (liver-stage data from Gural et al. (2018)). (A) Heatmap comparisons showing summed transcription of enriched Pfam domains in HPZ versus mLS, SPZ versus HPZ and SPZ versus mLS comparisons. All Pfam domains were statistically significantly enriched at *P*-value 0.05). All transcript data for stage up-regulated genes at False-Discovery Rate (FDR) 0.01). (B) Violin box-plot showing relative fold-change differences between SPZ and HPZ compared with SPZ and mLS for genes down-regulated in mLS compared with SPZ, but not down-regulated in HPZ compared with SPZ. (C) Ternary heatmap summarizing relative transcript abundance in each of SPZ, mLS and HPZ stages.

PF05424 and PF12139) and metabolism (PF00085, PF00118, PF00268, PF01066, PF01214 and PF01214). This includes genes involved in replication and merozoite formation ($n = 14$; including PVP01_0728900 (*msp1*), PVP01_0010670 (*msp3*) and PVP01_1446800 (*msp9*)), rhoptry function ($n = 9$; including PVP01_1469200 (*rnp3*), PVP01_1255000 (*rnp2*) and PVP01_1338500 (*rap1*)) and reticulocyte binding ($n = 10$ including PVP01_0534300 (rbp2c), PVP01_1402400 (rbp2a), PVP01_0701100 (rbp1b) and PVP01_0800700 (rbp2b)). Collectively, these data point towards the SPZ stage as being highly regulated and controlled at transcriptional, translational and chromatin levels, with the mLS stages directed toward replication, protein turn-over, reconfiguration of the proteins on the plasma membrane and metabolic activity.

Comparison of SPZs with HPZs does not indicate a similar magnitudinal change. The SPZ is enriched, relative to HPZs, in genes such as PVP01_1258000 (*gest*), PVP01_0418000 (*sera*), PVP01_1435400 (*celtos*), PVP01_0835600 (*csp*) and PVP01_0602100 (*uis4*). At a broad level, SPZ-enriched Pfam domains include a smaller number associated with translational repression/regulation (PF00076) or DNA/RNA binding (PF01428 and PF12756). Interestingly, SPZs are enriched in Pfam domains specifically associated with heterochromatin (H3K9me3) reading/interaction (PF02463, PF00628, PF13831 and PF13865). In contrast, HPZs were enriched, relative to SPZs, for genes including histone proteins (PVP01_1138700, PVP01_1131700 and PVP01_0905900), classic markers of metabolism (PVP01_MITO3300 and PVP01_MITO3400) and *lisp2*. Pfam data indicated largely similar domain enrichment trends as were seen for the mLS stage relative to SPZs, including a number of proteosomal (PF00227, PF00112, PF03981), vesicular transport (PF00996) and metabolic (PF00118, PF00268, PF01066, PF01214 a) associated functions. This supports HPZs being an arrested, rather than classically 'dormant', stage with active metabolism and protein turnover. HPZs are also enriched for Pfams associated with mRNA/tRNA regulation and turnover (PF04857, PF01612, PF00009 and PF01138) and glycine metabolism (PF01571 and PF00464) and acetyl-CoA production (PF02779 and PF00676).

Finally, although not the focus of this study, we looked at differential transcription between mLS and HPZ stages using the Gural et al. (2018) data, but using the same approaches employed here. In particular, we were interested in what insights these comparisons might provide in terms of SPZ differentiation or development into liver schizonts or HPZs (Supplementary Table S13). Among mLS up-regulated transcripts are genes associated with rhoptry function ($n = 11$; including PVP01_0107500, PVP01_1469200 and PVP01_1469200), cytoadherence to red-cells (PVP01_1401400 and PVP01_0734500), merozoite formation (PVP01_0728900 and PVP01_0612400) and exported proteins ($n = 6$; including PVP01_0504000, PVP01_0119200 and PVP01_0801600). Consistent with *P. cynomolgi* (Cubi et al., 2017), HPZ up-regulated transcripts include several key SPZ transcripts, specifically *uis4* (PVP01_0602100), *puf1* (PVP01_1015000) and *speld* (PVP01_0938800). At the Pfam domain level, mLS is enriched for metabolic (PF00317) and proteosomal (PF00112) domains also enriched in mLS or HPZs relative to SPZs above, as well as domains associated with merozoite formation (PF12948, PF07462), rhoptry function (PF0712), DNA/RNA binding (PF12756, PF10601 and PF02151) and cell division, development and DNA replication (PF06705, PF00533, PF00488, PF02460, PF07034, PF02181). In contrast, HPZs are enriched in Pfam domains that overlap notably with key SPZ transcripts including *etramps* (PF09716) and *puf* proteins (PF00806), as well as domains associated with calcium (PF08683) and nucleotide metabolism (PF06437). These data largely indicate that the HPZ bears similarity both to the SPZ and liver schizonts consistent with a stalled stage on the path to schizont development regulated by checkpoint signals that halt/restart normal schizont development, which has been proposed previously for this species (Westenberger et al., 2010).

