Single-cell transcriptomics to define *Plasmodium falciparum* stage-transition in the mosquito midgut

Mubasher Mohammed[1], Alexis Dziedziech[1*], Vaishnovi Sekar[2*], Medard Ernest[3], Thiago Luiz Alves E Silva[3], Balu Balan[4], S. Noushin Emami[1], Inna Biryukova[2], Marc R. Friedländer[2] Aaron Jex[4], Marcelo Jacobs Lorena[5], Johan Henriksson[6], Joel Vega Rodriguez[3#], Johan Ankarklev[1,7#]

[1] Department of Molecular Biosciences the Wenner Gren Institute, Stockholm University, Svante Arrhenius väg 20C, SE-106 91, Stockholm, Sweden
[2] Science for Life Laboratory, Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, Svante Arrhenius väg 20C, SE-106 91, Stockholm, Sweden
[3] Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12735 Twinbrook Parkway, Rockville, MD 20852, USA
[4] Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, VIC, Australia; Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia
[5] Department of Molecular Microbiology and immunology, Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe St, Baltimore, MD 21205, USA
[6] Laboratory for Molecular Infection Medicine Sweden (MIMS), Department of Molecular Biology, Umeå University, SE-901 87, Umeå, Sweden
[7] Microbial Single Cell Genomics, Department of Cell and Molecular Biology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Running Head: scRNA-seq of *P. falciparum* development in the mosquito midgut

* These authors contributed equally

Address correspondence to Johan Ankarklev, *johan.ankarklev@su.se* or Joel Vega Rodriguez, joel.vega-rodriguez@nih.gov

Abstract:

Malaria inflicts the highest rate of morbidity and mortality among the vector-borne diseases. The dramatic bottleneck of parasite numbers that occurs in the gut of the obligatory vector mosquito provides a promising target for novel control strategies. Using single-cell transcriptomics, we analyzed *Plasmodium falciparum* development in the mosquito gut, from unfertilized female gametes through the first 20 hours post blood feeding, including the zygote and ookinete stages. This study revealed the transcriptional trajectories of the ApiAP2 family of transcription factors, and of parasite stress genes in response to the harsh environment of the mosquito midgut. Further, employing structure-based functional predictions we found several upregulated genes predicted to encode intrinsically disordered proteins (IDPs), a category of proteins known for their importance in regulation of transcription, translation and protein-protein interactions. IDPs are known for their antigenic properties and may serve as suitable targets for antibody or peptide-based transmission suppression strategies.

Background:

Malaria remains a large global health burden, infecting approximately 224 million people each year and having a death toll exceeding 627,000, mainly children in Sub-Saharan Africa (*World Malaria Report 2021*, n.d.). Out of the five *Plasmodium* species known to infect humans, *P. falciparum* causes the vast majority of severe cases and deaths. *Plasmodium* parasites have a complex life cycle that extends between the human host and the female *Anopheles* mosquito. Parasite development within the mosquito begins during the ingestion of a blood meal containing male and female gametocytes. Gametocytes quickly differentiate into gametes in the mosquito midgut, which fuse to form a diploid zygote. The zygote differentiates into a motile banana-shaped ookinete that migrates within the blood bolus, traverses the midgut epithelium, and forms an oocyst on the basal side of the midgut. During the transition from the blood bolus to the hemocoel, the parasite is exposed to harsh conditions within the midgut, including immune factors from the human host, the midgut microbiome and oxidative stress from blood digestion (Billker et al., 1997; Smith et al., 2014). Subsequently, the ookinetes lodge into the basal lamina, where they develop into replicative oocysts and thousands of sporozoites are produced. Sporozoites are released from mature oocysts into the hemocoel, from where they invade the salivary glands and are delivered into a vertebrate host when the mosquito takes a blood meal (Graumans et al., 2020).

