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

Vivax Sporozoite Consortium[*] (Ivo Muller[1,2,3], Aaron R. Jex[1,3,4], Stefan H. I. Kappe[5], Sebastian A. Mikolajczak[5], Jetsumon Sattabongkot[7], Rapatbhorn Patrapuvich[6], Scott Lindner[8], Erika L. Flannery[5], Cristian Koepfli[1], Brendan Ansell[4], Anita Lerch[1], Samantha J Emery-Corbin[1], Sarah Charnaud[1], Jeffrey Smith[1], Nicolas Merrienne[2], Kristian E. Swearingen[5], Robert L. Moritz[9], Michaela Petter[10,11], Michael Duffy[10], Vorada Chuenchob[5]).

[*]**Group authorship – all authors are equal contributors (order per author contributions section below).**

1. Population Health and Immunity Division, The Walter and Eliza Hall Institute for Medical Research, 1G Royal Parade, Parkville, Victoria, 3052, Australia.
2. Malaria: Parasites & Hosts Unit, Institut Pasteur, 28 Rue de Dr. Roux, 75015, Paris, France.
3. Department of Medical Biology, The University of Melbourne, Victoria, 3010, Australia.
4. Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Corner of Park and Flemington Road, Parkville, Victoria, 3010, Australia.
5. Center for Infectious Disease Research, 307 Westlake Avenue North, Suite 500, Seattle, WA 98109, USA;
6. Department of Global Health, University of Washington, Seattle, WA 98195, USA.
7. Mahidol *Vivax* Research Center, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand.
8. Department of Biochemistry and Molecular Biology, Center for Malaria Research, Pennsylvania State University, University Park, PA 16802, USA.
9. Institute for Systems Biology, Seattle, WA, 98109, USA.
10. Department of Medicine Royal Melbourne Hospital, The Peter Doherty Institute, The University of Melbourne, 792 Elizabeth Street, Melbourne, Victoria 3000, Australia.
11. Institute of Microbiology, University Hospital Erlangen, Erlangen 91054, Germany

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. The salivary-gland sporozoite transcriptome is heavily composed of transcripts associated with functions needed for early infection of the vertebrate host and development within hepatocytes. Through comparisons to recently published proteome data for the *P. vivax* sporozoite, our study finds that although highly transcribed, these transcripts are not detectable as proteins and may be regulated through translational repression; a finding we test for a small subset of transcripts and proteins through immunofluorescent microscopy of sporozoites and liver stages in humanized mice. We identify 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 largely absent in replicating liver schizonts or mixed blood-stages. Common transcripts up-regulated in sporozoites and hypnozoites compared to mixed (i.e., schizont) liver-stages identify genes linked to dormancy/persistence in bacteria, amoebae and plants. We also characterise histone epigenetic modifications in the *P. vivax* sporozoite and explore their role in regulating transcription. Collectively, these data support the hypothesis that the sporozoite as a tightly programmed stage primed to infect the human host and identifies potential mechanisms for hypnozoite-formation that may be further explored in liver stage models.

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 [1]. Two major parasitic species contribute to the vast majority of human malaria, *Plasmodium falciparum* and *P. vivax*. Historically, *P. falciparum* has attracted the majority of global attention, due to its higher contribution to morbidity and

1

59   mortality. However, *P. vivax* is broadly distributed, more pathogenic than previously thought,
60   and is recognised as the key obstacle to malaria elimination in the Asia-Pacific and Americas
61   [2]. Unlike *P. falciparum*, *P. vivax* can establish long-lasting 'sleeper-cells' (= hypnozoites)
62   in the host liver that emerge weeks, months or years after the primary infection (= relapsing
63   malaria) [3]. Primaquine is the only approved drug that prevents relapse. However, the short
64   half-life, long dosage regimens and incompatibility of primaquine with glucose-6-phosphate-
65   dehydrogenase deficiency (which requires pre-screening of recipients [4]) makes it unsuitable
66   for widespread use. As a consequence, *P. vivax* is overtaking *P. falciparum* as the primary
67   cause of malaria in a number of co-endemic regions [5]. Developing new tools to diagnose,
68   treat and/or prevent hypnozoite infections is considered one of the highest priorities in the
69   malaria elimination research agenda [6].
70   When *Plasmodium* sporozoites are deposited by an infected mosquito, they likely
71   traverse the skin cells, enter the blood-stream and are trafficked to the host liver, as has been
72   shown in rodents [7]. The sporozoites' journey from skin deposition to hepatocytes takes less
73   than a few minutes [8]. Upon reaching the liver, sporozoites traverse Kupffer and endothelial
74   cells to reach the parenchyma, moving through several hepatocytes before invading a final
75   hepatocyte suitable for development [7, 9]. Within hepatocytes, these parasites replicate, and
76   undergo further development and differentiation to produce merozoites that emerge from the
77   liver and infect red blood cells. However, *P. vivax* sporozoites are able to commit to two
78   distinct developmental fates within the hepatocyte: they either immediately continue
79   development as replicating schizonts and establish a blood infection, or delay replication and
80   persist as hypnozoites. Regulation of this major developmental fate decision is not understood
81   and this represents a key gap in current knowledge of *P. vivax* biology and control.
82   Sporozoites prepare for mammalian host infection while still residing in the mosquito
83   salivary glands. It has been hypothesized that *P. vivax* sporozoites exist within an inoculum as
84   replicating 'tachysporozoites' and relapsing 'bradysporozoites' [10] and that these
85   subpopulations may have distinct developmental fates as schizonts or hypnozoites, thus
86   contributing to their relapse phenotype [10-12]. This observation is supported by the stability
87   of different hypnozoite phenotypes (ratios of hypnozoite to schizont formation) in *P. vivax*
88   infections of liver-chimeric mouse models [13]. To determine fates in the sporozoite stage
89   control of protein expression must take place. Studies using rodent malaria parasites have
90   identified genes [14] that are transcribed in sporozoites but translationally repressed (i.e.,
91   present as transcript but un- or under-represented as protein), via RNA-binding proteins [15],
92   and ready for immediate translation after the parasites' infection of the mammalian host cell
93   [13, 16]. It is therefore also possible that translational repression (i.e., the blocking of
94   translation of present and retained transcripts) and other mechanisms of epigenetic control
95   may contribute to the *P. vivax* sporozoite fate decision and hypnozoite formation, persistence
96   and activation. Supporting this hypothesis, histone methyltransferase inhibitors stimulate
97   increased activation of *P. cynomolgi* hypnozoites to become schizonts in macaque
98   hepatocytes [17, 18]. Epigenetic control of stage development is further evidenced in
99   *Plasmodium* through chromatin structure controlling expression of PfAP2-G, a specific
100   transcription factor that, in turn, regulates gametocyte (dimorphic sexual stages) development
101   in blood-stages [19]. It is well documented that *P. vivax* hypnozoite activation patterns
102   stratify with climate and geography [11] and recent modelling suggests transmission potential
103   selects for hypnozoite phenotype [20]. Clearly the ability for *P. vivax* to dynamically regulate
104   hypnozoite formation and relapse phenotypes in response to high or low transmission periods
105   in different climate conditions would confer a significant evolutionary advantage.
106   Unfortunately, despite recent advances [21] current approaches for *in vitro P. vivax*
107   culture do not support routine maintenance in the laboratory and tools to directly perturb gene
108   function are not established. This renders studies on *P. vivax*, particularly its sporozoites and
109   liver stages, exceedingly difficult. Although *in-vitro* liver stage assays and humanised mouse
110   models are being developed [13], 'omics analysis of *P. vivax* liver stage dormancy has until
111   recently [22] been impossible and even now is in its early stages. Recent characterization [23]
112   of liver-stage (hypnozoites and schizonts) of *P. cynomolgi* (a related and relapsing parasite in
113   macaques) provides valuable insight, but investigations in *P. vivax* directly are clearly

114 needed. The systems analysis of *P. vivax* sporozoites that reside in the mosquito salivary
115 glands and are poised for transmission and liver infection offer a key opportunity to gain
116 insight into *P. vivax* infection. *Plasmodium vivax* sporozoites have been explored previously
117 by microarray [24] and most recently, in a single RNA-seq replicate [25] and a study on
118 sporozoite activation {Roth, 2018 #66}. Epigenetic regulation in sporozoites has only been
119 explored in *P. falciparum* [26, 27]. Here, we present a detailed characterization of the *P. vivax*
120 sporozoite transcriptome and histone epigenome and use these data to better understand this
121 key infective stage and the role of sporozoite programming in invasion and infection of the
122 human host, and development within the host liver.
123
124 **RESULTS AND DISCUSSION**
125 Mosquito infections were generated by membrane feeding of blood samples taken from *P.*
126 *vivax* infected patients in western Thailand (n = 9). Approximately 3-15 million *P. vivax*
127 sporozoites were harvested per isolate from *Anopheles dirus* salivary glands. Using RNA-seq,
128 we detected transcription for 5,714 *P. vivax* genes (based on the *P. vivax* P01 gene models:
129 [28]) and obtained a high degree of coverage (4,930 with a mean counts per million (CPM) ≥
130 1.0; Figure S1 and Table S1 and S2). Among the most highly transcribed genes in the
131 infectious sporozoite stage are *csp* (c̲ircums̲porozoite protein), five *etramps* (e̲arly t̲ranscribed
132 m̲embrane p̲roteins), including *uis3* (u̲p-regulated in i̲nfective s̲porozoites), *uis4* and *lsap-1*
133 (l̲iver s̲tage a̲ssociated p̲rotein 1), a variety of genes involved in cell transversal and initiation
134 of invasion, including *celtos* (c̲ell traversal protein for o̲okinetes and s̲porozoites), *gest*
135 (g̲amete e̲gress and s̲porozoite t̲raversal protein), *spect1* (s̲porozoite protein e̲ssential for c̲ell
136 t̲raversal) and *siap-1* (s̲porozoite i̲nvasion a̲ssociated p̲rotein), and genes associated with
137 translational repression (*alba1*, *alba4* and *Puf2*). Collectively, these genes account for >1/3rd
138 of all transcripts in the sporozoite. Although we found only moderate agreement (R^2 = 0.35;
139 Figure S2) between our RNA-seq data and previous microarray data for *P. vivax* sporozoites
140 and blood-stages [24], improved transcript detection and quantitation is expected with the
141 increased technical resolution of RNA-seq over microarray. Supporting this, we find higher
142 correlation between RNA-seq data from *P. vivax* and *P. falciparum* (single replicate
143 sequenced herein for comparative purposes) sporozoite datasets (R^2 = 0.42), compared to
144 either species relative to published microarray data (Figure S2 and Table S3).
145   Although microarray supports the high transcription in sporozoites of genes such as
146 *uis4*, *csp*, *celtos* and several other *etramps*, 27% and 16% of the most abundant 1% of
147 transcribed genes in our sporozoite RNA-seq data are absent from the top decile or quartile
148 respectively in the existing *P. vivax* sporozoite microarray data [24]. Among these are genes
149 involved in early invasion/hepatocyte development, such as *lsap-1*, *celtos*, *gest* and *siap-1*, or
150 translational repression (e.g., *alba-1* and *alba-4*); orthologs of these genes are also in the top
151 percentile of transcripts in RNA-seq (see [26, 29]) and previous microarray data [30, 31] for
152 human-infecting *P. falciparum* and murine-infecting *P. yoelii* sporozoites, suggesting many
153 are indeed more abundant than previously characterized. A subset of representative
154 transcripts, including *Pv_AP2-X* (PVP01_0733100), *d13*, *gest*, *g10* (PVP01_1011100), 40S
155 ribosomal protein S27 (PVP01_1409300), *puf-2*, *zipco* and 14-3-3, were tested by qPCR for
156 their transcript abundance relative to *celtos* and *sera* (Figure 1A and Table S4). This
157 representative set differed markedly in their relative abundance between our RNAseq and
158 previous microarray data [24]. To control for batch effects introduced by collection of the
159 sporozoites used here for RNAseq, this testing was conducted in an additional six sample
160 replicates representing four additional clinical *P. vivax* isolates (PvSPZ-Thai13-16; with
161 PvSPZ-Thai16 tested in technical triplicate). The qPCR results agreed with the RNAseq data
162 for these transcripts (Table S4).
163
164 ***Transcription in* P. vivax *relative to other plasmodia* sporozoites*.** To gain insight into
165 species-specific aspects of the *P.vivax* transcriptome, we qualitatively compared these data
166 with available data for *P. falciparum* [27] and *P. yoelii* sporozoites (single replicate only) for
167 4,067 single-copy orthologs (SCO) (transcribed at ≥ 1 TPM in *P. vivax* infectious
168 sporozoites) shared with *P. falciparum* and *P. yoelli* (Table S5). Genes highly transcribed in