With this is mind, we looked at transcripts that are differentially transcribed in mLS, but not HPZs, relative to SPZs. There are 107 transcripts down-regulated in mLS relative to SPZs that are transcribed at approximately similar levels in both SPZs and HPZs (Fig. 3B and C). A common theme among many of these genes are their role in transcriptional, post-transcriptional, translational or post-translational regulation. Among transcriptional regulators are transcription factors including AP2-SP2 (PVP01_0303400) and three non-AP2-like transcription factors (PVP01_0306600, PVP01_0204300 and PVP01_1415800). Post-transcriptional controllers include several DNA/RNA-binding proteins (PVP01_1011000. PVP01_0932900, PVP01_0715300, PVP_1242600 and PVP01_0605200), RNA helicases (PVP01_1403600 and PVP01_1329800) and mRNA processing (PVP01_1443100 and PVP01_1458200) genes. Translational control includes several key regulators of translation initiation (PVP01_1467700), tRNA processing (PVP01_0318700 and PVP01_1017700) or ribosomal function/biogenesis (PVP01_1443700, PVP01_0421400, PVP01_1117200 and PVP01_0215100). Post-translational control includes two methyltransferases (PVP01_1428800 and PVP01_1465200), including CARM1, which methylates of H3R17 and, in mice, prevents differentiation in embryonic stem cells (Wu et al., 2009), and a putative histone methylation reading enzyme, EEML2 (PVP01_1014100). The remaining genes in this group have three noteworthy and largely overlapping themes: (i) an association with calcium binding, metabolism or signalling, (ii) a role in organellar metabolism and (iii) homologs in other organisms, including a variety of prokaryotes and eukaryotes, with key roles in germination, dormancy and persistent non-replicating stages. The latter-most function is clearly intriguing in the context of HPZ formation and activation. These genes include a homolog of dihydrolipoamide acyltransferase (aka 'sucB'), which is essential for growth in *Mycobacterium tuberculosis* (Shi and Ehrt, 2006) and a key regulator in persistent *Escherichia coli* stages (Ma et al., 2010). Another example is gamete fusion factor HAP2, which, despite the name, has been shown to regulate dormancy in eukaryotes ranging from plants (Schrader et al., 2004; Yazawa and Kamada, 2007) to amoebae (Wood et al., 2017).

In addition to data for *P. vivax*, two transcriptomic studies are now available for *P. cynomolgi* (Voorberg-van der Wel et al., 2017; Zanghi et al., 2018) that compare mixed/schizont stage parasites with small-form "HPZs". Cubi et al. (2017) noted an ApiAP2 (dubbed "AP2-Q"; PCYB_102390) as transcriptionally up-regulated in *P. cynomolgi* HPZs and proposed this as a potential HPZ marker. AP2s also feature among transcribed genes in *P. vivax* liver stages, with one, PVP01_0916300, significantly up-regulated in HPZs compared with mLS. Intriguingly, PVP01_0916300 is among the top quartile of *P. vivax* SPZ transripts (TPM = 104) and enriched relative to blood stages. The *P. vivax* ortholog of *Pc*-AP2-Q (PVP01_1016100) is among the genes detectable as a transcript but not protein in *P. vivax* SPZs. This may point to a translationally repressed signal in SPZs to regulate HPZ formation. However, as *Pv*-AP2-Q is transcribed at an abundance (~50 TPM) at or below which ~50% of *P. vivax* genes are detectable as transcripts but not as proteins (Fig. 1B), this could as likely result from LC–MS detection sensitivity. Further, although AP2-Q was reported as specific to HPZ forming *Plasmodium* spp. (Cubi et al., 2017), it is indeed found in non-HPZ producing species such as *Plasmodium knowlesi*, *Plasmodium gallinaceum* and *Plasmodium inui* (Voorberg-van der Wel et al., 2017). Up-regulation of AP2-Q transcripts is not observed for HPZs in subsequent transcriptomic studies of *P. cynomolgi* (Voorberg-van der Wel et al., 2017) or *P. vivax* (Gural et al., 2018), nor did we see such an up-regulation here.

Voorberg van der Wel et al. (2017) noted transcription of a range of AP2s in *P. cynomolgi* liver stages, but did not find any to be up-regulated in HPZs.

3.3. Chromatin epigenetics in P. vivax SPZs

As noted above, transcriptomic data for SPZs, and their comparison with liver and blood stages, implicate histone epigenetics in SPZ biology and liver stage differentiation. This concept has been alluded to in recent liver-stage studies of *P. cynomolgi* (Dembele et al., 2014; Cubi et al., 2017) that propose methyltransferases as having a potential role in HPZ formation. This is intriguing but requires detailed study in *P. vivax*. It is presently not possible to generate sufficient *P. vivax* liver stage material to support ChIP-seq studies. Here, we take a first step in this process and characterize major histone marks in *P. vivax* SPZs and explore their relationship to transcript abundance.

3.3.1. Histone modifications in P. vivax SPZs

We identified 1506, 1999 and 5262 ChIP-seq peaks stably represented in multiple *P. vivax* SPZ replicates and associated with H3K9me3, H3K4me3 and H3K4me3 histone marks, respectively (Fig. 4 and Supplementary Figs. S14-S19). Consistent with observations in *P. falciparum* SPZs (Gomez-Diaz et al., 2017; Zanghi et al., 2018), H3K9me3 'heterochromatin' marks primarily clustered in telomeric and subtelometric regions (Fig. 4) and H3K4me3 and H3K9ac marks were distributed broadly across the chromosomes outside of the telomeric regions and did not overlap with regions under H3K9me3 suppression.