While *P. falciparum* undergoes large scale asexual replication in the human host, sexual recombination occurs only in the mosquito vector. This enables genetic crossing and spread of genetic factors such as drug tolerance, into the parasite

progeny (Vaughan et al., 2015). The human-to-mosquito transmission phase is one of the major bottlenecks in the parasite's lifecycle, in part due to the limited number of gametocytes taken up by the mosquito, but also due to immune factors from the human blood, the mosquito microbiome and the effective innate immune response elicited by the mosquito (Smith et al., 2014). Moreover, the anopheline mosquito has evolved to express a battery of antimicrobial defenses including long non-coding RNAs, nitric oxide, prophenoloxidases, and anti-microbial peptides (Dimopoulos et al., 2001; Padrón et al., 2014). These events contribute to the large bottleneck in parasite numbers during early development in the mosquito midgut, making these stages a crucial target for transmission-blocking interventions (Griffin et al., 2010). In addition, recent advances in malaria control have bolstered the interest in malaria intervention strategies linked to the reduction of transmission, including vaccines that target sexual, sporogenic, and/or mosquito stage antigens in order to interrupt malaria transmission (SSM-VIMT) (Sauerwein & Bousema, 2015). Although *P. falciparum* development in the mosquito midgut is considered a prime target for the development of effective transmission-blocking interventions, little is currently known about the transcriptional program that controlling these processes, with surprisingly few proteins characterized and annotated. In recent years, single-cell RNA sequencing (scRNA-seq) has been used extensively to resolve cellular heterogeneity, discern key transcriptionally regulated biological processes and predict transcript and protein expression patterns and functions in multiple stages during the life cycle of *P. falciparum* (Howick et al., 2019; Ngara et al., 2018; Poran et al., 2017; Real et al., 2021; Reid et al., 2018). Nonetheless, a detailed and comprehensive transcriptional map of key *P. falciparum* developmental stages in the mosquito midgut is still missing (Figure 1A).

In this study, we use scRNA-seq to explore the *P. falciparum* developmental dynamics from unfertilized female gametes, through the zygote and the ookinete stages. All parasite cells were carefully isolated by micromanipulation from infected *Anopheles gambiae* mosquitoes. We sought to elucidate the timing of expression of genes connected to cytoskeletal modifications, invasion, and meiosis. Computational analyses identified the timing of expression for members of the ApiAP2 family of transcription factors throughout the developmental timeline. Further, expression of a set of genes containing the upstream binding motif of the stage-specific ApiAP2-O were connected to the expression of this transcription factor. Moreover, we provide insights into the connections of the parasite's stress responses during the mosquito midgut stage, with morphogenic changes and evolution of cell fate. Finally, we identify several highly expressed genes with non-annotated function, that are predicted to be intrinsically disordered proteins (IDPs) and novel immunogenic candidates for antibody-based therapies. The elucidation of genetic factors involved in *P. falciparum*'s differentiation in the mosquito midgut provides vital insights into the biology of these vulnerable developmental stages while identifying new targets for the control of malaria transmission.

Results:

Unsupervised clustering defines a developmental trajectory and transcriptional heterogeneity among the early developmental timepoints

To enable high-resolution transcriptional profiling of *P. falciparum* as it develops in the lumen of the mosquito midgut, we infected the African main malaria vector *An. gambiae* with mature stage V gametocytes of the NF54 genetic background. Mosquito midguts were dissected at 2, 4, 8, 12, and 20 h post infection (p.i.), homogenized to

release the blood bolus content, followed by Hoechst (DNA) and anti-Pf25 antibody staining to label the parasite sexual stages (female gametes, zygotes and ookinetes). Parasites do not develop synchronously in the midgut of the mosquito and therefore, we sought to visualize parasites prior to isolation through DNA and Pfs25 staining, and select only parasites with intact morphology. Individual parasites were collected by micro-capillary manipulation within 30 minutes post dissection and homogenization (Figure 1B), and cDNA libraries were prepared for each isolated parasite using a modified version of the Smart-seq2 protocol (Ngara et al., 2018; Reid et al., 2018) using 24 cycles for cDNA amplification.