169   salivary-gland sporozoites of all three species include *celtos*, *gest*, *trap*, *siap1*, *spect1* and
170   *puf2*. There are 696 *P. vivax* genes shared as orthologs between *P. vivax* P01 and *P. vivax*
171   Sal1 lacking a defined SCO in *P. falciparum* or *P. yoelli* transcribed at a mean of ≥ 1 TPM in
172   *P. vivax* salivary-gland sporozoites (Table S6). Prominent among these are *vir* (n=25) and *Pv-*
173   *fam* (41 fam-e, 16 fam-b, 14 fam-a, 8 fam-d and 3 fam-h) genes, as well as hypothetical
174   proteins or proteins of unknown function (n=212) and, interestingly, a number of 'merozoite
175   surface protein' 3 and 7 homologs (n=5 of each). Both *msp3* and *msp7* have undergone
176   significant expansion in *P. vivax* relative to *P. falciparum* and *P. yoelii* [32] and may have
177   repurposed functions in sporozoites. In addition, there are 69 *P. vivax* P01 genes lacking a
178   defined ortholog in *P. vivax* Sal1, *P. falciparum* or *P. yoelli* transcribed at ≥ 1 TPM in
179   infectious *P. vivax* sporozoites; most of which are *Plasmodium* interspersed repeat (PIR)
180   genes [32] found in telomeric regions of the P01 assembly and likely absent from the Sal1
181   assembly but present in the Sal1 genome, indicating the improved coverage of telomeric
182   regions in P01 relative to Sal1.
183
184   ***P. vivax* sporozoites transcriptome compared with proteome.** We compared relative
185   protein abundance presented in a recently published *P. vivax* sporozoite proteome [33] to
186   relative transcript abundance from the current study (Figure 1B and Table S7). The proteome
187   study incorporated data from the same PvSPZ-Thai1 and PvSPZ-Thai5 isolates tested by
188   RNAseq here. We identified 2,402 *P. vivax* genes transcribed in the sporozoite (CPM > 1) for
189   which no protein expression was detected. Although many of these are lowly transcribed and
190   likely below the detection sensitivity of LC-MS proteomics, others are among the most highly
191   transcribed genes in the sporozoite, indicating these may be under translational repression.
192   Translational repression, the mechanism through which transcripts are held in stasis
193   by RNA binding proteins, has been demonstrated to have important functional roles in the
194   transition of *Plasmodium* spp. between the vertebrate to invertebrate host. More than 700
195   genes have been identified as translationally repressed in *Plasmodium berghei* ('rodent
196   malaria') gametocytes based on DOZI (DEAD box RNA helicase "<u>d</u>evelopment <u>o</u>f <u>z</u>ygote
197   <u>i</u>nhibited") pulldowns [34]. Translational repression mechanisms mediated through Puf-2
198   have been explored in sporozoites of several *Plasmodium* species and regulate some of the
199   most abundant transcripts in the sporozoite, such as *uis-3* and *uis-4*. UIS3 and UIS4 are the
200   best characterized proteins under translational repression by Puf-2 in sporozoites [35] and are
201   essential for liver-stage development [14].
202   In considering genes that may be translationally repressed (i.e., transcribed but not
203   translated) in the *P. vivax* sporozoite, we confine our observations to those transcripts
204   representing the top decile of transcript abundance to ensure their lack of detection as proteins
205   was not due to limitations in the detection sensitivity of the proteomic dataset. Approximately
206   1/3rd of transcripts in the top decile of transcriptional abundance (n = 170 of 558) in *P. vivax*
207   sporozoites were not detectable as peptides in multiple replicates (**Figure 1B** and Table S7).
208   Of these 170 putatively repressed transcripts, 156 and 154 have orthologs in *P. falciparum*
209   and *P. yoelii* respectively, with 89 and 118 of these also not detected as proteins in *P.*
210   *falciparum* and *P. yoelii* salivary-gland sporozoites [36] despite being highly transcribed in
211   these stages (see [26, 29]; Tables S3-S5), and 133 (78.2%) having no detectable sporozoite
212   expression (>1 unique peptide count) in LC-MS data deposited for any species in PlasmoDB
213   (Table S8). In contrast, 106 of these putatively repressed transcripts with orthologs in other
214   *Plasmodium* species (Table S8) for which proteomic data is available in PlasmoDB, are
215   detectable (>1 unique peptide count) by LC-MS methods in at least one other life-cycle stage,
216   indicating against a technical issue (e.g., inability to be trypsin-digested) preventing their
217   detection in the *P. vivax* sporozoite proteome [33]. In addition to *uis3* and *uis4*, genes
218   involved in liver stage development and detectable as transcripts but not proteins in the *P.*
219   *vivax* sporozoites include *lsap1* (<u>l</u>iver <u>s</u>tage <u>a</u>ssociated <u>p</u>rotein 1), *zipco* (<u>ZIP</u> domain-
220   <u>co</u>ntaining protein), several other *etramps* (PVP01_1271000, PVP01_0422600,
221   PVP01_0504800 and PVP01_0734800), *pv1* (<u>p</u>arasitophorous <u>v</u>acuole protein 1) and *lisp1*
222   and *lisp2* (PVP01_1330800 and PVP01_0304700). Also notable among genes detectable as
223   transcripts but not proteins in sporozoites is a putative 'Yippee' homolog (PVP01_0724100).

4

224   Yippee is a DNA-binding protein that, in humans (YPEL3), suppresses cell growth [37] and
225   is regulated through histone acetylation [38], making it noteworthy in the context of *P. vivax*
226   hypnozoite developmental arrest.
227   Although verifying each putatively repressed transcript will require further empirical
228   data, our system level approach is supported by immunofluorescent microscopy (Figure 1C)
229   of UIS4, LISP1, EXP1 and ACP (PVP01_0416300). These represent one known and three
230   putative (i.e., newly proposed here) translationally repressed genes in *P. vivax* sporozoites,
231   and are compared to TRAP and BiP (which are both transcribed and expressed as protein in
232   the *P. vivax* sporozoite; Table S8). The *lisp1* gene is an interesting find. In *P. berghei*, *lisp1* is
233   essential for rupture of the PVM during liver stage development allowing release of the
234   merozoite into the host blood stream. *Pv-lisp1* is ~350-fold and ~1,350-fold more highly
235   transcribed in *P. vivax* sporozoites compared to sporozoites of either *P. falciparum* or *P.*
236   *yoelii* (see Table S5). IFAs using LISP1 specific mAbs (Figure 1C) show that this protein is
237   undetectable in sporozoites but clearly expressed at 7 days post-infection in liver schizonts.
238
239   **Up-regulated transcripts in *P. vivax* sporozoites relative to other life-cycle stages.**
240   Recently completed studies of the transcriptome of *P. vivax* for sporozoite activation [39], as
241   well as, liver [22] and asexual blood-stages [40] support comparative transcriptomic study of
242   sporozoites, their biology and transcriptional regulation over the *P. vivax* life-cycle. Recently
243   published data for activate sporozoites from Roth et al [39] was significantly lower depth
244   coverage, with ~0.03 to 0.6M reads mapped the *P. vivax* P01 coding domains; compared with
245   0.7 to 15.3 M, 2.4 to 10.6 M and 18.7 to 57.6M mapped reads for salivary sporozoite, liver-
246   stages [22] and asexual blood stages [40] respectively. This lower coverage could not be
247   compensated for through data normalization and therefore data from Roth et al [39] was not
248   included in our quantitative analyses, although qualitatively, many of the highly transcribed
249   genes in Roth et al [39] sporozoites were among the highly transcribed genes in salivary
250   sporozoites from the present study. The remaining RNAseq data presents an analytical
251   challenge in that each (sporozoites, liver-stages and blood stages) is produced in a separate
252   study and may be influenced by technical batch effects that cannot be differentiated from
253   biologically meaningful changes. To address this, we first examined *P. vivax* transcripts in a
254   previous microarray study of multiple *P. vivax* life-cycle stages [24], including sporozoites
255   and several blood-stages, to identify genes that may be transcriptionally stable across the life-
256   cycle. We identified ~160 genes with low transcriptional variability between sporozoites and
257   blood-stages that covered the breadth of transcript abundance levels in Westenberger et al
258   [24]. These include genes typically associated with "house-keeping" functions, such as
259   ribosomal proteins, histones, translation initiation complex proteins and various chaperones
260   (see Figure S3 and Table S8). We assessed transcription of these 160 genes among the current
261   and recently published RNA-seq data for *P. vivax* and all were of similarly low variability
262   (Figure S4). This suggests that any batch effect between the studies is sufficiently lower than
263   the biological differences between each life-cycle stage, allowing informative comparisons.
264   We then combined all published RNAseq-based, transcriptomic data available for *P. vivax*
265   [22, 39, 40] with the salivary sporozoite data generated here.
266
267   **P. vivax *sporozoite relative to blood-stage transcriptome.*** To identify transcripts up-
268   regulated in sporozoites, we first compared the *P. vivax* sporozoite transcriptome to RNA-seq
269   data for *P. vivax* blood-stages [40] (Figure 2 and Figures S7-S10). We identified 1,672 up
270   (Table S9); Interactive Glimma Plot - Supplementary Data 1) and 1,958 down-regulated
271   (Table S9); Interactive Glimma Plot - Supplementary Data 1) transcripts (FDR $\leq$ 0.05;
272   minimum 2-fold change in Counts per Million (CPM)) and explored patterns among these
273   differentially transcribed genes (DTGs) by protein family (Figure 2C and Table S10) and
274   Gene Ontology (GO) classifications (Table S11). RNA recognition motifs (RRM-1 and
275   RRM-6) and helicase domains (Helicase-C and DEAD box helicases) are over-represented
276   (p-value <0.05) among transcripts up-regulated in sporozoites, consistent with translational
277   repression through ribonucleoprotein (RNP) granules [41]. Transcripts encoding nucleic acid
278   binding domains, such as bromodomains (PF00439; which can also bind lysine-acetylated

5

279   proteins), zinc fingers (PF13923) and EF hand domains (PF13499) are also enriched in
280   sporozoites. Included among these proteins are a putative ApiAP2 transcription factor
281   (PVP01_1211900) and a homologue of the *Drosophila* zinc-binding protein 'Yippee'
282   (PVP01_0724100). Thrombospondin-1 like repeats (TSR: PF00090) and von Willebrand
283   factor type A domains (PF00092) are enriched in sporozoites as well. In *P. falciparum*
284   sporozoites, genes enriched in TSR domains are important in invasion of the mosquito
285   salivary gland (e.g., *trap*) and secretory vesicles released by sporozoites upon entering the
286   vertebrate host (e.g., *csp*) [42]. By comparison, genes up-regulated in blood-stages are
287   enriched for *vir* gene domains (PF09687 and PF05796), Tryptophan-Threonine-rich
288   *Plasmodium* antigens (PF12319; which are associated with merozoites [43]), markers of cell-
289   division (PF02493; [44]) protein production/degradation (PF00112, PF10584, PF00152,
290   PF09688 and PF00227) and ATP metabolism (PF08238 and PF12774). 47 of the 343
291   transcripts unique to *P. vivax* sporozoites relative to *P. falciparum* or *P. yoelii* are up-
292   regulated in sporozoites compared to *P. vivax* blood stages. Nine of these are in the top decile
293   of transcription, and include a Pv-fam-e (PVP01_0525200), a Pf-fam-b homolog
294   (PVP01_0602000) and 7 proteins of unknown function. A further nine have an ortholog in *P.*
295   *cynomolgi* (which also forms hypnozoites) but not the closely related *P. knowlesi* (which does
296   not form hypnozoites) and include '*msp7*'-like (PVP01_1219600, PVP01_1220300 and
297   PVP01_1219900), '*msp3*'-like (PVP01_1031300), Pv-fam-e genes (PVP01_0302100,
298   PVP01_0524500 and PVP01_0523400), a serine-threonine protein kinase (PVP01_0207300)
299   and a RecQ1 helicase homolog (PVP01_0717000). Notably, the *P. cynomolgi* ortholog of
300   PVP01_0207300, PCYB_021650, is transcriptionally up-regulated in hypnozoites relative to
301   replicating schizonts [23], indicating a target of significant interest when considering
302   hypnozoite formation and/or biology and suggesting that the list here may contain other genes
303   important in hypnozoite biology.
304
305   **P. vivax *sporozoites are enriched in translational repressors.*** In *Plasmodium*, translational
306   repression regulates key life-cycle transitions coinciding with switching between the
307   mosquito and the mammalian host (either as sporozoites or gametocytes) [41]. For example,
308   although *uis4* is the most abundant transcript in the infectious sporozoite ([24, 31]; Table S2),
309   UIS4 is translationally repressed in this stage [15] and only expressed after hepatocyte
310   invasion [45]. In sporozoites, it is thought that PUF2 binds to mRNA transcripts and prevents
311   their translation [29], and SAP1 stabilises the repressed transcripts and prevents their
312   degradation [45]. Consistent with this, *Puf2* and *SAP1* (PVP01_0947600) are up-regulated in
313   the sporozoite relative to blood-stages. Indeed, *Puf2* (PVP01_0526500) is among the top
314   percentile of transcripts in salivary sporozoites and expressed at high levels in the proteome
315   [33]. However, our data implicate other genes that may act in translational repression in *P.*
316   *vivax* sporozoites, many of which are already known to be involved in translational repression
317   in other *Plasmodium* stages and other protists [41]. Among these are *alba-2* and *alba-4*, both
318   of which are among the top 2% of genes transcribed in sporozoites and ~14 to 20-fold more
319   highly transcribed in sporozoites relative to blood-stages; ALBA-2 is in the top 100 most
320   abundant proteins in the *P. vivax* sporozoite proteome [33]. In addition, *P. vivax* sporozoites
321   are enriched for genes encoding RRM-6 RNA helicase domains. Intriguing among these are
322   HoMu (homolog of Musashi; top decile of sporozoite proteins by abundance [33]) and ptbp
323   (polypyrimidine tract binding protein). Musashi regulates eukaryotic stem cell differentiation
324   through translational repression [46] and HoMu localizes with DOZI and CITH in
325   *Plasmodium* gametocytes [47]. PTBP is linked to mRNA stability, splice regulation and
326   translational initiation [48] and may perform a complementary role to SAP1.
327
328   **P. vivax *sporozoite relative to Plasmodium spp. liver stage transcriptomes.*** New advances in
329   *P. vivax* liver culture has allowed recent publication of mixed stage and hypnozoite-enriched
330   transcriptomes [22]. This is an early, yet highly valuable, study and, due no doubt to the
331   difficulty in generating the material, is limited to biological duplicates. Noting this, although
332   we undertake differential transcriptomic studies of this dataset here, we recognize that
333   additional biological replication is needed and have used a higher burden of significance