3.3.2. Genes under histone regulation

We explored an association between these histone marks and the transcriptional behaviour of protein coding genes (Fig. 4 and Supplementary Fig. S19 and Supplementary Tables S14–S19). Four hundred and eighty-five coding genes stably intersected with an H3K9me3 mark; all were located near the ends of the chromosomal scaffolds (i.e., are (sub)telomeric). On average, these genes were transcribed at ~30 fold lower levels (mean 0.7 TPM) than genes not stably intersected by H3K9me3 marks. These data clearly support the function of this mark in transcriptional silencing. This is largely consistent with observations in *P. falciparum* SPZs (Gomez-Diaz et al., 2017), however, in contrast to *P. falciparum* SPZs (Zanghi et al., 2018), we observed no genes within the heterochromatin dense region that lacked a stable H3K9me3 signal or were transcribed at notable levels (i.e., above ~5 TPM). Whether this relates to differences in epigenetic control between the species is not clear. We note that (sub)telomeric genes are overall transcriptionally silent in *P. vivax* SPZs relative to blood stages (Fig. 4 and Supplementary Tables S20 and S21). Consistent with observations in *P. falciparum* (Lopez-Rubio et al., 2009), the bulk of these genes include complex protein families such as *vir* and *Pv-fam* genes, which are described to function primarily in blood stages. Also notable among the genes are several reticulocyte-binding proteins including RBP2, 2a, 2b and 2c; *Pv*-RBP2b meditates reticulocyte invasion in *P. vivax* blood stages (Gruszczyk et al., 2018).

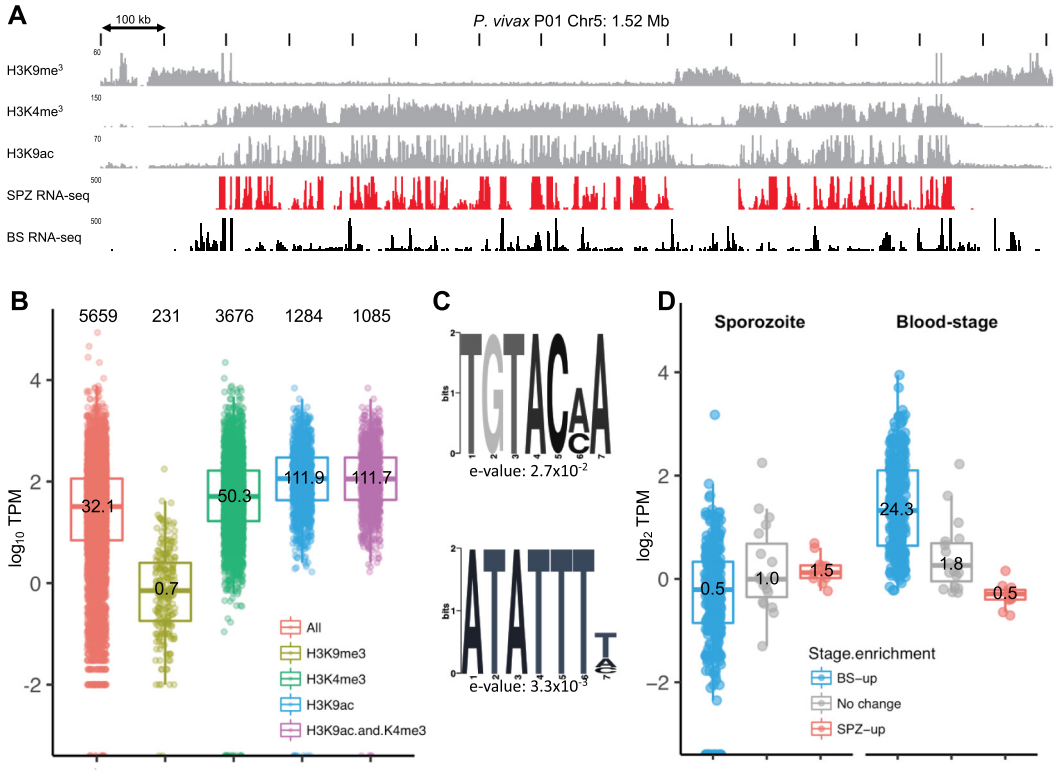

Fig. 4. Histone epigenetics relative to transcriptional behaviour in salivary gland *Plasmodium vivax* sporozoites. (A) Representative H3K9me3, H3K4me3 and H3K9ac ChIP-seq data (grey) from a representative chromosome (*P. vivax* P01 Chr5) relative to mRNA transcription in salivary gland sporozoites (black) and blood stages (black). Small numbers to top left of each row show the data ranges. (B) Salivary gland sporozoite transcription relative to the nearest stable histone epigenetic marks. Numbers at the top of the figure represent total genes included in each category. Numbers within the box plot represent the mean transcription in transcripts per million (TPM). (C) Sequence motifs enriched within 1 kb upstream of the inferred Transcription Start Site of highly transcribed (top 10%) relative to lowly transcribed genes associated with H3K9ac marks in salivary gland sporozoites. (D) Relative transcription of (sub)telomeric genes in *P. vivax* salivary gland sporozoites and blood stages categorized by gene sets up-regulated in blood stages (BS, blue), salivary sporozoites (SPZ, red) or not stage enriched (grey). Numbers in each box show mean transcription in TPM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Outside of the telomeres and subtelomeres, H3K4me3 marks are stably associated with the intergenic regions of 3676 genes. Stable H3K9ac marks were identified within 1 kb of the transcriptional start site (TSS) of 1284 coding genes, with 1085 of these also stably marked by H3K4me3 (Fig. 4B). The average transcription of these genes is 50, 112 and 112 TPMs respectively (72, 160 and 160-fold higher than H3K9me3 marked genes). Gene-by-gene observations show that H3K9ac and H3K4me3 marks cluster densely in the 1000 kb up- and down-stream of the start and stop codons, respectively, of transcribed genes, but are much less dense within coding regions of these genes (Supplementary Fig. S15). This pattern directly correlates with transcription and contrasts H3K9me3 marks, which are distributed across the length of the gene at even density and are correlated with a lack of transcription. These data support the role of H3K9ac and H3K4me3 in transcriptional activation in *P. vivax*. The lower transcriptional abundance of H3K4me3 marked, compared with H3K9ac or H3K9ac and H3K4me3 marked, genes suggest these marks work synergistically and that H3K9ac is possibly the better of the two, as a single mark indicator of transcriptional activity in *P. vivax*. This is consistent with recent observations in *P. falciparum* SPZs (Gomez-Diaz et al., 2017).