A total of 180 single cells were collected, of which 125 single cells were retained for subsequent analyses after filtering out cells with poor read counts. The remaining cells were quality controlled based on mRNA count and number of genes expressed (Supplementary Figure 1A). The data from each time point were validated by analyzing the expression levels of known marker genes for specific midgut developmental stages, including G27/25, p230, CDPK4, P25 and P28 (Bennink et al., 2016). Marker genes associated with ookinete development or midgut invasion like CTRP, WARP and SOAP, showed significantly elevated expression levels towards the end of the time course (Figure 1C). Further, we integrated and normalized our data with a recent single-cell data set published by Real and colleagues (Real et al., 2021) in order to validate the coverage of the hitherto unexplored *Plasmodium* developmental stages. As seen in Figure 1A, our late, 20h timepoint aligns with the majority of cells from the Real *et al.* dataset, which were collected at 24 h post mosquito infection, while our earlier timepoints form separate clusters (Figure 1A and Supplementary Figure 1B).

A total of 2,000 genes with significant variance were selected for downstream analyses, while 2,835 did not show significant variance (Supplementary Figure 1C).

Further, among the genes with the highest average change, we found several non-annotated genes and genes that were previously shown to have essential roles during parasite development in the midgut, such as GAMER, PIMMS1, PIMMS57 and SOPT (Supplementary Figure 1D) (Akinosoglou et al., 2015; Ukegbu et al., 2020). Next, we clustered the cells in an unsupervised manner using UMAP non-linear dimensionality reduction, which identified five distinct clusters that largely correspond to the time points at which parasites were collected (Figure 1D and E, see Methods for details). The UMAP displays transcriptional variability between and within isolated time points. Early timepoints, in particular, were computationally predicted to diverge by cell type within each time point (Supplementary Figure 2A).

Pseudotime alignment indicates heterogeneity among the early timepoints

Given the heterogeneity of the identified cells, we sought to determine a biologically relevant temporality to the cell clusters. Since the non-fertilized female gametes (NFFGs) were cultivated *in vitro* and therefore not exposed to stresses from the mosquito midgut, we performed a cell trajectory pseudotime analysis that included non-fertilized female gametes. The results indicated that the inclusion of these cells created a second trajectory, which appeared to distort the natural developmental trajectory as visualized in (Supplementary Figure 2B). Therefore, we excluded the 0-hour (*in vitro*) time point from subsequent pseudotime analyses. To infer a pseudotime alignment for the five clusters identified by Seurat analyses, we reconstructed the cell trajectory in terms of their relative developmental stage. Clusters were reordered with C3 indicated to be the earliest developmental stage, followed by C1, C0, C2 and ending with C4 as the terminal cluster (Figure 2A). Further, when the data was ordered along a temporal plane, i.e., the pseudotime (pt) axis, three distinct clusters, pt0, pt1 and pt2

(Figure 2A, right) were delineated. The pt lineage showed general similarity with the sampling timepoints, although the early timepoints showed a trend of asynchronicity where some of the cells aligned with a more progressed stage of development while others aligned with a differing transcriptional state, which we hypothesize may indicate stalled development (Figure 2B, left). Further, the Seurat clusters showed greater spread over the pseudotime within a single group and more overlap between groups in comparison to the pt clusters, which were more distinct (Figure 2B, middle *vs*. right). While the pt clusters showed more mutual exclusivity than Seurat clustering, the Seurat clustering indicated more distinct intermediary points in the transition from the early zygote to the ookinete. We utilized the Seurat cluster variance when exploring the role of key genes in transcriptional programing while we relied on slingshot pseudotime clusters for determining distinct biological processes. The GO enrichment for the pt clusters revealed similar GO terms to the Seurat clustering, pointing to an ordered and predictable transcriptional regulation of gene expression programs. For example, pt0 was related to “response to stimulus", “localization” and “reproduction”, pt1 was related to “DNA replication”, “signaling”, and “translation”, while pt2 was related to “entry into host”, “invasion”, and “response to abiotic stimulus” (Figure 2C). Based on the imaging data acquired during the isolation of individual parasites, we saw distinct cell morphologies throughout the time of collection (Figure 2D). When comparing the nucleus mean volume, fluorescence intensity and mRNA abundance, we saw variation across the 5 Seurat clusters with C1 containing higher nuclear mean volume and fluorescence intensity, which coincides with the post meiotic tetraploidy (4N) (Supplementary Figure 3 A, B and C). To identify the genetic signature for each cluster, we generated the scaled average expression of the top10 differentially expressed genes for the Seurat clusters (C0-C4) and the top 20 differentially expressed genes for

the pt clusters (pt0-2) (Supplementary Figure 4 A and B, respectively). Thus, based on morphology, nuclear DNA and GO term analyses, clusters C3 and C1 align with pt0 and represent early zygote development, C0 aligns with pt1 and represents the maturing zygote to early ookinete development, whereas C2 and C4 align with pt2 and represent mid-to-late ookinete development.