6

334 (FDR ≤ 0.01 and ≥ 2-fold change) than used with blood-stages. Nevertheless, these
335 comparisons identified 1,015 and 856 sporozoite up-regulated transcripts relative to mLS and
336 HPZs respectively and 1,007 and 1,079 transcripts up-regulated in mLS and HPZs relative to
337 sporozoites respectively (Figures 3 and S11-S13, Table S12 and S13 and Interactive Glimma
338 Plot - Supplementary Data 1).
339 Compared to mLS transcriptomes, sporozoites are enriched for many of the
340 transcripts similarly up-regulated in comparison to blood-stages (e.g., *uis4*, *celtos*, *puf2*, *siap1*
341 and *plp-1*). More broadly, SPZ up-regulated transcripts over-represent (p-value ≤ 0.05) Pfam
342 domains (Figure 3A) associated with transcriptional regulation (PF00176, PF01096, PF01661
343 and PF08711), translational repression/regulation (PF00076, PF00279, PF01008, PF01873,
344 PF01917, PF02847, PF02854, PF13893 and PF14259), DNA/RNA binding (PF0097,
345 PF13445, PF13639, PF14570 and PF15247) and chromatin regulation (PF00271, PF00850
346 and PF13489). In contrast, the mLS transcriptome is enriched in genes involved in replication
347 and merozoite formation [n = 14; including PVP01_0728900 (*msp1*), PVP01_0010670
348 (*msp3*) and PVP01_1446800 (*msp9*)], rhoptry function [n = 9; including PVP01_1469200
349 (*rnp3*), PVP01_1255000 (*rnp2*) and PVP01_1338500 (*rap1*)] and reticulocyte binding [n=10
350 including PVP01_0534300 (rbp2c), PVP01_1402400 (rbp2a), PVP01_0701100 (rbp1b) and
351 PVP01_0800700 (rbp2b)]. These data are further enriched for Pfam domains associated with
352 cell division (PF00493), merozoite formation (PF07133 and PF12984), proteasome function
353 (PF00227, PF00400, PF00656, PF01344, PF01398 and PF03981), protein export / vesicle
354 function (PF00350 and PF00996), membrane proteins (PF01105, PF03011, PF05424 and
355 PF12139) and metabolism (PF00085, PF00118, PF00268, PF01066, PF01214 and PF01214).
356 Collectively, in addition to markers consistent with sporozoite or merozoite formation, these
357 data point towards the sporozoite stage as being highly regulated and controlled at
358 transcriptional, translational and chromatin levels, with the mLS stages representing a release
359 of this control allowing replication, protein turn-over, reconfiguration of the proteins on the
360 plasma membrane and metabolic activity.
361 Comparison of sporozoites with HPZs does not indicate a similar release of control,
362 or at least that any release is more specific than for mLS. The sporozoite is enriched, relative
363 to HPZs, in genes such as PVP01_1258000 (*gest*), PVP01_0418000 (*sera*), PVP01_1435400
364 (*celtos*), PVP01_0835600 (*csp*) and PVP01_0602100 (*uis4*). At a broad level, sporozoite
365 enriched Pfam domains include a smaller number associated with translational
366 repression/regulation (PF00076) or DNA/RNA binding (PF01428 and PF12756).
367 Interestingly, sporozoites are enriched in Pfam domains specifically associated with
368 heterochromatin (H3K9me3) reading/interaction (PF02463, PF00628, PF13831 and
369 PF13865). Ours (see below) and previous epigenetic studies of *Plasmodium* sporozoites [27]
370 find dense heterochromatin in the telomeric to subtelomeric regions of the chromosome,
371 which is more transcriptionally active in blood-stages [49]. Others have noted an up-
372 regulation of methyl/acetyltransferases in *P. cynomolgi* HPZs [23] and/or shown
373 methyltransferase inhibitors stimulate hypnozoite activation *in vitro* [17]. The potential that
374 histone epigenetics of sporozoites has a role in or changes with liver-stage development and
375 the formation of liver schizonts or HPZs is intriguing but requires detailed study of the
376 chromatin of liver-stage parasites, which is not presently available for *P. vivax*. In contrast,
377 HPZs were enriched, relative to sporozoites, for genes including histone proteins
378 (PVP01_1138700, PVP01_1131700 and PVP01_0905900) and classic markers of metabolism
379 (PVP01_MITO3300 and PVP01_MITO3400) and *lisp2*. Pfam data indicated– largely similar
380 domain enrichment trends as were seen for the mLS stage relative to sporozoites, including a
381 number of proteosomal (PF00227, PF00112, PF03981), vesicular transport (PF00996) and
382 metabolic (PF00118, PF00268, PF01066, PF01214 a) associated functions. This supports
383 HPZs being an arrested, rather than classically 'dormant', stage with active metabolism and
384 protein turn-over. HPZs are also enriched for Pfams associated with mRNA/tRNA regulation
385 and turnover (PF04857, PF01612, PF00009 and PF01138) and glycine metabolism (PF01571
386 and PF00464) and acetyl-CoA production (PF02779 and PF00676).
387 Finally, although not the focus of this study, we looked at differential transcription
388 between mLS and HPZ stages using the Gural et al [22] data, but using the same approaches

389 as employed here. In particular, we were interested in what these comparisons might provide
390 in terms of sporozoite differentiation or development into liver schizonts or HPZs (Table
391 S14). Among mLS up-regulated transcripts are genes associated with rhoptry function (n =
392 11; including PVP01_0107500, PVP01_1469200 and PVP01_1469200), cytoadherence to
393 red-cells (PVP01_1401400 and PVP01_0734500), merozoite formation (PVP01_0728900
394 and PVP01_0612400) and exported proteins (n = 6; including PVP01_0504000,
395 PVP01_0119200 and PVP01_0801600). Consistent with *P. cynomolgi* [23], HPZ up-
396 regulated transcripts include several key sporozoite transcripts, specifically *uis4*
397 (PVP01_0602100), *puf1* (PVP01_1015000) and *speld* (PVP01_0938800). At the Pfam
398 domain level, mLS is enriched for metabolic (PF00317) and proteosomal (PF00112) domains
399 also enriched in mLS or HPZs relative to sporozoites above, as well as domains associated
400 with merozoite formation (PF12948, PF07462), rhoptry function (PF0712), DNA/RNA
401 binding (PF12756, PF10601 and PF02151) and cell division, development and DNA
402 replication (PF06705, PF00533, PF00488, PF02460, PF07034, PF02181). In contrast, HPZs
403 are enriched in Pfam domains that overlap notably with key sporozoite markers, including
404 *etramps* (PF09716) and *puf* proteins (PF00806), as well as domains associated with calcium
405 (PF08683) and nucleotide metabolism (PF06437). These data largely indicate that the
406 hypnozoite bears similarity both to the sporozoite and liver schizonts consistent with a stalled
407 stage on the path to schizont development regulated by checkpoint signals that halt/restart
408 normal schizont development, which has been proposed previously for this species [24].
409   With this is mind, we looked at transcripts that are differentially transcribed in mLS,
410 but not HPZs, relative to SPZs. There are 107 transcripts down-regulated in mLS relative to
411 SPZs that are transcribed at roughly similar levels in both SPZs and HPZs (Figure 3B). A
412 common theme among many of these genes are their role in transcriptional, post-
413 transcriptional, translational or post-translational regulation. Among transcriptional regulators
414 are transcription factors including AP2-SP2 (PVP01_0303400) and three non-AP2-like
415 transcription factors (PVP01_0306600, PVP01_0204300 and PVP01_1415800). Post-
416 transcriptional controllers include several DNA/RNA-binding proteins (PVP01_1011000.
417 PVP01_0932900, PVP01_0715300, PVP_1242600 and PVP01_0605200), RNA helicases
418 (PVP01_1403600 and PVP01_1329800) and mRNA processing (PVP01_1443100 and
419 PVP01_1458200) genes. Translational control includes several key regulators of translation
420 initiation (PVP01_1467700), tRNA processing (PVP01_0318700 and PVP01_1017700) or
421 ribosomal function/biogenesis (PVP01_1443700, PVP01_0421400, PVP01_1117200 and
422 PVP01_0215100). Post-translational control includes two methyltransferases
423 (PVP01_1428800 and PVP01_1465200), including CARM1, which methylates of H3R17
424 and, in mice, prevents differentiation in embryonic stem cells [50], and a putative histone
425 methylation reading enzyme, EEML2 (PVP01_1014100). The remaining genes in this group
426 have three noteworthy and largely overlapping themes: (1) an association with calcium
427 binding, metabolism or signalling, (2) a role in organellar metabolism and (3) homologs in
428 other organisms, including a variety of prokaryotes and eukaryotes, with key roles in
429 germination, dormancy and persistent non-replicating stages. The latter most function is
430 clearly intriguing in the context of HPZ formation and activation. These genes include a
431 homolog of dihydrolipoamide acyltransferase (aka 'sucB'), which is essential for growth in
432 *Mycobacterium tuberculosis* [51] and a key regulator in persistent *Escherichia coli* stages
433 [52]. Another example is gamete fusion factor HAP2, which, despite the name, has been
434 shown to regulate dormancy in eukaryotes ranging from plants [53, 54] to amoebae [55].
435   In addition to data for *P. vivax*, two transcriptomic studies are now available for *P.
436 cynomolgi* [27, 56] that compare mixed/schizont stage parasites with small-form
437 "hypnozoites". In comparing *P. cynomolgi* liver-stage RNA-seq and *P. vivax* liver-stage
438 microarray data, Cubi et al [23] noted a moderate to good level of agreement (R^2 = 0.50) as
439 evidence of *P. cynomolgi* being predictive and representative of *P. vivax*. However, Voorberg
440 van der Wel et al [56] explored congruence between their and the Cubi et al [23] studies and
441 found generally good agreement among schizonts and overall relatively poor agreement
442 among hypnozoites from each study. This highlights the complexity of these datasets and
443 indicates caution in comparing the current data to *P. vivax*. ApiAP2 transcription factors

8

444   feature prominently in each liver-stage transcriptomic study for *P. cynomolgi* [23, 56] and *P.*
445   *vivax* [22]. Cubi et al [23] noted an ApiAP2 (dubbed "AP2-Q"; PCYB_102390) as
446   transcriptionally up-regulated in *P. cynomolgi* hypnozoites and proposed this as a potential
447   hypnozoite marker. We note that the *P. vivax* ortholog of *Pc*-AP2-Q (PVP01_1016100) is
448   among the genes detectable as a transcript but not protein in *P. vivax* sporozoites. This may
449   point to a translationally repressed signal in sporozoites to regulate hypnozoite formation.
450   However, as *Pv*-AP2-Q is transcribed at an abundance (~50 TPM) at or below which ~50% of
451   *P. vivax* genes are detectable as transcripts but not as proteins (Figure 1B), this could as likely
452   result from LC-MS detection sensitivity. Further, although AP2-Q was reported as specific to
453   hypnozoite forming *Plasmodium* species [23], it is indeed found in non-hypnozoite producing
454   species, such as *P. knowlesi*, *P. gallinaceum* and *P. inui* [56]. Up-regulation of AP2-Q
455   transcripts is not observed for hypnozoites in subsequent transcriptomic studies of *P.*
456   *cynomolgi* [56] or *P. vivax* [22], nor do we see such an up-regulation here. Voorberg van der
457   Wel et al [56] note transcription of a range of AP2s in *P. cynomolgi* liver stages, but do not
458   find any to be up-regulated in hypnozoites. AP2s also feature among transcribed genes in *P.*
459   *vivax* liver stages, with one, PVP01_0916300, significantly up-regulated in hypnozoites. We
460   note that PVP01_0916300 is up-regulated in *P. vivax* sporozoites relative to blood-stages and
461   found in the top quartile of transcripts by abundance (TPM = 104).

463   **Chromatin epigenetics in *P. vivax* sporozoites.** As noted above, transcriptomic data for
464   sporozoites, and their comparison with liver and blood-stages, implicate histone epigenetics
465   as having an important role in sporozoite biology and liver stage differentiation. This concept
466   has been alluded to in recent liver-stage studies of *P. cynomolgi* [17, 23] that propose
467   methyltransferases as having a potential role in hypnozoite formation. No epigenetic data are
468   currently available for any *P. vivax* life-cycle stage. Studies of *P. falciparum* blood-stages
469   have identified the importance of histone modifications as a primary epigenetic regulator [57,
470   58] and characterized key markers of heterochromatin (H3K9me3) and
471   euchromatin/transcriptional activation (H3K4me3 and H3K9ac). Recently, these marks have
472   been explored with the maturation of *P. falciparum* sporozoites in the mosquito [26]. Here,
473   we characterize major histone marks in *P. vivax* sporozoites and assess their relationship to
474   transcript abundance.

476   *Histone modifications in* **P. vivax** *sporozoites.* Using ChIP-seq, we identified 1,506, 1,999
477   and 5,262 ChIP-seq peaks stably represented in multiple *P. vivax* sporozoite replicates and
478   associated with H3K9me3, H3K9ac and H3K4me3 histone marks respectively (Figure 4 and
479   S14-S19). Peak width, spacing and stability differed with histone mark type (Figures S15 and
480   S16). H3K4me3 covered the greatest breadth of the genome (36.0% of all bases) and was the
481   most stable among replicates, with ~84% of bases associated with an H3K4me3 found in
482   multiple biological replicates. By comparison H3K9me3 marks were least stable, with 46% of
483   bases associated with this mark found in just one replicate. Consistent with observations in *P.*
484   *falciparum*, H3K9me3 'heterochromatin' marks primarily clustered in telomeric and
485   subtelometric regions (**Figure 4**). In contrast, the 'euchromatin' / transcriptionally open
486   histone marks, H3K4me3 and H3K9ac, were distributed in chromosome central regions and
487   did not overlap with regions under H3K9me3 suppression. Both H3K9me3 and H3K4me3
488   marks were reasonably uniformly distributed (mean peak spacing ~500bp for each) within
489   their respective regions of the genome. In contrast, H3K9ac peaks were spaced further apart
490   (mean: ~2kb), but also with a greater variability in spacing (likely reflecting their association
491   with promoter regions [59]; Figure S17 and S19). The instability of H3K9me3 may reflect its
492   use in *Plasmodium* for regulating expression of contingency genes from multigene families,
493   whose members have overlapping and redundant functions [49] and confer phenotypic
494   plasticity [60].