Interestingly, H3K9ac-marked genes ranged in transcriptional activity from the most abundantly transcribed genes to many in the lower 50% and even lowest decile of transcription. This suggests more contributes to transcriptional activation in *P. vivax* SPZs than, simply, gene accessibility through chromatin regulation. Specific activation by a transcription factor (e.g., ApiAP2s; De Silva et al., 2008) is the obvious candidate. To explore this, we compared upstream regions (within 1 kb of the inferred TSS or up to the 3′ end of the next gene upstream, whichever was less) of highly (top 10%) and lowly (bottom 10%) transcribed H3K9ac marked genes for over-represented sequence motifs in the highly expressed genes that might coincide with known ApiAP2 transcription factor binding sites (Painter et al., 2011). We identified these based on the location of the nearest stable H3K9ac peak relative to the transcription start site for each gene (Supplementary Fig. S12). In most instances, these peaks were within 100 bp of the TSS and, consistent with data from *P. falciparum* (Cui et al., 2007), *P. vivax* promoters appear to be no more than a few hundred to a maximum of 1000 bp upstream of the TSS. Exploring these regions, we identified two over-represented motifs: TGTACMA (e-value $2.7e^{-2}$) and ATATTTH (e-value $3.3e^{-3}$) (Fig. 4C). TGTAC is consistent with the known binding site for *Pf*-AP2-G, which regulates sexual differentiation in gametocytes (Kafsack et al., 2014), but its *P. vivax* ortholog (PVP01_1418100) is neither highly transcribed nor expressed in SPZs. ATATTTH is similar to the binding motif for *Pf*-AP2-L (AATTTCC), a transcription factor that is important for liver stage development in *P. berghei* (Iwanaga et al., 2012). In contrast to AP2-G, *Pv*-AP2-L (PVX_081180) is in the top 10% of transcription and expression in *P. vivax* SPZs and up-regulated relative to blood stages. In *P. vivax* SPZs, the ATATTTH motif is associated with a number of highly transcribed genes, including *lisp1* and *uis2-4*, known to be regulated by AP2-L in *P. berghei* (Iwanaga et al., 2012) as well as many of the most highly transcribed, H3K9ac marked genes including two *etramps* (PVP01_0734800 and PVP01_0504800), several RNA-binding proteins including *Puf2*, *ddx5*, a putative ATP-dependent RNA helicase DBP1 (PVP01_1429700), and a putative *bax1* inhibitor (PVP01_1465600). Interestingly, a number of highly transcribed and translationally repressed genes associated with the ATATTTH motif, including *uis4*, *siap2* and *pv1*, are not stably marked by H3K9ac in all replicates (i.e., there is significant variation in the placement of the H3K9ac peak or their presence/absence among replicates for these genes). It may be that additional histone modifications, for example H3K27me, H3R17me3 or H2A or H4 modifications, are involved in regulating transcription of these genes. Certainly the incorporation of the H2A.Z histone variant, which is present in intergenic regions of *P. falciparum* (Petter et al., 2011), and controls temperature responses in plants (Boden et al., 2013), is intriguing as a potential mark regulating SPZ fate in