Pseudotime ordering identifies gene programs that drive *P. falciparum* development in the midgut

To gain insight into the transcriptional programing for the clusters identified over the pseudotime, we performed hierarchical clustering and gene ontology analysis of the top 200 differentially expressed genes across all timepoints (Figure 3A, Supplementary Table 1). We observed a sequence of biological events corresponding to the development from early zygote to ookinetes over the pseudotime, as supported by the top scoring GO terms which included “DNA Replication” in pt0, “Reproduction” in pt1 and “Entry into host” and “Negative regulation of metabolic process” in pt2. After performing a gene set enrichment analysis over the pseudotime, increasingly fewer cells showed enrichment for metabolic processes. The average expression of each significantly upregulated program was assessed throughout the developmental time course (Supplementary Figure 5 A and Supplementary Table 2). To further define patterns of biological processes among the gene modules that make up the pseudotime clusters, we also ran over-representation analyses of the pt clusters’ (pt0 to pt2) gene ontology terms (Supplementary Figure 5 B, C and D).

Given that fertilization and meiosis occur in the midgut under extreme oxidative stress conditions, we selected known meiosis and DNA repair genes to further our

understanding of these processes during the midgut developmental stages (Figure 3B). A distinct set of early expressed genes, including APE1, FEN 1, MSH2-1, MLH, PF3D7_1140300 and PF3D7_1437200, were included in DNA repair and DNA replication. A subsequent set of genes that correspond to developmental progression through meiosis in the early zygote, included the expression of MISFIT (male-inherited sporulation factor important for transmission), FEN1, a DNA repair enzyme, and DMC1, a meiotic recombination protein, and MLH, a DNA mismatch repair protein. MISFIT may also contribute to cell cycle progression and replication dynamics in conjunction with NEK4 and ORC1 leading to the development of the early ookinete (Bushell et al., 2009). A third set of genes expressed at an even more advanced stage of development include: RAD23 involved in DNA repair, GyrA, DNA topoisomerase activity, as well as MSH2-2 involved in DNA recombination, Capd3, meiotic chromosome condensation, TERT, telomere maintenance and SMC1 involved in sister chromatid segregation. In summary, the high-resolution single-cell dataset enables the description of expression timing of key genes involved in meiosis and its coordination. Interestingly, we observe a difference between ORC1 and ORC5 expression, which may indicate how the timing of the different subunits of the origin recognition complex in *P. falciparum* are orchestrated in their expression.

Towards the end of the pseudotime, there are seven genes, including RAD2, RAD5 and PF3D7_1250800, involved in DNA repair that may be expressed in response to oxidative stress caused by heme released from hemoglobin digestion, or alternatively expressed in late ookinetes in preparation for the subsequent mass replication during sporogony. The additional four genes RFC2, involved in DNA replication, SMC5, structural maintenance of chromosomes, GEX1, karyogamy and PlasMei2, linked to schizogony, are likely to function during sporogony in oocysts.