496   *Genes under histone regulation*. We explored an association between these histone marks
497   and the transcriptional behaviour of protein coding genes (Figure 4 and S19 and Tables S15-
498   S20). 485 coding genes stably intersected with an H3K9me3 mark; all are located near the

9

499   ends of the chromosomal scaffolds (i.e., are (sub)telomeric). On average, these genes are
500   transcribed at ~30 fold lower levels (mean 0.7 TPM) than genes not stably intersected by
501   H3K9me3 marks. These data clearly support the function of this mark in transcriptional
502   silencing. This is largely consistent with observations in *P. falciparum* sporozoites [26],
503   however, in contrast to *P. falciparum* sporozoites where a single *var* gene was described to
504   lack heterochromatin structure [27] we observe no genes within heterochromatin dense region
505   that lacked a stable H3K9me3 signal or were transcribed at notable levels (i.e., above ~5
506   TPM). Whether this relates to differences in epigenetic control between the species is not
507   clear. We note that (sub)telomeric genes are overall transcriptionally silent in *P. vivax*
508   sporozoites relative to blood-stages (Figure 4 and Tables S21 and S22). Consistent with
509   observations in *P. falciparum* [57], the bulk of these genes include complex protein families,
510   such as *vir* and *Pv-fam* genes, which are so far described to function primarily in blood-
511   stages. Also notable among the genes are several reticulocyte-binding proteins, including
512   RBP2, 2a, 2b and 2c. This transcriptional silence in telomeric and subtelomeric regions was
513   recently observed in *P. falciparum* sporozoites [27].

514   Outside of the telomeres and subtelomeres, H3K4me3 marks are stably associated
515   with the intergenic regions of 3,676 genes. H3K9ac marks are also identified within 1kb of
516   the transcriptional start site (TSS) of 1,284 coding genes, with 1085 of these stably marked
517   also by H3K4me3 (Figure 4B). The average transcription of these genes is 50, 112 and 112
518   TPMs respectively (72, 160 and 160-fold higher than H3K9me3 marked genes). Gene-by-
519   gene observations show that H3K9ac and H3K4me3 marks cluster densely in the 1000kb up
520   and down-stream of the start and stop codon respectively of transcribed genes, but are much
521   less dense within coding regions of these genes (Figure S15). This pattern directly correlates
522   with transcription and contrasts H3K9me3 marks, which are distributed across the length of
523   the gene at even density and are correlated with a lack of transcription. These data support the
524   role of H3K9ac and H3K4me3 in transcriptional activation in *P. vivax*. The lower
525   transcriptional abundance of H3K4me3 marked, compared with H3K9ac or H3K9ac and
526   H3K4me3 marked genes suggest these marks work synergistically and that H3K9ac is
527   possibly the better of the two, as a single mark indicator of transcriptional activity in *P. vivax*.
528   This is consistent with recent observations in *P. falciparum* sporozoites [26].

529   Interestingly, H3K9ac-marked genes ranged in transcriptional activity from the most
530   abundantly transcribed genes to many in the lower 50% and even lowest decile of
531   transcription. This suggests more contributes to transcriptional activation in *P. vivax*
532   sporozoites than, simply, gene accessibility through chromatin regulation. Specific activation
533   by a transcription factor (e.g., ApiAP2s [61]) is the obvious candidate. To explore this, we
534   compared upstream regions (within 1kb of the TSS or up to the 3' end of the next gene
535   upstream, whichever was less) of highly (top 10%) and lowly (bottom 10%) transcribed
536   H3K9ac marked genes for over-represented sequence motifs in the highly expressed genes
537   that might coincide with known ApiAP2 transcription factor binding sites [62]. We identified
538   these based on the location of the nearest stable H3K9ac peak relative to the transcription
539   start site for each gene (Figure S12). In most instances, these peaks were within 100bp of the
540   TSS and, consistent with data from *P. falciparum* [59], *P. vivax* promoters appear to be no
541   more than a few hundred to a maximum of 1000 bp upstream of the TSS. Exploring these
542   regions, we identified two over-represented motifs: TGTACMA (e-value $2.7e^{-2}$) and
543   ATATTTH (e-value $3.3e^{-3}$) (Fig. 2D). TGTAC is consistent with the known binding site for
544   *Pf*-AP2-G, which regulates sexual differentiation in gametocytes [63], but its *P. vivax*
545   ortholog (PVP01_1418100) is neither highly transcribed nor expressed in sporozoites.
546   ATATTTH is similar to the binding motif for *Pf*-AP2-L (AATTTCC), a transcription factor
547   that is important for liver stage development in *P. berghei* [64]. In contrast to AP2-G, *Pv*-
548   AP2-L (PVX_081180) is in the top 10% of transcription and expression in *P. vivax*
549   sporozoites and up-regulated relative to blood-stages. In *P. vivax* sporozoites, the ATATTTH
550   motif is associated with a number of highly transcribed genes, including *lisp1* and *uis2-4*,
551   known to be regulated by AP2-L in *P. berghei* [64] as well as many of the most highly
552   transcribed, H3K9ac marked genes, including two *etramps* (PVP01_0734800 and
553   PVP01_0504800), several RNA-binding proteins, including *Puf2*, *ddx5*, a putative ATP-

554 dependent RNA helicase DBP1 (PVP01_1429700), and a putative *bax1* inhibitor
555 (PVP01_1465600). Interestingly, a number of highly transcribed and translationally repressed
556 genes associated with the ATATTTH motif, including *uis4*, *siap2* and *pv1*, are not stably
557 marked by H3K9ac in all replicates (i.e., there is significant variation in the placement of the
558 H3K9ac peak or their presence/absence among replicates for these genes). It may be that
559 additional histone modifications, for example H3K27me, H3R17me3 or H2A or H4
560 modifications, are involved in regulating transcription of these genes. Certainly the
561 incorporation of the H2A.Z histone variant, which is present in intergenic regions of *P.*
562 *falciparum (*Petter et al 2011*)*, and controls temperature responses in plants [65] is intriguing
563 as a potential mark regulating sporozoite fate in *P. vivax* considering the association between
564 hypnozoite activation rate and climate [11], as is H3R17me3 in consideration of the
565 enrichment of markers/readers of this modification in HPZs noted above and the role of this
566 mark in cell fate progression in other species [50].
567
CONCLUSIONS
568
569 We provide the first comprehensive study of the transcriptome and epigenome of mature
570 *Plasmodium vivax* sporozoites and undertake detailed comparisons with recently published
571 proteomic data for *P. vivax* sporozoites [33] and transcriptomic data for *P. vivax* mixed and
572 hypnozoite-enriched liver-stages [22] and mixed blood-stages [40]. These data support the
573 proposal that the sporozoite is a highly-programmed stage that is primed for invasion of and
574 development in the host hepatocyte. Cellular regulation, including at transcription,
575 translational and epigenetic levels, appears to play a major role in shaping this stage (which
576 continues on in some form in hypnozoites), and many of the genes proposed here as being
577 under translational repression are involved in hepatocyte infection and early liver-stage
578 development (Figure 5). We highlight a major role for RNA-binding proteins, including
579 PUF2, ALBA2/4 and, intriguingly, 'Homologue of Musashi' (HoMu). We find that
580 transcriptionally, the hypnozoite appears to be a transition point between the sporozoite and
581 replicating schizonts, having many of the dominant sporozoite transcripts and retaining high
582 transcription of a number of key regulatory pathways involved in transcription, translation
583 and chromatin configuration (including histone arginine methylation). A consistent theme in
584 the study is the prominence of a number of genes that have a role in numerous eukaryotic
585 systems in cell fate determination and differentiation (e.g., HoMu, Yippee and CARM1) and
586 overlap with dormancy and/or persistent cell states in bacteria, protists or higher eukaryotes
587 (e.g., bacterial sucB and gamete fusion protein HAP2). These data do not point to one single
588 programming switch for dormancy or liver developmental fate in *P. vivax*, but present a
589 number of intriguing avenues for exploration in subsequent studies, particularly in model
590 species such as *P. cynomolgi*. Our study contributes to understanding the early stages of
591 hepatocyte infection and the developmental switch between liver trophozoite and hypnozoite
592 formation. We also identify potential avenues for rationally prioritizing targets underpinning
593 liver-stage differentiation for functional evaluation in humanized mouse and simian models
594 for relapsing *Plasmodium* species and identifying novel avenues to understand and eradicate
595 liver-stage infections.
596
MATERIALS AND METHODS
597
598 **Ethics Statement.** Collection of venous blood from human patients with naturally acquired
599 vivax infection for the current study was approved by the Ethical Review Committee of the
600 Faculty of Tropical Medicine, Mahidol University (Human Subjects Protocol number TMEC
601 11-033) with the informed written consent of each donor individual. All mouse tissue used in
602 the current study was from preserved infected tissues generated previously [13]. All mouse
603 infection work in [13] was carried out at the Centre for Infectious Diseases Research (CIDR)
604 in Seattle, USA, under direct approval of the CIDR Institutional Animal Care and Use
605 Committee (IACUC) and performed in strict accordance with the recommendations in the
606 Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, USA.
607 The Centre for Infectious Disease Research Biomedical Research Institute has an Assurance

608   from the Public Health Service (PHS Assurance number is A3640-01) through the Office of
609   Laboratory Animal Welfare (OLAW) for work approved by its IACUC.
610
611   **Material collection, isolation and preparation.** Nine field isolates (PvSpz-Thai 1 to 9),
612   representing symptomatic blood-stage malaria infections were collected as venous blood (20
613   mL) from patients presenting at malaria clinics in Tak and Ubon Ratchatani provinces in
614   Thailand. Each isolate was used to establish, infections in *Anopheles dirus* colonized at
615   Mahidol University (Bangkok) by membrane feeding [13], after 14-16 days post blood
616   feeding, ~3-15 million sporozoites were harvested per field isolate from the salivary glands of
617   up to 1,000 of these mosquitoes as per [66] and shipped in preservative (trizol (RNA/DNA) or
618   1% paraformaldehyde (DNA for ChiP-seq) to the Walter and Eliza Hall Institute (WEHI).
619
620   **Transcriptomics sequencing and differential analysis.** Upon arrival at WEHI, messenger
621   RNAs were purified from an aliquot (~0.5-1 million sporozoites) of each *P. vivax* field isolate
622   as per [40] and subjected to RNA-seq on Illumina NextSeq using TruSeq library construction
623   chemistry as per the manufacturer's instructions. Raw reads for each RNA-seq replicate are
624   available through the Sequence Read Archive (XXX-XXX). Sequencing adaptors were
625   removed and low quality reads trimmed and filtered using Trimmomatic v. 0.36 [67]. To
626   remove host contaminants, processed reads were aligned, as single-end reads, to the
627   *Anopholes dirus* wrari2 genome (VectorBase version W1) using Bowtie2 [68] (--very-
628   sensitive preset). All non-host reads were then aligned to the manually curated transcripts of
629   the *P. vivax* P01 genome (http://www.genedb.org/Homepage/P*vivax*P01; [28]) using RSEM
630   [69] (pertinent settings: --bowtie2 --bowtie2-sensitivity-level very_sensitive --calc-ci --ci-
631   memory 10240 --estimate-rspd --paired-end). Transcript abundance for each gene in each
632   replicate was calculated by RSEM as raw count, posterior mean estimate expected counts
633   (pme-EC) and transcripts per million (TPM).
634   Transcriptional abundance in *P. vivax* sporozoites was compared qualitatively (by
635   ranked abundance) with previously published microarray data for *P. vivax* salivary-gland
636   sporozoites [24]. As a further quality control, these RNA-seq data were compared also with
637   previously published microarray data for *P. falciparum* salivary-gland sporozoites [30], as
638   well as RNA-seq data from salivary-gland sporozoites generated here for *P. falciparum*
639   (single replicate generated from *P. falciparum* 3D7 lab cultures isolated from *Anopholes*
640   *stephensi* and processed as above) and previously published for *P. yoelii* [29]. RNA-seq data
641   from these additional *Plasmodium* species were (re)analysed from raw reads and
642   transcriptional abundance for each species was determined (raw counts and pme-EC and TPM
643   data) as described above using gene models current as of 04-10-2016 (PlasmoDB release
644   v29). Interspecific transcriptional behaviour was qualitatively compared by relative ranked
645   abundance in each species using TPM data for single copy orthologs (SCOs; defined in
646   PlasmoDB) only, shared between *P. vivax* and *P. faliciparum* or shared among *P. vivax*, *P.*
647   *falciparum* and *P. yoelii*.
648   To define transcripts that were up-regulated in sporozoites, we remapped raw reads
649   representing early (18-24 hours post-infection (HPI)), mid (30-40 HPI) and late (42-46 HPI)
650   *P. vivax* blood-stage infections recently published by Zhu *et al* [40] to the *P. vivax* P01
651   transcripts using RSEM as above. All replicate data was assessed for mapping metrics,
652   transcript saturation and other standard QC metrics using QualiMap v 2.1.3 [70]. Differential
653   transcription between *P. vivax* salivary-gland sporozoites and mixed blood-stages [40] was
654   assessed using pme-EC data in EdgeR [71] and limma [72] (differential transcription cut-off:
655    $\geq$ 2-fold change in counts per million (CPM) and a False Discovery Rate (FDR) $\leq$ 0.05).
656   Pearson Chi squared tests were used to detect over-represented Pfam domains and Gene
657   Ontology (GO) terms among differentially transcribed genes in sporozoites (Bonferroni-
658   corrected $p < 0.05$), based on gene annotations in PlasmoDB (release v29).
659   We also compared transcription of the sporozoite stages to recently published liver-
660   stage data from Gural et al [22] as per the sporozoite to blood-stage comparisons above, with
661   the following modifcations: (1) EC values were normalized using the 'upper quartile' method
662   instead of TMM, (2) differential transcription was assessed using a quasi-likelihood