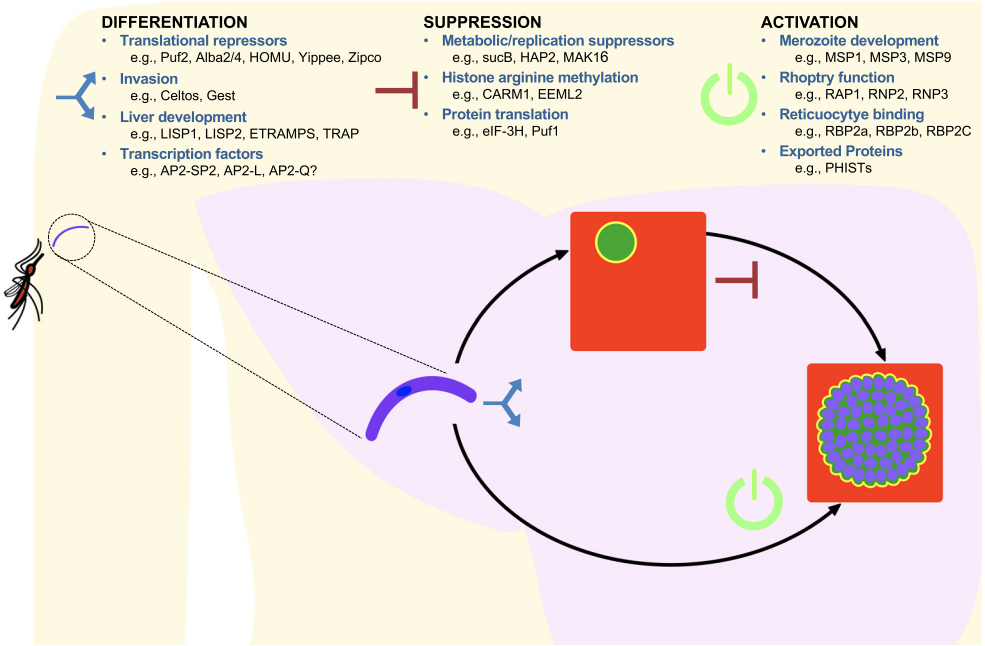

Fig. 5. Schematic of potential mechanisms underpinning development in differentiation of *Plasmodium vivax* sporozoites during liver-stage infection as hypnozoites and schizonts. We suggest differentiation programming at different points in development; first, schizont or hypnozoite fate possibly encoded in the sporozoite as epigenetic signals or translationally repressed transcripts; secondly, suppression signals that halt progression of the hypnozoite to schizont stage and support persistence; and finally activation signals signified by a release in chromatin, (post)transcriptional and (post)translational control leading to up-regulation of replication, metabolic and protein export pathways.

P. vivax considering the association between HPZ activation rate and climate (White, 2011), as is H3R17me3 in consideration of the enrichment of markers/readers of this modification in HPZs noted above and the role of this mark in cell fate progression in other species (Wu et al., 2009).

3.4. Conclusions

We provide a comprehensive study of the transcriptome and epigenome of mature *Plasmodium vivax* SPZs and undertake detailed comparisons with recently published proteomic data for *P. vivax* SPZs (Swearingen et al., 2017) and transcriptomic data for *P. vivax* mixed and HPZ-enriched liver-stages (Gural et al., 2018) and mixed blood-stages (Zhu et al., 2016). These data support the proposal that the SPZ is a highly programmed stage that is primed for invasion of, and development in, the host hepatocyte. Cellular regulation including at transcriptional, translational and epigenetic levels, appears to play a major role in shaping this stage (which continues on in some form in HPZs), and many of the genes proposed here as being under translational repression are involved in hepatocyte infection and early liver stage development (Fig. 5). We highlight a major role for RNA-binding proteins, including PUF2, ALBA2/4 and, intriguingly, 'Homologue of Musashi' (HoMu). We find that transcriptionally, the HPZ appears to be a transition point between the SPZ and replicating schizonts, having many of the dominant SPZ transcripts and retaining high transcription of a number of key regulatory pathways involved in transcription, translation and chromatin configuration (including histone arginine methylation). A consistent theme in the study is the prominence of a number of genes that have a role in cell fate determination and differentiation (e.g., HoMu, Yippee and CARM1) and overlap with dormancy and/or persistent cell states in bacteria, protists or higher eukaryotes (e.g., bacterial sucB and gamete fusion protein HAP2). These data do not point to one single programming switch for dormancy or liver development in *P. vivax*, but present a number of intriguing avenues for exploration in subsequent studies, particularly in model species such as *P. cynomolgi*. Our study contributes to understanding the early stages of hepatocyte infection and the developmental switch between liver trophozoite and HPZ formation. We also identify potential avenues for rationally prioritizing targets underpinning liver stage differentiation for functional evaluation in humanized mouse and simian models for relapsing *Plasmodium* spp. and identifying novel avenues to understand and eradicate liver-stage infections.

Acknowledgements

The authors acknowledge funding from the National Health and Medical Research Council (NHMRC, Australia; APP1021544, 1043345 and 1092789), the Australian Research Council (ARC), the Victorian State Government Operational Infrastructure Support, Australia and Australian Government National Health and Medical Research Council Independent Research Institute Infrastructure Support Scheme, the Ian Potter Foundation, Australia the National Institute of Health, USA, the Bill and Melinda Gates Foundation, USA, the US Department of Defense (W81XWH-15-1-0249) and the Office of the Assistant Secretary of Defence for Health Affairs (USA) through the Peer Reviewed Medical Research Program (PRMRP).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2019.02.007.