Considering the extensive morphological changes that occur as *P. falciparum* develops from early zygotes to mature ookinetes, we sought to better understand the coordination of its cytoskeletal gene network (Figure 3C). We found that ADF2, MyoK, and DCX, involved in actin depolymerization, cytoskeletal regulation and protein polymerization, respectively, were highly expressed in pt0, as were Pi3k and Rab1a, indicating a potential role for endosome recycling in the early development of zygotes. Pt1 showed upregulation of VPS51 and DHHC8, potentially increasing vesicle sorting and protein transport, followed by the upregulation of DHHC3, which is involved in protein palmitoylation and FIKK7.1, involved in protein phosphorylation and signal transduction. Thus, pt1 likely portrays the coordination of vesicles and protein transport that aid in parasite survival in the gut and development from early to mid/late ookinetes. Finally, cells in pt2 significantly expressed other regulators of actin (ADF2) and cell division (MyoJ) as well as ARP-dependent DNA helicase Q1 (RECQ1), which plays a role in DNA double-strand break repair and DNA unwinding (Claessens et al., 2018), possibly in preparation for the subsequent sporogony. Further, late ookinete proteins like Cap380 (oocyst capsule protein) essential for oocyst formation (Srinivasan et al., 2008) and PSOP7 (putative secreted ookinete protein) linked to midgut epithelial invasion (Ecker et al., 2008), were both highly expressed in preparation of epithelia invasion and subsequent oocyst development (Figure 3C). Thus, our data shines new light on the complex transcriptional orchestration of cytoskeletal changes as zygotes differentiate into ookinetes, and as ookinetes prepare for oocyst formation.

We also analyzed the timing of expression of ookinete invasion genes, which are almost exclusively expressed within pt2. These include well-known ookinete marker genes such as; Gamer, SOPT, WARP and PIMMS43 (Figure 3D). Interestingly, Clag9,

a cyto-adherent molecule, known to form complex with RHOPH1 during erythrocyte invasion and nutrient uptake (Schureck et al., 2021), was found to be upregulated in late ookinetes.

Defining the temporality of expression of ApiAP2 family members

Members of the ApiAP2 transcription factor family, which in *P. falciparum* consists of 27 genes, regulate differentiation and stage progression events (Jeninga et al., 2019). We observed significant upregulation of expression in four out of the five AP2-O genes expressed during parasite development in the mosquito midgut, AP2-O4 did not show expression at any time point (Figure 4A). AP2-O was found to be expressed in the C0 and C1 clusters, indicating high expression in the early zygote and in agreement with a previous report showing AP2-O transcription in stage V female gametocytes (Kaneko et al., 2015). AP2-O2, AP2-O3 and AP2-O5 show expression in C0 and C2 (Figure 4A), during the zygote to ookinete transition. Thus, the timing of expression of the AP2-O genes supports a role in the regulation of genes linked to ookinete function, as previously shown by functional analyses (Li et al., 2021; Modrzynska et al., 2017, p. 2; Zhang et al., 2017). Interestingly, we observe additional AP2 genes that appear to be upregulated in a significant proportion (>50%) of cells within their respective clusters (Figure 4A). These include two additional AP2 genes in C2 (PF3D7_1239200 and PF3D7_1342900) and four AP2 genes in C4 (PF3D7_0420300, PF3D7_0934400, PF3D7_0802100 and PF3D7_1139300), none of which has previously been linked to any developmental stage. In agreement with our data, PF3D7_0934400 was shown to be highly expressed among ookinetes in the single-cell data by Real and colleagues. Also, PF3D7_0622900: AP2Tel, which is significantly upregulated in C4, appears to

show low to intermediate expression among ookinetes in the same dataset (Real et al., 2021).

To gain further insight into gene regulation during *P. falciparum* zygote to ookinete development, we made use of previously published ChIP-seq data on *P. berghei* parasites to identify genes regulated by AP2-O (Kaneko et al., 2015). The target genes from this study were used to find orthologs in the *P. falciparum* genome that showed significant upregulation in our scRNA-seq data. To validate the orthologs, we investigated the presence of the AP2-O binding motif [TC][AG]GC[TC][AG] (Jeninga et al., 2019; Kaneko et al., 2015) within a 1 kb region upstream of the start codon of differentially expressed genes (Figure 4B). We found that a small number of genes putatively regulated by AP2-O, including *p12*, PF3D7_1349900, PF3D7_0518000 and *cpr*, were upregulated at the beginning of the pseudotime followed by downregulation at an earlier or similar timepoint as the expression of AP2-O (Figure 4C). The *p12* gene belongs to the 6-cys protein family expressed in blood-stage parasites, and was detected in the proteome of merozoites (Sanders et al., 2005). The *cpr* gene has been predicted to encode NADPH-cytochrome P450 reductase, with a predicted role in oxidation-reduction (GO:0055114, AmiGO, geneontology.org). PF3D7_1349900 and PF3D7_0518000, are both unknown proteins but have previously been shown to be transcribed in mature gametocytes and to some level in ookinetes (López-Barragán et al., 2011).