12

663 generalize linear model (instead of a linear model) and (3) an FDR threshold for significance
664 of $\leq$ 0.01 was used instead of $\leq$ 0.05. These differences related to specific attributes of the
665 liver-stage dataset, particularly the small number of replicates (n = 2) per experiment
666 condition. Data visualization and interactive R-shiny plots were produced in R using the
667 ggplot[73], ggplot2 [74], gplots(heatmap.2) [75] and Glimma [76] packages.
668
669 **Assessment of Sporozoite RNA-seq transcriptome by selective RT-qPCR:** Extracted RNA
670 was DNase treated (Sigma D5307) as per manufacturer recommendations. RNA was
671 quantified using the TapeStation High Sensitivity RNA kit (Agilent). Two intron-spanning
672 primer pairs were designed per gene of interest using Primer3 and BLAST. Primer pairs were
673 tested in two concentrations (0.75ng and 2.83ng per reaction) to determine efficiency and
674 specificity. Product was run on a 1% agarose gel with ethidium bromide. Primer pairs
675 indicating non-specific priming were removed. The resulting 11 primer pairs were used on
676 four sporozoite samples; VUBR06, VUNL23, VUBR24, VTTY84. RNA was reverse
677 transcribed (Sensifast, Bioline) and used at 0.75ng per reaction, run on a Roche LightCycler
678 480 II. Melt curves were assessed and products were run on a gel to ensure specificity again.
679 Cp threshold was set automatically. ΔCp value was calculated as target gene – comparator
680 gene (SERA and CelTOS were used). Data were log transformed and fold change calculated.
681
682 RT-qPCR Primers were as follows:

Name	Gene	Forward Primer	Reverse Primer
RPS27	PVX_122245	ACCACCTTGTTTAGCCATGC	TAATTTGCACTTTCCACCCGTT
D13	PVX_089510	CTGTACACGCACGAGCTGGC	CAGCTCCTTGACGCCACTG
G10	PVX_080110	ACGAGCTGTACTACAAGCGGA	TTTCTCCTGCACCAGGTAGTC
AP2	PVX_086995*	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

683
684 * denotes single exon gene
685
686 **Salivary-gland sporozoite and liver-stage immunofluorescence assays (IFAs).** IFAs were
687 performed as per [13] using preserved, vivax infected mouse liver tissue generated previously
688 for that study. In [13], female FRG [fumarylacetoacetate hydrolase (F), recombination
689 activation gene 2 (R), interleukin-2 receptor subunit gamma (G)] triple KO mice engrafted
690 with human hepatocytes (FRG KO huHep) were purchased from Yecuris Corporation
691 (Oregon, USA). Mice were infected through intravenous injection into the tail with $3.5 \times$
692 10^5 to 1×10^6 sporozoites isolated from the salivary glands of infected mosquitoes in 100 μl
693 of RPMI media. Liver stages for the current study were obtained from 10μm formalin fixed
694 paraffin embedded day 7 liver stages generated previously [13] from FRG knockout huHep
695 mice;[13] these were deparaffinized prior to staining. Fresh salivary-gland sporozoites were
696 fixed in acetone per [13]. All cells were incubated twice for 3 minutes in Xylene, then 100%
697 Ethanol, and finally once for 3 minutes each in 95%, 70%, and 50% Ethanol. The cells were
698 rinsed in DI water and permeabilized immediately in 1XTBS, containing Triton X-100 and
699 30% hydrogen peroxide. The cells were blocked in 5% milk in 1XTBS. The hepatocytes were
700 stained overnight with a rabbit polyclonal LISP1 antibody (A), a rabbit polyclonal UIS4

701 antibody (B), and a rabbit polyclonal BIP antibody (C) in blocking buffer. The cells were
702 washed with 1XTBS and the primary antibodies were detected with goat anti-rabbit Alexa
703 Fluor 488 antibody (Life Technologies). The cells were washed in 1XTBS. The hepatocytes
704 were rinsed in KMNO4 and washed in 1XTBS. The cells were incubated in DAPI for 5
705 minutes.
706
707 **Histone ChIP sequencing and analysis.** Aliquots of 2 – 6 million freshly isolated
708 sporozoites were fixed with 1% paraformaldehyde for 10 min at 37°C and the reaction
709 subsequently quenched by adding glycine to a final concentration of 125 mM. After three
710 washes with PBS, sporozoite pellets were stored at -80°C and shipped to Australia. Nuclei
711 were released from the sporozoites by dounce homogenization in lysis buffer (10 mM Hepes
712 pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EDTA, 1 mM DTT, 1x EDTA-free protease
713 inhibitor cocktail (Roche), 0.25% NP40). Nuclei were pelleted by centrifugation at 21,000 g
714 for 10 min at 4°C and resuspended in SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris
715 pH 8.1, 1x EDTA-free protease inhibitor cocktail). Chromatin was sheared into 200–1000 bp
716 fragments by sonication for 16 cycles in 30 sec intervals (on/off, high setting) using a
717 Bioruptor (Diagenode) and diluted 1:10 in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-
718 100, 1.2 mM EDTA, 16.7 mM Tris pH 8.1, 150 mM NaCl). Chromatin was precleared for 1
719 hour with protein A/G sepharose (4FastFlow, GE Healthcare) equilibrated in 0.1% BSA
720 (Sigma-Aldrich, USA) in ChIP dilution buffer. Chromatin from 3×10^5 nuclei was taken
721 aside as input material. Chromatin from approximately 3×10^6 sporozoite nuclei was used for
722 each ChIP. ChIP was carried out over night at 4°C with 5 µg of antibody (H3K9me3 (Active
723 Motif), H3K4me3 (Abcam), H3K9ac (Upstate), H4K16ac (Abcam)) and 10 µl each of
724 equilibrated protein A and G sepharose beads (4FastFlow, GE Healthcare). After washes in
725 low-salt, high-salt, LiCl, and TE buffers (EZ-ChIP Kit, Millipore), precipitated complexes
726 were eluted in 1% SDS, 0.1 M NaHCO$_3$. Cross-linking of the immune complexes and input
727 material was reversed for 6 hours at 45°C after addition of 500 mM NaCl and 20 µg/ml of
728 proteinase K (NEB). DNA was purified using the MinElute® PCR purification kit (Qiagen)
729 and paired-end sequenced on Illumina NextSeq using TruSeq library construction chemistry
730 as per the manufacturer's instructions. Raw reads for each ChIP-seq replicate are available
731 through the Sequence Read Archive (XXX-XXX).
732 Fastq files were checked for quality using fastqc
733 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapter sequences were
734 trimmed using cutadapt [77]. Paired end reads were mapped to the *P. vivax* P01 strain
735 genome annotation using Bowtie2 [68]. The alignment files were converted to Bam format,
736 sorted and indexed using Samtools [78]. ChIP peaks were called relative to input using
737 MACS2[79] in paired end mode with a q value less than or equal to 0.01. Peaks and peak
738 summits were converted to sorted BED files. Bedtools intersect[80] was used to identify
739 genes that intersected H3K9me3 peaks and Bedtools closest was used to identify genes that
740 were closest to and downstream of H3K9ac and H3K4me3 peak summits.
741
742 **Sequence motif analysis.** Conserved sequence motifs were identified using the program
743 DREME [81]. Only genes in the top decile of transcription showing no evidence of protein
744 expression in multiple salivary-gland sporozoite replicates were considered as putatively
745 translationally repressed (n = 170). We queried coding regions and regions upstream of the
746 transcriptional start site (TSS) for each gene, defined by Zhu *et al* [40] and/or predicted here
747 from all RNA-seq data using the Tuxedo suite [82], for enriched sequence motifs in
748 comparison to 170 genes found to be in the top decile of both transcriptional and expressional
749 abundance in the same sporozoite replicates. In searching for motifs associated with highly
750 transcribed genes with stable H3K9ac marks within 1kb of the TSS (or up to the 3' end of the
751 next gene upstream), we compared H3K9ac marked genes in the top decile of transcription to
752 the same number of H3K9ac marked genes in the bottom decile of transcription. In both
753 instances, an e-value threshold of 0.05 was considered the minimum threshold for statistical
754 significance.
755

756 **References**
757 1.   Organization WH. World Malaria Report 2015. WHO, Geneva. 2015.
758 2.   Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, et al.
759   Shrinking the malaria map: progress and prospects. Lancet. 2010;376
760   9752:1566-78. doi:10.1016/S0140-6736(10)61270-6.
761 3.   Price RN, Douglas NM and Anstey NM. New developments in *Plasmodium vivax*
762   malaria: severe disease and the rise of chloroquine resistance. Curr Opin Infect
763   Dis. 2009;22 5:430-5. doi:10.1097/QCO.0b013e32832f14c1.
764 4.   Baird KJ. Malaria caused by *Plasmodium vivax*: recurrent, difficult to treat,
765   disabling, and threatening to life - averting the infectious bite preempts these
766   hazards. Pathogens and global health. 2013;107:475-9.
767 5.   Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J and Cui L. *Plasmodium
768   vivax* transmission: chances for control? Trends in parasitology. 2004;20 4:192-
769   8.
770 6.   Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, et al. Key gaps
771   in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite.
772   Lancet Infect Dis. 2009;9 9:555-66.
773 7.   Lindner SE, Miller JL and Kappe SH. Malaria parasite pre-erythrocytic infection:
774   preparation meets opportunity. Cell Microbiol. 2012;14 3:316-24.
775 8.   Shin SC, Vanderberg JP and Terzakis JA. Direct infection of hepatocytes by
776   sporozoites of *Plasmodium berghei*. J Protozool. 1982;29 3:448-54.
777 9.   Mota MM, Pradel G, Vanderberg JP, Hafalla JC, Frevert U, Nussenzweig RS, et al.
778   Migration of *Plasmodium* sporozoites through cells before infection. Science.
779   2001;291 5501:141-4.
780 10.   Lysenko AJ, Beljaev A and Rybalka V. Population studies of *Plasmodium vivax*: 1.
781   The theory of polymorphism of sporozoites and epidemiological phenomena of
782   tertian malaria. Bull WHO. 1977;55 5:541.
783 11.   White NJ. Determinants of relapse periodicity in *Plasmodium vivax* malaria.
784   Malar J. 2011;10:297.
785 12.   Price RN, Tjitra E, Guerra CA, Yeung S, White NJ and Anstey NM. Vivax malaria:
786   neglected and not benign. Amer J Trop Med Hyg. 2007;77 6 Suppl:79-87.
787 13.   Mikolajczak SA, Vaughan AM, Kangwanrangsan N, Roobsoong W, Fishbaugher M,
788   Yimamnuaychok N, et al. *Plasmodium vivax* liver stage development and
789   hypnozoite persistence in human liver-chimeric mice. Cell Host Microbe.
790   2015;17 4:526-35.
791 14.   Mueller A-K, Camargo N, Kaiser K, Andorfer C, Frevert U, Matuschewski K, et al.
792   *Plasmodium* liver stage developmental arrest by depletion of a protein at the
793   parasite–host interface. Proc Nat'l Acad Sci USA. 2005;102 8:3022-7.
794 15.   Silvie O, Briquet S, Muller K, Manzoni G and Matuschewski K. Post-
795   transcriptional silencing of UIS4 in *Plasmodium berghei* sporozoites is important
796   for host switch. Mol Microbiol. 2014;91 6:1200-13.
797 16.   Mackellar DC, O'Neill MT, Aly AS, Sacci JB, Jr., Cowman AF and Kappe SH.
798   *Plasmodium falciparum* PF10_0164 (ETRAMP10.3) is an essential
799   parasitophorous vacuole and exported protein in blood stages. Eukaryot Cell.
800   2010;9 5:784-94.
801 17.   Dembele L, Franetich JF, Lorthiois A, Gego A, Zeeman AM, Kocken CH, et al.
802   Persistence and activation of malaria hypnozoites in long-term primary
803   hepatocyte cultures. Nat Med. 2014;20 3:307-12.
804 18.   Malmquist NA, Moss TA, Mecheri S, Scherf A and Fuchter MJ. Small-molecule
805   histone methyltransferase inhibitors display rapid antimalarial activity against
806   all blood stage forms in *Plasmodium falciparum*. Proc Nat'l Acad Sci USA.
807   2012;109 41:16708-13.
808 19.   Josling GA and Llinas M. Sexual development in *Plasmodium* parasites: knowing
809   when it's time to commit. Nat Rev Microbiol. 2015;13 9:573-87.