References

Aly, A.S., Lindner, S.E., MacKellar, D.C., Peng, X., Kappe, S.H., 2011. SAP1 is a critical post-transcriptional regulator of infectivity in malaria parasite sporozoite stages. Mol. Microbiol. 79, 929–939.
Auburn, S., Bohme, U., Steinbiss, S., Trimarsanto, H., Hostetler, J., Sanders, M., Gao, Q., Nosten, F., Newbold, C.I., Berriman, M., Price, R.N., Otto, T.D., 2016. A new *Plasmodium vivax* reference sequence with improved assembly of the subtelomeres reveals an abundance of pir genes. Wellcome Open Res. 1, 4.
Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics (Oxford, England) 27, 1653–1659.
Baird, K.J., 2013. Malaria caused by *Plasmodium vivax*: recurrent, difficult to treat, disabling, and threatening to life - averting the infectious bite preempts these hazards. Pathogens Global Health 107, 475–479.
Boden, S.A., Kavanova, M., Finnegan, E.J., Wigge, P.A., 2013. Thermal stress effects on grain yield in *Brachypodium distachyon* occur via H2A.Z-nucleosomes. Genome Biol. 14, R65.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30, 2114–2120.
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., Crabb, B.S., Del Portillo, H.A., Essien, K., Feldblyum, T.V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang'a, S., Kooij, T.W., Korsinczky, M., Meyer, E.V., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stoeckert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W., Fraser-Liggett, C.M., 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 455, 757–763.
Cubi, R., Vembar, S.S., Biton, A., Franetich, J.F., Bordessoulles, M., Sossau, D., Zanghi, G., Bosson-Vanga, H., Benard, M., Moreno, A., Dereuddre-Bosquet, N., Le Grand, R., Scherf, A., Mazier, D., 2017. Laser capture microdissection enables transcriptomic analysis of dividing and quiescent liver stages of *Plasmodium* relapsing species. Cell. Microbiol. 19, e12735.
Cui, L., Lindner, S., Miao, J., 2015. Translational regulation during stage transitions in malaria parasites. Ann. New York Acad. Sci. 1342, 1–9.
Cui, L., Miao, J., Furuya, T., Li, X., Su, X.Z., Cui, L., 2007. PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. Eukaryot. Cell 6, 1219–1227.
De Silva, E.K., Gehrke, A.R., Olszewski, K., León, I., Chahal, J.S., Bulyk, M.L., Llinás, M., 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. Proc. Natl. Acad. Sci. 105, 8393–8398.
Dembele, L., Franetich, J.F., Lorthiois, A., Gego, A., Zeeman, A.M., Kocken, C.H., Le Grand, R., Dereuddre-Bosquet, N., van Gemert, G.J., Sauerwein, R., Vaillant, J.C., Hannoun, L., Fuchter, M.J., Diagana, T.T., Malmquist, N.A., Scherf, A., Snounou, G., Mazier, D., 2014. Persistence and activation of malaria hypnozoites in long-term primary hepatocyte cultures. Nat. Med. 20, 307–312.
Duffy, M.F., Selvarajah, S.A., Josling, G.A., Petter, M., 2014. Epigenetic regulation of the *Plasmodium falciparum* genome. Brief Funct. Genomics 13, 203–216.
Feachem, R.G., Phillips, A.A., Hwang, J., Cotter, C., Wielgosz, B., Greenwood, B.M., Sabot, O., Rodriguez, M.H., Abeyasinghe, R.R., Ghebreyesus, T.A., Snow, R.W., 2010. Shrinking the malaria map: progress and prospects. Lancet 376, 1566–1578.
Gomez-Diaz, E., Yerbanga, R.S., Lefevre, T., Cohuet, A., Rowley, M.J., Ouedraogo, J.B., Corces, V.G., 2017. Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in *Anopheles gambiae*. Sci. Rep. 7, 40655.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29, 644.
Gruszczyk, J., Kanjee, U., Chan, L.J., Menant, S., Malleret, B., Lim, N.T.Y., Schmidt, C.Q., Mok, Y.F., Lin, K.M., Pearson, R.D., Rangel, G., Smith, B.J., Call, M.J., Weekes, M.P., Griffin, M.D.W., Murphy, J.M., Abraham, J., Sriprawat, K., Menezes, M.J., Ferreira, M.U., Russell, B., Renia, L., Duraisingh, M.T., Tham, W.H., 2018. Transferrin receptor 1 is a reticulocyte-specific receptor for *Plasmodium vivax*. Science 359, 48–55.
Guerreiro, A., Deligianni, E., Santos, J.M., Silva, P.A., Louis, C., Pain, A., Janse, C.J., Franke-Fayard, B., Carret, C.K., Siden-Kiamos, I., Mair, G.R., 2014. Genome-wide RIP-Chip analysis of translational repressor-bound mRNAs in the *Plasmodium* gametocyte. Genome Biol. 15, 493.
Gural, N., Mancio-Silva, L., Miller, A.B., Galstian, A., Butty, V.L., Levine, S.S., Patrapuvich, R., Desai, S.P., Mikolajczak, S.A., Kappe, S.H.I., Fleming, H.E., March, S., Sattabongkot, J., Bhatia, S.N., 2018. *In vitro* culture, drug sensitivity, and transcriptome of *Plasmodium vivax* hypnozoites. Cell Host Microbe 23, (395–406) e394.
Iwanaga, S., Kaneko, I., Kato, T., Yuda, M., 2012. Identification of an AP2-family protein that is critical for malaria liver stage development. PLoS One 7, e47557.
Josling, G.A., Llinas, M., 2015. Sexual development in *Plasmodium* parasites: knowing when it's time to commit. Nat. Rev. Microbiol. 13, 573–587.
Kafsack, B.F., Rovira-Graells, N., Clark, T.G., Bancells, C., Crowley, V.M., Campino, S. G., Williams, A.E., Drought, L.G., Kwiatkowski, D.P., Baker, D.A., 2014. A transcriptional switch underlies commitment to sexual development in human malaria parasites. Nature 507, 248.
Kelley, K.D., Miller, K.R., Todd, A., Kelley, A.R., Tuttle, R., Berberich, S.J., 2010. YPEL3, a p53-regulated gene that induces cellular senescence. Cancer Res. 70, 3566–3575.

Kennedy, M., Fishbaugher, M.E., Vaughan, A.M., Patrapuvich, R., Boonhok, R., Yimamnuaychok, N., Rezakhani, N., Metzger, P., Ponpuak, M., Sattabongkot, J., Kappe, S.H., Hume, J.C., Lindner, S.E., 2012. A rapid and scalable density gradient purification method for *Plasmodium* sporozoites. Malar J. 11, 421.

Kim, A., Popovici, J., Vantaux, A., Samreth, R., Bin, S., Kim, S., Roesch, C., Liang, L., Davies, H., Felgner, P., Herrera, S., Arevalo-Herrera, M., Menard, D., Serre, D., 2017. Characterization of *P. vivax* blood stage transcriptomes from field isolates reveals similarities among infections and complex gene isoforms. Sci. Rep. 7, 7761.