To compare the genes expressed in our scRNA-seq dataset and containing the AP2-O binding motif with the orthologs in the study by Kaneko and colleagues (Kaneko et al., 2015), we only included genes expressed by at least 25% of the cells in each pt cluster. Due to the nature of dropouts in single-cell sequencing data and the conservative threshold set in our analysis, our list of genes is not as extensive as the

one by Kaneko et al. (Kaneko et al., 2015). However, our analysis shows a significant overlap with the target genes validated in Kaneko *et al*. We detect significant expression of genes involved in midgut epithelium invasion and oocyst formation, including the secreted proteins; CelTOS, CHT1, CTRP, GAMER, PSOP1, SOAP and WARP, the plasma membrane associated protein; P25, the oocyst development protein Cap380, and microneme proteins; PLP3, PLP4. Furthermore, the pellicular protein TTL, histone proteins H2B, H3 and H3.3, DNA replication proteins MCM9 and CDPK3, which are essential for midgut infection, are all expressed in our dataset. Finally, PIMMS43, previously described as having a function in parasite immune evasion and sporogonic development (Ukegbu et al., 2020) and MUS81 involved in DNA-repair (Lee et al., 2014) were also expressed.

Response to stress during mosquito midgut development

The stress response elicited by *P. falciparum* is an essential mechanism for adaptation to host environmental factors across the lifecycle (Duran-Bedolla et al., 2017). We investigated the transcriptional activity of known cellular stress marker genes and found transcriptional differences between the Seurat clusters, C0-C4 (Figure 5A). The cells in the C0 cluster showed higher average expression of nPrx, which encodes a nuclear peroxiredoxin that has been suggested to protect nuclear components from oxidative damage or be involved in DNA repair during the asexual blood stage (Richard et al., 2011). Further, genes such as TRXR, PF3D7_0213500, 1-CysPxn and DHX36, are specifically expressed during the later developmental stages C4 (Figure 5A). We also compared the stress genes in non-fertilized female gametes (NFFGs) which showed a smaller population of cells expressing stress related genes compared to C0 (Figure 5A and B). The genes expressed in *in vitro* generated NFFGs, likely include

353 genes expressed through a preprogrammed default stress pathway in comparison to
354 the larger number of heterogeneous genes expressed in C0. The population of cells in
355 C0 likely demonstrate two trajectories for cells either fated to properly maturate into an
356 ookinete or parasites that arrested and failed to mature.
357
358 **Protein modelling of stage-specific non-annotated genes**
359 Due to the ability of membrane proteins to induce immune responses as well as their
360 importance in nutrient/chemical uptake (Kirk & Horner, 1995), we sought to define and
361 characterize non-annotated membrane proteins, which make up approximately 33%
362 of the *Plasmodium* genome (Böhme et al., 2019). Such proteins could serve as targets
363 in novel antibody-based therapies. First, from the pt clusters, we identified annotated
364 and non-annotated genes that were significantly upregulated (P adjusted value <0.05)
365 (Figure 6A). From pt1 and pt2, 41 and 111 non-annotated gene candidates made the
366 described cut-off, respectively. The tables in Figures 6B and 6C denote the presence
367 or absence of predicted TM domain and/or a signal peptide. Of note, all the selected
368 non-annotated genes in Figures 6B and 6C are classified as intrinsically disordered
369 proteins (IDPs), all of which had a minimum average log fold expression of 0.5
370 (Supplementary Table 5). These proteins are characterized by the presence of large
371 segments of disordered structure under normal physiological conditions. Distortion
372 exceeding 15.5% of the entire protein secondary structure were classified as IDPs
373 (Figure 6D, Supplementary Figure 6). IDPs have been described as highly
374 immunogenic (Guy et al., 2015) and may be candidates for novel antibody or peptide-
375 based therapeutics. Molecular simulation (Supplementary Figure 7) shows how protein
376 structure can be predicted for therapeutic molecules. This is an important tool for the
377 identification of immunogenic proteins to be used for vaccine development.