810   20.   White MT, Karl S, Battle KE, Hay SI, Mueller I and Ghani AC. Modelling the
811   contribution of the hypnozoite reservoir to *Plasmodium vivax* transmission. Elife.
812   2014;3.
813   21.   Roobsoong W, Tharinjaroen CS, Rachaphaew N, Chobson P, Schofield L, Cui L, et
814   al. Improvement of culture conditions for long-term in vitro culture of
815   *Plasmodium vivax*. Malaria J. 2015;14 1:1.
816   22.   Gural N, Mancio-Silva L, Miller AB, Galstian A, Butty VL, Levine SS, et al. In vitro
817   culture, drug sensitivity, and transcriptome of *Plasmodium vivax* hypnozoites.
818   Cell Host Microbe. 2018;23 3:395-406 e4.
819   23.   Cubi R, Vembar SS, Biton A, Franetich JF, Bordessoulles M, Sossau D, et al. Laser
820   capture microdissection enables transcriptomic analysis of dividing and
821   quiescent liver stages of *Plasmodium* relapsing species. Cell Microbiol. 2017.
822   24.   Westenberger SJ, McClean CM, Chattopadhyay R, Dharia NV, Carlton JM,
823   Barnwell JW, et al. A systems-based analysis of *Plasmodium vivax* lifecycle
824   transcription from human to mosquito. PLoS Negl Trop Dis. 2010;4 4:e653.
825   25.   Kim A, Popovici J, Vantaux A, Samreth R, Bin S, Kim S, et al. Characterization of P.
826   vivax blood stage transcriptomes from field isolates reveals similarities among
827   infections and complex gene isoforms. Sci Rep. 2017;7 1:7761.
828   26.   Gomez-Diaz E, Yerbanga RS, Lefevre T, Cohuet A, Rowley MJ, Ouedraogo JB, et al.
829   Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression
830   during development in *Anopheles gambiae*. Sci Rep. 2017;7:40655.
831   27.   Zanghi G, Vembar SS, Baumgarten S, Ding S, Guizetti J, Bryant JM, et al. A specific
832   PfEMP1 is expressed in *P. falciparum* sporozoites and plays a role in hepatocyte
833   infection. Cell Rep. 2018;22 11:2951-63.
834   28.   Auburn S, Bohme U, Steinbiss S, Trimarsanto H, Hostetler J, Sanders M, et al. A
835   new *Plasmodium vivax* reference sequence with improved assembly of the
836   subtelomeres reveals an abundance of pir genes. Wellcome Open Res. 2016;1:4.
837   29.   Lindner SE, Mikolajczak SA, Vaughan AM, Moon W, Joyce BR, Sullivan WJ, Jr., et
838   al. Perturbations of *Plasmodium* Puf2 expression and RNA-seq of Puf2-deficient
839   sporozoites reveal a critical role in maintaining RNA homeostasis and parasite
840   transmissibility. Cell Microbiol. 2013;15 7:1266-83.
841   30.   Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, et al. Global
842   analysis of transcript and protein levels across the *Plasmodium falciparum* life
843   cycle. Genome research. 2004;14 11:2308-18.
844   31.   Mikolajczak SA, Silva-Rivera H, Peng X, Tarun AS, Camargo N, Jacobs-Lorena V, et
845   al. Distinct malaria parasite sporozoites reveal transcriptional changes that
846   cause differential tissue infection competence in the mosquito vector and
847   mammalian host. Mol Cell Biol. 2008;28 20:6196-207.
848   32.   Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative
849   genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature.
850   2008;455 7214:757-63.
851   33.   Swearingen KE, Lindner SE, Flannery EL, Vaughan AM, Morrison RD,
852   Patrapuvich R, et al. Proteogenomic analysis of the total and surface-exposed
853   proteomes of *Plasmodium vivax* salivary gland sporozoites. PLoS Negl Trop Dis.
854   2017;11 7:e0005791.
855   34.   Guerreiro A, Deligianni E, Santos JM, Silva PA, Louis C, Pain A, et al. Genome-wide
856   RIP-Chip analysis of translational repressor-bound mRNAs in the *Plasmodium*
857   gametocyte. Genome Biol. 2014;15 11:493.
858   35.   Silvie O, Briquet S, Müller K, Manzoni G and Matuschewski K.
859   Post-transcriptional silencing of UIS4 in Plasmodium berghei sporozoites is
860   important for host switch. Molecular microbiology. 2014;91 6:1200-13.
861   36.   Lindner SE, Swearingen KE, Harupa A, Vaughan AM, Sinnis P, Moritz RL, et al.
862   Total and putative surface proteomics of malaria parasite salivary gland
863   sporozoites. Mol Cell Proteomics : MCP. 2013;12 5:1127-43.

864   37.   Kelley KD, Miller KR, Todd A, Kelley AR, Tuttle R and Berberich SJ. YPEL3, a p53-
865   regulated gene that induces cellular senescence. Cancer Res. 2010;70 9:3566-75.
866   38.   Tuttle R, Simon M, Hitch DC, Maiorano JN, Hellan M, Ouellette J, et al. Senescence-
867   associated gene YPEL3 is downregulated in human colon tumors. Ann Surg
868   Oncol. 2011;18 6:1791-6.
869   39.   Roth A, Adapa SR, Zhang M, Liao X, Saxena V, Goffe R, et al. Unraveling the
870   *Plasmodium vivax* sporozoite transcriptional journey from mosquito vector to
871   human host. Sci Rep. 2018;8 1:12183.
872   40.   Zhu L, Mok S, Imwong M, Jaidee A, Russell B, Nosten F, et al. New insights into
873   the *Plasmodium vivax* transcriptome using RNA-Seq. Sci Rep. 2016;6:20498.
874   41.   Kramer S. RNA in development: how ribonucleoprotein granules regulate the life
875   cycles of pathogenic protozoa. WIR: RNA. 2014;5 2:263-84.
876   42.   Tucker RP. The thrombospondin type 1 repeat superfamily. Int J Biochem Cell
877   Biol. 2004;36 6:969-74.
878   43.   Ntumngia FB, Bouyou-Akotet MK, Uhlemann AC, Mordmuller B, Kremsner PG
879   and Kun JF. Characterisation of a tryptophan-rich *Plasmodium falciparum*
880   antigen associated with merozoites. Mol Biochem Parasitol. 2004;137 2:349-53.
881   44.   Gubbels MJ, Vaishnava S, Boot N, Dubremetz JF and Striepen B. A MORN-repeat
882   protein is a dynamic component of the *Toxoplasma gondii* cell division
883   apparatus. J Cell Sci. 2006;119 Pt 11:2236-45.
884   45.   Aly AS, Lindner SE, MacKellar DC, Peng X and Kappe SH. SAP1 is a critical post-
885   transcriptional regulator of infectivity in malaria parasite sporozoite stages. Mol
886   Microbiol. 2011;79 4:929-39.
887   46.   Okano H, Imai T and Okabe M. Musashi: a translational regulator of cell fate.
888   Journal of Cell Science. 2002;115 7:1355-9.
889   47.   Cui L, Lindner S and Miao J. Translational regulation during stage transitions in
890   malaria parasites. Annals N Y Acad Sci. 2015;1342 1:1-9.
891   48.   Lasko P. Gene regulation at the RNA layer: RNA binding proteins in intercellular
892   signaling networks. Sci STKE. 2003;179:RE6.
893   49.   Guizetti J and Scherf A. Silence, activate, poise and switch! Mechanisms of
894   antigenic variation in *Plasmodium falciparum.* Cell microbiol. 2013;15 5:718-26.
895   50.   Wu Q, Bruce AW, Jedrusik A, Ellis PD, Andrews RM, Langford CF, et al. CARM1 is
896   required in embryonic stem cells to maintain pluripotency and resist
897   differentiation. Stem Cells. 2009;27 11:2637-45.
898   51.   Shi S and Ehrt S. Dihydrolipoamide acyltransferase is critical for *Mycobacterium*
899   *tuberculosis* pathogenesis. Infect Immun. 2006;74 1:56-63.
900   52.   Ma C, Sim S, Shi W, Du L, Xing D and Zhang Y. Energy production genes sucB and
901   ubiF are involved in persister survival and tolerance to multiple antibiotics and
902   stresses in Escherichia coli. FEMS Microbiol Lett. 2010;303 1:33-40.
903   53.   Schrader J, Moyle R, Bhalerao R, Hertzberg M, Lundeberg J, Nilsson P, et al.
904   Cambial meristem dormancy in trees involves extensive remodelling of the
905   transcriptome. Plant J. 2004;40 2:173-87.
906   54.   Yazawa K and Kamada H. Identification and characterization of carrot HAP
907   factors that form a complex with the embryo-specific transcription factor C-
908   LEC1. J Exp Bot. 2007;58 13:3819-28.
909   55.   Wood FC, Heidari A and Tekle YI. Genetic Evidence for Sexuality in
910   Cochliopodium (Amoebozoa). J Hered. 2017;108 7:769-79.
911   56.   Voorberg-van der Wel A, Roma G, Gupta DK, Schuierer S, Nigsch F, Carbone W, et
912   al. A comparative transcriptomic analysis of replicating and dormant liver stages
913   of the relapsing malaria parasite *Plasmodium cynomolgi.* Elife. 2017;6
914   57.   Lopez-Rubio J-J, Mancio-Silva L and Scherf A. Genome-wide analysis of
915   heterochromatin associates clonally variant gene regulation with perinuclear
916   repressive centers in malaria parasites. Cell host & microbe. 2009;5 2:179-90.

917 58. Duffy MF, Selvarajah SA, Josling GA and Petter M. Epigenetic regulation of the
918 *Plasmodium falciparum* genome. Brief Funct Genomics. 2014;13 3:203-16.
919 59. Cui L, Miao J, Furuya T, Li X, Su XZ and Cui L. PfGCN5-mediated histone H3
920 acetylation plays a key role in gene expression in *Plasmodium falciparum*.
921 Eukaryot Cell. 2007;6 7:1219-27.
922 60. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, de Pouplana LR, et al.
923 Transcriptional variation in the malaria parasite *Plasmodium falciparum*.
924 Genome research. 2012;22 5:925-38.
925 61. De Silva EK, Gehrke AR, Olszewski K, León I, Chahal JS, Bulyk ML, et al. Specific
926 DNA-binding by apicomplexan AP2 transcription factors. Proc Nat'l Acad Sci.
927 2008;105 24:8393-8.
928 62. Painter HJ, Campbell TL and Llinás M. The Apicomplexan AP2 family: integral
929 factors regulating *Plasmodium* development. Mol Biochem Parasitol. 2011;176
930 1:1-7.
931 63. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, et
932 al. A transcriptional switch underlies commitment to sexual development in
933 human malaria parasites. Nature. 2014;507 7491:248.
934 64. Iwanaga S, Kaneko I, Kato T and Yuda M. Identification of an AP2-family protein
935 that is critical for malaria liver stage development. PLoS One. 2012;7 11:e47557.
936 65. Boden SA, Kavanova M, Finnegan EJ and Wigge PA. Thermal stress effects on
937 grain yield in Brachypodium distachyon occur via H2A.Z-nucleosomes. Genome
938 Biol. 2013;14 6:R65.
939 66. Kennedy M, Fishbaugher ME, Vaughan AM, Patrapuvich R, Boonhok R,
940 Yimamnuaychok N, et al. A rapid and scalable density gradient purification
941 method for *Plasmodium* sporozoites. Malar J. 2012;11:421.
942 67. Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina
943 sequence data. Bioinformatics (Oxford, England). 2014;30 15:2114-20.
944 68. Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat
945 Methods. 2012;9 4:357-9.
946 69. Li B and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data
947 with or without a reference genome. BMC bioinformatics. 2011;12 1:323.
948 70. Okonechnikov K, Conesa A and Garcia-Alcalde F. Qualimap 2: advanced multi-
949 sample quality control for high-throughput sequencing data. Bioinformatics
950 (Oxford, England). 2016;32 2:292-4.
951 71. Nikolayeva O and Robinson MD. edgeR for differential RNA-seq and ChIP-seq
952 analysis: an application to stem cell biology. Methods Mol Biol. 2014;1150:45-
953 79.
954 72. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
955 differential expression analyses for RNA-sequencing and microarray studies.
956 Nucleic Acids Res. 2015;43 7:e47-e.
957 73. Wickham H. ggplot: An implementation of the Grammar of Graphics in R, 2006. R
958 package version 04 0.
959 74. Wickham H and Chang W. ggplot2: An implementation of the Grammar of
960 Graphics. R package version 07, URL: http://CRAN R-project org/package=
961 ggplot2. 2008.
962 75. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots:
963 Various R programming tools for plotting data. R package version. 2009;2 4:1.
964 76. Law CW, Alhamdoosh M, Su S, Smyth GK and Ritchie ME. RNA-seq analysis is
965 easy as 1-2-3 with limma, Glimma and edgeR. F1000Research. 2016;5.
966 77. Martin M. Cutadapt removes adapter sequences from high-throughput
967 sequencing reads. EMBnet journal. 2011;17 1:pp. 10-2.
968 78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
969 Alignment/Map format and SAMtools. Bioinformatics (Oxford, England).
970 2009;25 16:2078-9.

18

971 79. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-
972   based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9 9:R137.
973 80. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing
974   genomic features. Bioinformatics (Oxford, England). 2010;26 6:841-2.
975 81. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data.
976   Bioinformatics (Oxford, England). 2011;27 12:1653-9.
977 82. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene
978   and transcript expression analysis of RNA-seq experiments with TopHat and
979   Cufflinks. Nat Protoc. 2012;7 3:562-78.

980 **Figure Legends:**

981 **Fig. 1** Transcriptional activity of the *P. vivax* sporozoite and evidence for translational
982 repression. **a** Relative transcript abundance of key marker genes for sporozoites inferred by
983 RNA-seq and qPCR (here) relative to previously published microarray data [24]; **b** Relative
984 proportion of genes detectable as transcripts and proteins or transcripts only in RNA-seq and
985 previously published proteomic data. Dashed line shows cut-off used in the current study for
986 putatively repressed transcripts. Immunofluorescent staining of select proteins either known
987 (UIS4) or predicted here (LISP1, EXP1 and ACP) to be translationally repressed in
988 sporozoites in **c** sporozoite stages. CSP, mTIP as known positive controls and TRAP and BIP
989 as exerpimental positive controls and **d** liver stages (schizonts) at 7 days post-infection in
990 HuHep mice. Liver expression of EXP1 and ACP has been demonstrated by IFA in
991 Mikolajczak et al [13], using the same antibodies as used here.