Kramer, S., 2014. RNA in development: how ribonucleoprotein granules regulate the life cycles of pathogenic protozoa. WIR. RNA 5, 263–284.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Lasko, P., 2003. Gene regulation at the RNA layer: RNA binding proteins in intercellular signaling networks. Sci. STKE 179, RE6.

Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K., Ritchie, M.E., 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research 5.

Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J., 2004. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Res. 14, 2308–2318.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., S. Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25, 2078–2079.

Lindner, S.E., Mikolajczak, S.A., Vaughan, A.M., Moon, W., Joyce, B.R., Sullivan Jr., W. J., Kappe, S.H., 2013a. Perturbations of *Plasmodium* Puf2 expression and RNA-seq of Puf2-deficient sporozoites reveal a critical role in maintaining RNA homeostasis and parasite transmissibility. Cell. Microbiol. 15, 1266–1283.

Lindner, S.E., Miller, J.L., Kappe, S.H., 2012. Malaria parasite pre-erythrocytic infection: preparation meets opportunity. Cell. Microbiol. 14, 316–324.

Lindner, S.E., Swearingen, K.E., Harupa, A., Vaughan, A.M., Sinnis, P., Moritz, R.L., Kappe, S.H., 2013b. Total and putative surface proteomics of malaria parasite salivary gland sporozoites. Mol. Cell. Proteom. 12, 1127–1143.

Lopez-Rubio, J.-J., Mancio-Silva, L., Scherf, A., 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5, 179–190.

Lysenko, A.J., Beljaev, A., Rybalka, V., 1977. Population studies of *Plasmodium vivax*: 1. The theory of polymorphism of sporozoites and epidemiological phenomena of tertian malaria. Bull. WHO 55, 541.

Ma, C., Sim, S., Shi, W., Du, L., Xing, D., Zhang, Y., 2010. Energy production genes sucB and ubiF are involved in persister survival and tolerance to multiple antibiotics and stresses in *Escherichia coli*. FEMS Microbiol. Lett. 303, 33–40.

Mackellar, D.C., O'Neill, M.T., Aly, A.S., Sacci Jr., J.B., Cowman, A.F., Kappe, S.H., 2010. *Plasmodium falciparum* PF10_0164 (ETRAMP10.3) is an essential parasitophorous vacuole and exported protein in blood stages. Eukaryot. Cell 9, 784–794.

Malmquist, N.A., Moss, T.A., Mecheri, S., Scherf, A., Fuchter, M.J., 2012. Small-molecule histone methyltransferase inhibitors display rapid antimalarial activity against all blood stage forms in *Plasmodium falciparum*. PNAS 109, 16708–16713.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10–12.

Mikolajczak, S.A., Silva-Rivera, H., Peng, X., Tarun, A.S., Camargo, N., Jacobs-Lorena, V., Daly, T.M., Bergman, L.W., de la Vega, P., Williams, J., 2008. Distinct malaria parasite sporozoites reveal transcriptional changes that cause differential tissue infection competence in the mosquito vector and mammalian host. Mol. Cell. Biol. 28, 6196–6207.

Mikolajczak, S.A., Vaughan, A.M., Kangwanrangsan, N., Roobsoong, W., Fishbaugher, M., Yimamnuaychok, N., Rezakhani, N., Lakshmanan, V., Singh, N., Kaushansky, A., Camargo, N., Baldwin, M., Lindner, S.E., Adams, J.H., Sattabongkot, J., Kappe, S. H., 2015. *Plasmodium vivax* liver stage development and hypnozoite persistence in human liver-chimeric mice. Cell Host Microbe 17, 526–535.

Mota, M.M., Pradel, G., Vanderberg, J.P., Hafalla, J.C., Frevert, U., Nussenzweig, R.S., Nussenzweig, V., Rodriguez, A., 2001. Migration of *Plasmodium* sporozoites through cells before infection. Science 291, 141–144.

Mueller, A.-K., Camargo, N., Kaiser, K., Andorfer, C., Frevert, U., Matuschewski, K., Kappe, S.H., 2005. *Plasmodium* liver stage developmental arrest by depletion of a protein at the parasite–host interface. PNAS 102, 3022–3027.

Mueller, I., Galinski, M.R., Baird, J.K., Carlton, J.M., Kochar, D.K., Alonso, P.L., del Portillo, H.A., 2009. Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. Lancet Infect. Dis. 9, 555–566.

Nikolayeva, O., Robinson, M.D., 2014. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. Methods Mol. Biol. 1150, 45–79.

Okano, H., Imai, T., Okabe, M., 2002. Musashi: a translational regulator of cell fate. J. Cell. Sci. 115, 1355–1359.

Okonechnikov, K., Conesa, A., Garcia-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics (Oxford, England) 32, 292–294.

Painter, H.J., Campbell, T.L., Llinás, M., 2011. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. Mol. Biochem. Parasitol. 176, 1–7.

Price, R.N., Douglas, N.M., Anstey, N.M., 2009. New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. Curr. Opin. Infect. Dis. 22, 430–435.

Price, R.N., Tjitra, E., Guerra, C.A., Yeung, S., White, N.J., Anstey, N.M., 2007. Vivax malaria: neglected and not benign. Am. J. Trop. Med. Hyg. 77, 79–87.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) 26, 841–842.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucl. Acids Res. 43. e47 e47.

Roobsoong, W., Tharinjaroen, C.S., Rachaphaew, N., Chobson, P., Schofield, L., Cui, L., Adams, J.H., Sattabongkot, J., 2015. Improvement of culture conditions for long-term in vitro culture of *Plasmodium vivax*. Malaria J. 14, 1.