992

993 **Fig. 2** Differential transcription between *Plasmodium vivax* salivary-gland sporozoites and
994 blood-stages. **A** BCV plot showing separation between blood-stage (black) and salivary-gland
995 sporozoite (red) biological replicates. **B** Volcano plot of distribution of fold-changes (FC) in
996 transcription between blood-stages and salivary-gland sporozoites relative to statistical
997 significance threshold (False Discovery Rate (FDR) ≤ 0.05). Positive FC represents up-
998 regulated transcription in the sporozoite stage. **C** Mirror plot showing pFam domains
999 statistically significantly (FDR ≤ 0.05) over-represented in salivary-gland sporozoite up-
1000 regulated (red) or blood-stage up-regulated (black) transcripts. Scale bar truncated for
1001 presentation. * - 55 PRESAN domains are in this dataset. ** - 99 Vir domains are in this
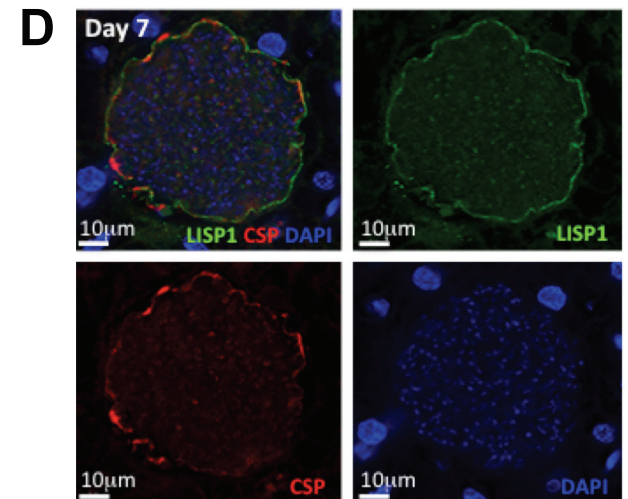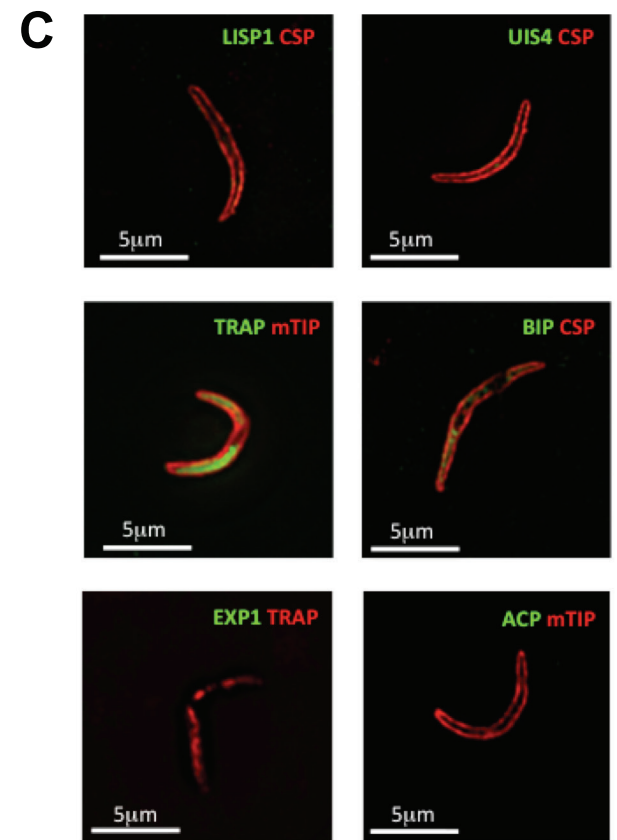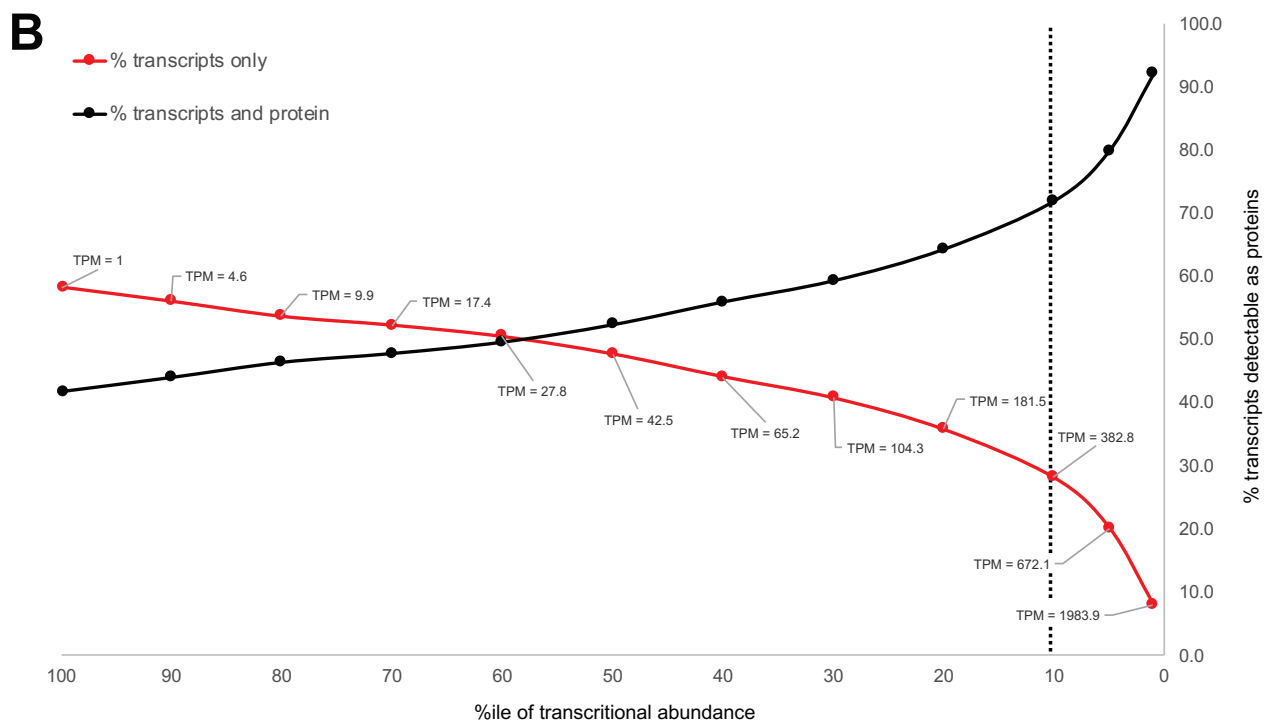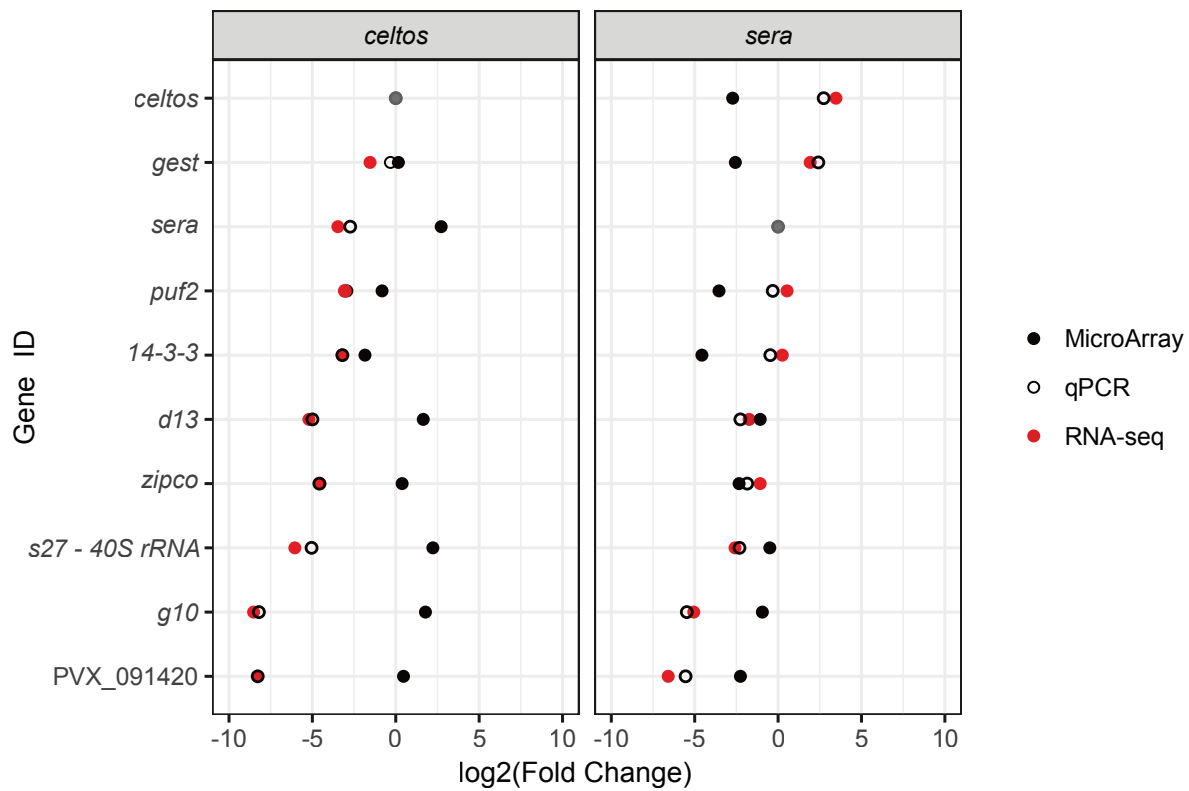1002 dataset.

1003

1004 **Fig. 3** Differential transcription between *P. vivax* sporozoites (SPZ), mixed (mLS) and
1005 hypnozoite (HPZ) enriched liver stages (liver-stage data from Gural et al [22]). **A** Heatmap
1006 comparisons showing summed transcription of enriched Pfam domains in HPZ vs mLS (left),
1007 SPZ vs HPZ (top middle) and SPZ vs mLS (top right) comparisons. All Pfam domains
1008 statistically significantly enriched at p-value 0.05). All transcript data for stage up-regulated
1009 genes at FDR 0.01). **B** Violin box-plot showing relative fold-change differences between SPZ
1010 and HPZ compared with SPZ and mLS for genes down-regulated in mLS compared to SPZ,
1011 but not down-regulated in HPZ compared to SPZ. **C** Ternary heatmap summarizing relative
1012 transcript abundance in each of SPZ, mLS and HPZ stages.
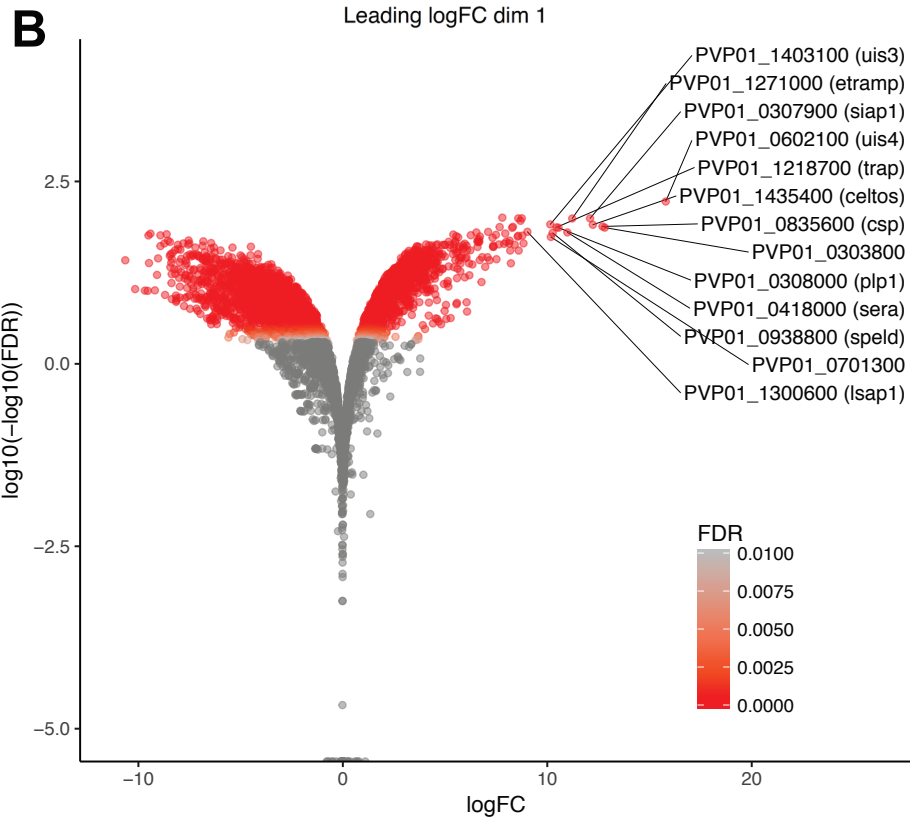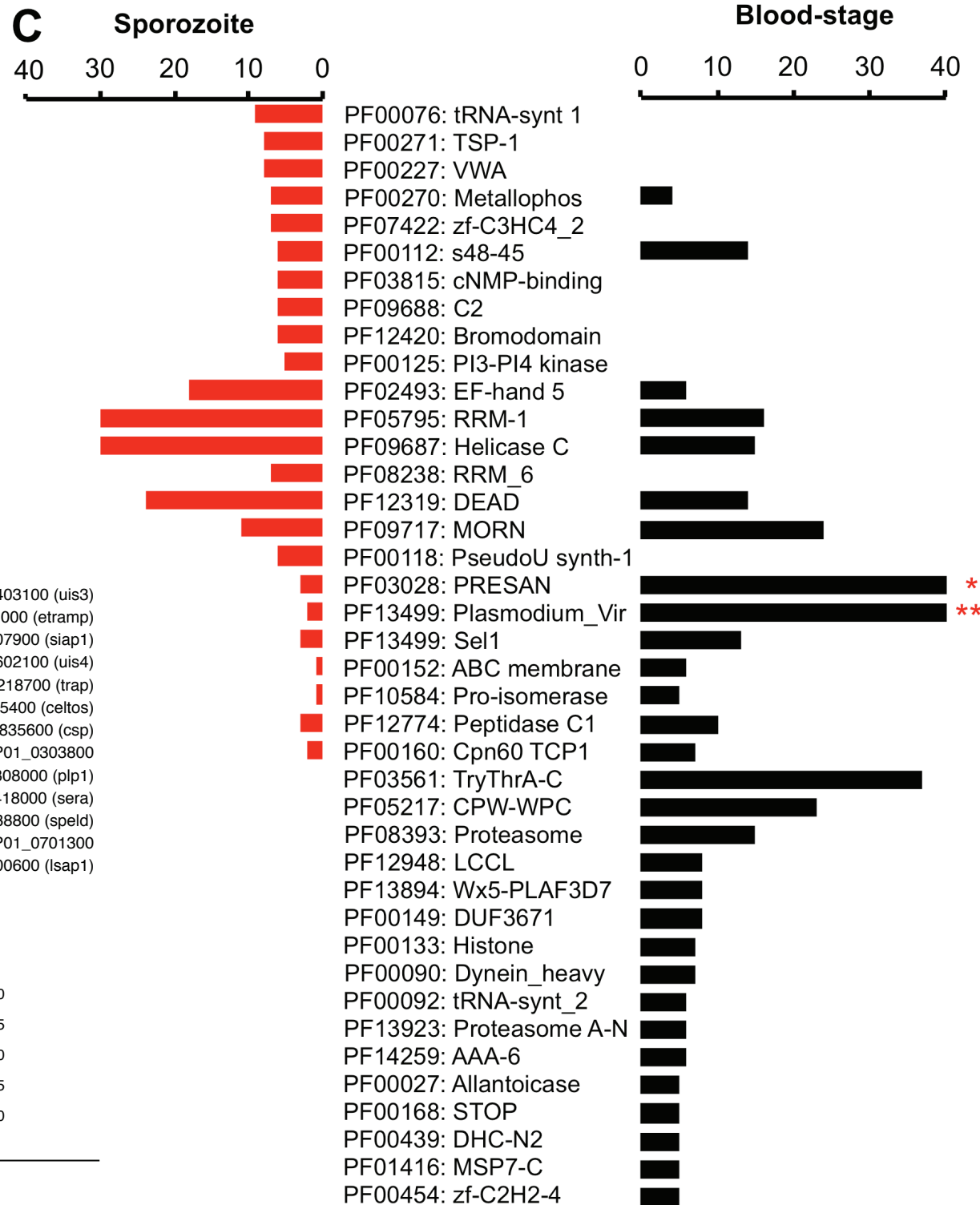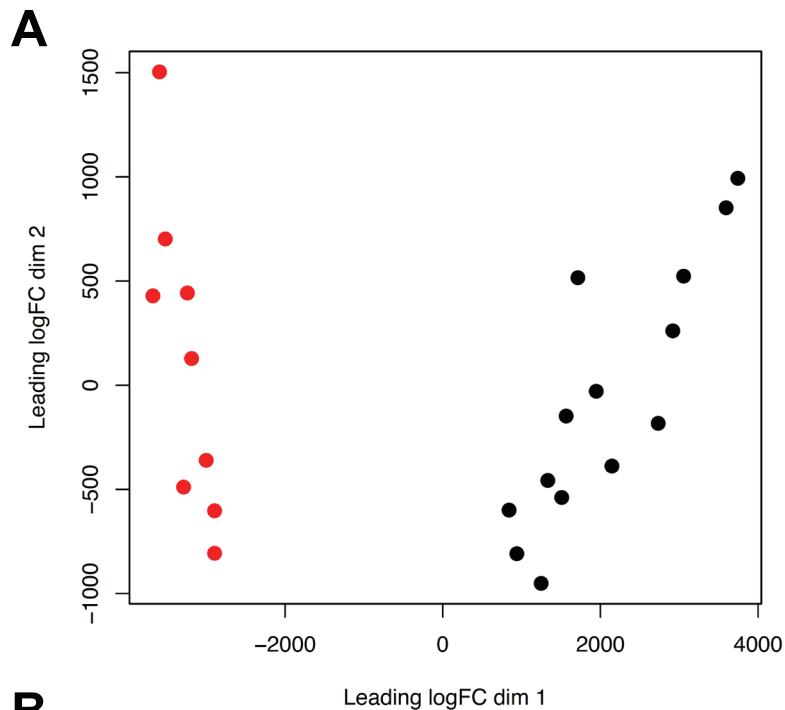
1013

1014 **Fig. 4** Histone epigenetics relative to transcriptional behaviour in salivary-gland sporozoites.
1015 **a** Representative H3K9me3, H3K4me3 and H3K9ac ChIP-seq data (grey) from a
1016 representative chromosome (*P. vivax* P01 Chr5) relative to mRNA transcription in salivary-
1017 gland sporozoites (black) and blood-stages (black). Small numbers to top left of each row
1018 show data range. **b** Salivary-gland sporozoite transcription relative to nearest stable histone
1019 epigenetic marks. Numbers at the top of the figure represent total genes included in each
1020 category. Numbers within in box plot represent mean transcription in transcripts per million
1021 (TPM). **c** Sequence motifs enriched within 1kb upstream of the Transcription Start Site of
1022 highly transcribed (top 10%) relative to lowly transcribed genes associated with H3K9ac
1023 marks in salivary-gland sporozoites. **d** Relative transcription of (sub)telomeric genes in *P.*
1024 *vivax* salivary-gland sporozoites and blood-stages categorized by gene sets up-regulated in
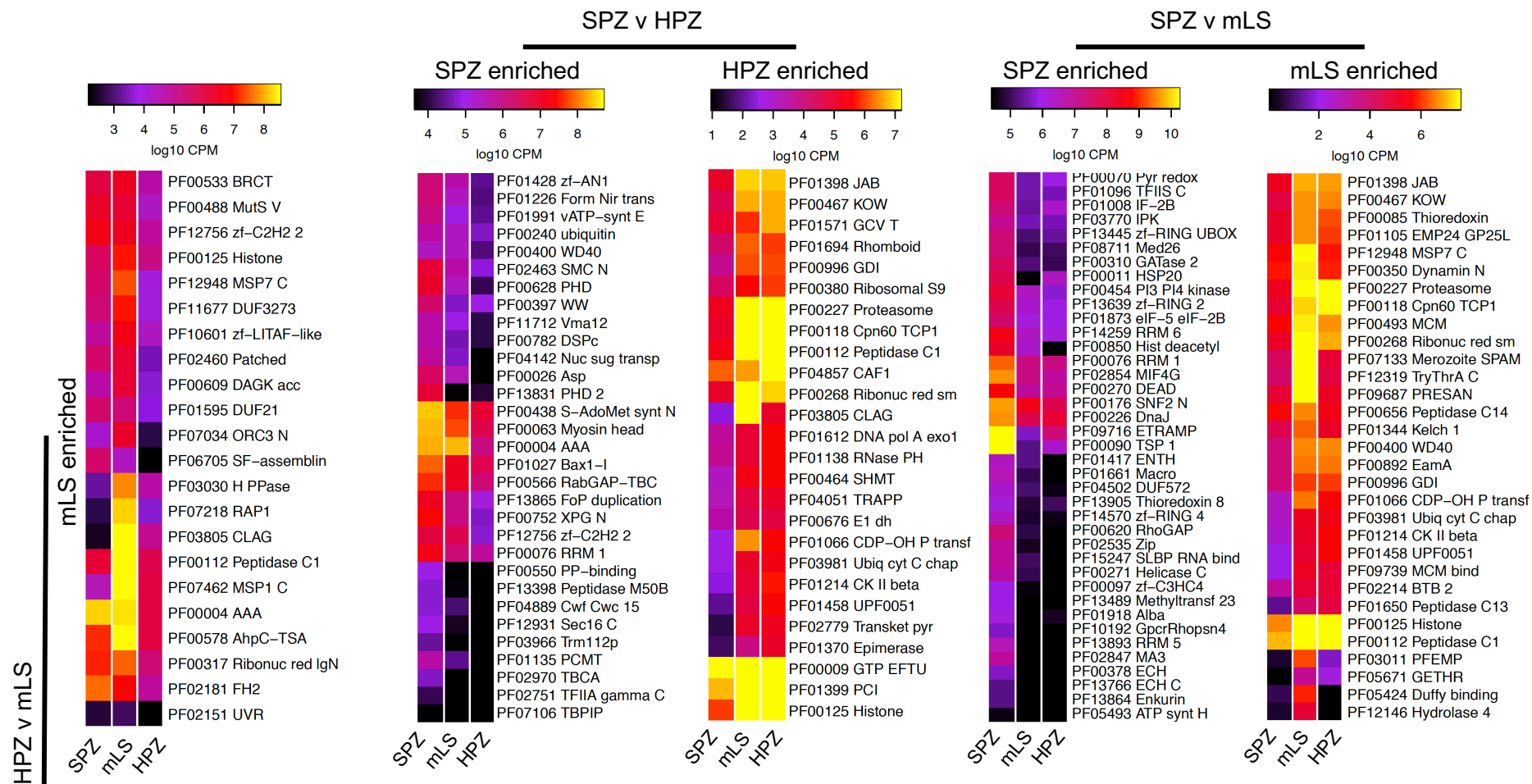
1025   blood-stages (blue), salivary sporozoites (red) or not stage enriched (grey). Numbers in each
1026   box show mean transcription in TPM.
1027
1028   **Fig. 5** Schematic of potential mechanisms underpinning development in differentiation of *P.*
1029   *vivax* sporozoites during liver-stage infection as hypnozoites and schizonts. We suggest
1030   differentiation programming at different points in development; first, schizont or hypnozoite
1031   fate possibly encoded in the sporozoite as epigenetic signals or translationally repressed
1032   transcripts; secondly, suppression signals that halt progression of the hypnozoite to schizont
1033   stage and support persistence; and finally activation signals signified by a release in
1034   chromatin, (post)transcriptional and (post)translational control leading to up-regulation of
1035   replication, metabolic and protein export pathways.

A

B

% transcripts only
% transcripts and protein

TPM = 1
TPM = 4.6
TPM = 9.9
TPM = 17.4
TPM = 27.8
TPM = 42.5
TPM = 65.2
TPM = 104.3
TPM = 181.5
TPM = 382.8
TPM = 672.1
TPM = 1983.9

%ile of transcritional abundance

% transcripts detectable as proteins

C

LISP1 CSP
UIS4 CSP
TRAP mTIP
BIP CSP
EXP1 TRAP
ACP mTIP
5μm

D

Day 7
LISP1 CSP DAPI
LISP1
CSP
DAPI
10μm

A — Leading logFC dim 1 (x-axis) vs Leading logFC dim 2 (y-axis)

B — logFC (x-axis) vs log10(−log10(FDR)) (y-axis)

PVP01_1403100 (uis3)
PVP01_1271000 (etramp)
PVP01_0307900 (siap1)
PVP01_0602100 (uis4)
PVP01_1218700 (trap)
PVP01_1435400 (celtos)
PVP01_0835600 (csp)
PVP01_0303800
PVP01_0308000 (plp1)
PVP01_0418000 (sera)
PVP01_0938800 (speld)
PVP01_0701300
PVP01_1300600 (lsap1)

FDR
0.0100
0.0075
0.0050
0.0025
0.0000

C

Sporozoite / Blood-stage

PF00076: tRNA-synt 1
PF00271: TSP-1
PF00227: VWA
PF00270: Metallophos
PF07422: zf-C3HC4_2
PF00112: s48-45
PF03815: cNMP-binding
PF09688: C2
PF12420: Bromodomain
PF00125: PI3-PI4 kinase
PF02493: EF-hand 5
PF05795: RRM-1
PF09687: Helicase C
PF08238: RRM_6
PF12319: DEAD
PF09717: MORN
PF00118: PseudoU synth-1
PF03028: PRESAN *
PF13499: Plasmodium_Vir **
PF13499: Sel1
PF00152: ABC membrane
PF10584: Pro-isomerase
PF12774: Peptidase C1
PF00160: Cpn60 TCP1
PF03561: TryThrA-C
PF05217: CPW-WPC
PF08393: Proteasome
PF12948: LCCL
PF13894: Wx5-PLAF3D7
PF00149: DUF3671
PF00133: Histone
PF00090: Dynein_heavy
PF00092: tRNA-synt_2
PF13923: Proteasome A-N
PF14259: AAA-6
PF00027: Allantoicase
PF00168: STOP
PF00439: DHC-N2
PF01416: MSP7-C
PF00454: zf-C2H2-4

A — *P. vivax* P01 Chr5: 1.52 Mb

100 kb

H3K9me^3 (60)
H3K4me^3 (150)
H3K9ac (70)
SPZ RNA-seq (500)
BS RNA-seq (500)

B

5659   231   3676   1284   1085

$\log_{10}$ TPM

32.1   0.7   50.3   111.9   111.7

All
H3K9me3
H3K4me3
H3K9ac
H3K9ac.and.K4me3

C

TGTACaA
e-value: 2.7x10^{-2}

ATATTT
e-value: 3.3x10^{-3}

D

Sporozoite   Blood-stage

$\log_2$ TPM

0.5   1.0   1.5   24.3   1.8   0.5

Stage.enrichment
BS–up
No change
SPZ–up

DIFFERENTIATION
- **Translational repressors**
  e.g., Puf2, Alba2/4, HOMU, Yippee, Zipco
- **Invasion**
  e.g., Celtos, Gest
- **Liver development**
  e.g., LISP1, LISP2, ETRAMPS, TRAP
- **Transcription factors**
  e.g., AP2-SP2, AP2-L, AP2-Q?

SUPPRESSION
- **Metabolic/replication suppressors**
  e.g., sucB, HAP2, MAK16
- **Histone arginine methylation**
  e.g., CARM1, EEML2
- **Protein translation**
  e.g., eIF-3H, Puf1

ACTIVATION
- **Merozoite development**
  e.g., MSP1, MSP3, MSP9
- **Rhoptry function**
  e.g., RAP1, RNP2, RNP3
- **Reticuocytye binding**
  e.g., RBP2a, RBP2b, RBP2C
- **Exported Proteins**
  e.g., PHISTs