Roth, A., Adapa, S.R., Zhang, M., Liao, X., Saxena, V., Goffe, R., Li, S., Ubalee, R., Saggu, G.S., Pala, Z.R., Garg, S., Davidson, S., Jiang, R.H.Y., Adams, J.H., 2018. Unraveling the *Plasmodium vivax* sporozoite transcriptional journey from mosquito vector to human host. Sci. Rep. 8, 12183.

Sattabongkot, J., Tsuboi, T., Zollner, G.E., Sirichaisinthop, J., Cui, L., 2004. *Plasmodium vivax* transmission: chances for control? Trends Parasitol. 20, 192–198.

Schrader, J., Moyle, R., Bhalerao, R., Hertzberg, M., Lundeberg, J., Nilsson, P., Bhalerao, R.P., 2004. Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome. Plant J. 40, 173–187.

Shi, S., Ehrt, S., 2006. Dihydrolipoamide acyltransferase is critical for *Mycobacterium tuberculosis* pathogenesis. Infect. Immun. 74, 56–63.

Shin, S.C., Vanderberg, J.P., Terzakis, J.A., 1982. Direct infection of hepatocytes by sporozoites of *Plasmodium berghei*. J. Protozool. 29, 448–454.

Silvie, O., Briquet, S., Muller, K., Manzoni, G., Matuschewski, K., 2014a. Post-transcriptional silencing of UIS4 in *Plasmodium berghei* sporozoites is important for host switch. Mol. Microbiol. 91, 1200–1213.

Silvie, O., Briquet, S., Müller, K., Manzoni, G., Matuschewski, K., 2014b. Post-transcriptional silencing of UIS4 in *Plasmodium berghei* sporozoites is important for host switch. Mol. Microbiol. 91, 1200–1213.

Swearingen, K.E., Lindner, S.E., Flannery, E.L., Vaughan, A.M., Morrison, R.D., Patrapuvich, R., Koepfli, C., Muller, I., Jex, A., Moritz, R.L., Kappe, S.H.I., Sattabongkot, J., Mikolajczak, S.A., 2017. Proteogenomic analysis of the total and surface-exposed proteomes of *Plasmodium vivax* salivary gland sporozoites. PLoS Negl. Trop. Dis. 11, e0005791.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578.

Tucker, R.P., 2004. The thrombospondin type 1 repeat superfamily. Int. J. Biochem. Cell. Biol. 36, 969–974.

Tuttle, R., Simon, M., Hitch, D.C., Maiorano, J.N., Hellan, M., Ouellette, J., Termuhlen, P., Berberich, S.J., 2011. Senescence-associated gene YPEL3 is downregulated in human colon tumors. Ann. Surg. Oncol. 18, 1791–1796.

Voorberg-van der Wel, A., Roma, G., Gupta, D.K., Schuierer, S., Nigsch, F., Carbone, W., Zeeman, A.M., Lee, B.H., Hofman, S.O., Faber, B.W., Knehr, J., Pasini, E., Kinzel, B., Bifani, P., Bonamy, G.M.C., Bouwmeester, T., Kocken, C.H.M., Diagana, T.T., 2017. A comparative transcriptomic analysis of replicating and dormant liver stages of the relapsing malaria parasite *Plasmodium cynomolgi*. Elife 6, e29605.

Westenberger, S.J., McClean, C.M., Chattopadhyay, R., Dharia, N.V., Carlton, J.M., Barnwell, J.W., Collins, W.E., Hoffman, S.L., Zhou, Y., Vinetz, J.M., Winzeler, E.A., 2010. A systems-based analysis of *Plasmodium vivax* lifecycle transcription from human to mosquito. PLoS Negl. Trop. Dis. 4, e653.

White, M.T., Karl, S., Battle, K.E., Hay, S.I., Mueller, I., Ghani, A.C., 2014. Modelling the contribution of the hypnozoite reservoir to *Plasmodium vivax* transmission. Elife 3, e04692.

White, N.J., 2011. Determinants of relapse periodicity in *Plasmodium vivax* malaria. Malar J. 10, 297.

WHO, 2015. World Malaria Report 2015. WHO, Geneva.

Wood, F.C., Heidari, A., Tekle, Y.I., 2017. Genetic evidence for sexuality in *Cochliopodium* (Amoebozoa). J. Hered. 108, 769–779.

Wu, Q., Bruce, A.W., Jedrusik, A., Ellis, P.D., Andrews, R.M., Langford, C.F., Glover, D. M., Zernicka-Goetz, M., 2009. CARM1 is required in embryonic stem cells to maintain pluripotency and resist differentiation. Stem Cells 27, 2637–2645.

Yazawa, K., Kamada, H., 2007. Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1. J. Exp. Bot. 58, 3819–3828.

Zanghi, G., Vembar, S.S., Baumgarten, S., Ding, S., Guizetti, J., Bryant, J.M., Mattei, D., Jensen, A.T.R., Renia, L., Goh, Y.S., Sauerwein, R., Hermsen, C.C., Franetich, J.F., Bordessoulles, M., Silvie, O., Soulard, V., Scatton, O., Chen, P., Mecheri, S., Mazier, D., Scherf, A., 2018. A Specific PfEMP1 Is expressed in *P. falciparum* sporozoites and plays a role in hepatocyte infection. Cell Rep. 22, 2951–2963.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhu, L., Mok, S., Imwong, M., Jaidee, A., Russell, B., Nosten, F., Day, N.P., White, N.J., Preiser, P.R., Bozdech, Z., 2016. New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. Sci. Rep. 6, 20498.