Discussion

The transcriptional regulation of *P. falciparum* development in the mosquito midgut remains understudied, with the availability of only a single-cell atlas of *P. falciparum* mosquito stages that includes mature ookinetes (Real et al., 2021). Using scRNA-seq, our data defines the genetics underlying the differentiation of a female gamete to an ookinete. We characterized five transcriptionally unique *P. falciparum* cell states as the parasite develops in the midgut lumen *An. gambiae*, its natural mosquito vetor. We also defined three distinct molecular signatures that orchestrate cell type transitions in the mosquito midgut. Genes related to DNA replication and metabolic processes were identified in early zygotes; reproduction, localization, and motility in intermediate stages, while early to mid-ookinetes express genes related to entry into its host and to down-regulation of metabolic processes. *Plasmodium* development in the mosquito involves a series of DNA replications and meiotic divisions. After fertilization, the zygote undergoes a round of DNA replication that increases ploidy to 4N, followed by a conventional two-step meiotic division (Janse et al., 1986). This is later followed by endomitotic division in the oocyst to produce the sporozoites. We observed the differential expression of genes involved in DNA replication and repair, as well as in meiosis, in three distinct expression programs; early, mid and late midgut development. The early program includes genes like NEK2, NEK4 and MISFIT that has been shown to regulate DNA replication in preparation for meiosis (Guttery et al., 2012). The mid developmental program includes genes like SMC1, CAPD3 and RAD23, involved in chromosomal segregation and DNA repair during meiosis (Hirano, 2016; Kalitsis et al., 2017; Madura & Prakash, 1990). The late program includes proteins expressed exclusively in ookinetes like SMC5, and RAD5 involved in DNA replication and repair before and during mitosis (Hirano, 2016; Ortiz-Bazán et al.,

2014). Of interest, PlasMei2, which in *P. yoelii* have been shown to be expressed exclusively in the liver stage where it controls chromosome segregation (Dankwa et al., 2016), was detected in the late program possibly indicating differences in expression with *P. yoelii*. Thus, together with other late program genes, PlasMei2 may have a role during the mitosis of the oocyst sporogony. Differential gene expression analysis determined a single cellular lineage for isolated cells, most of which aligned with the time points of isolation, although a number of individual cells showed temporal transcriptional dynamics aligning with earlier or later time points.

In *P. falciparum*, most transcription factors belong to the Apetala2 (AP2) family which consists of 27 genes. Some of these transcription factors are named according to the life cycle stage where they are expressed and exert their function, for example ApiAp2-g and its role in gametocyte differentiation (Ralph & Cortés, 2019). The progression through the life cycle in the mosquito midgut is dependent on the temporal regulation of ApiAP2 transcription factors, as indicated by the differential expression in pseudotime clusters. With the exception of AP2-O4, all AP2-O genes showed expression in our dataset, and the high resolution provided by single-cell transcriptomics, allowed the definition of the timing of AP2-O gene expression through the studied time course. Moreover, additional ApiAP2 genes appear to be upregulated in a significant proportion of cells (>50%). Two ApiAP2 genes were found in C2 (PF3D7_1239200 and PF3D7_1342900) and four genes in C4 (PF3D7_0420300, PF3D7_0934400, PF3D7_0802100 and PF3D7_1139300), none of which had previously been linked to a developmental stage. In agreement with our data, PF3D7_0934400 was shown to be highly expressed among ookinetes in the single-cell data by Real and colleagues (Real et al., 2021). Also, PF3D7_0622900: AP2Tel, which is significantly upregulated in C4 in our dataset, showed low to intermediate

428 expression among ookinetes in the dataset of Real and colleagues and high
429 expression among a portion of the oocyst-sporozoites. This could be indicative of its
430 involvement in regulating DNA replication genes in oocysts. AP2Tel has been
431 described to be expressed throughout the asexually replicating blood stage in *P.*
432 *falciparum,* where it was shown to bind directly to telomeric repeats of all 14
433 chromosomes (Sierra-Miranda et al., 2017). In addition, AP2Tel (AP2-SP3) has been
434 shown to have a role in sporozoite release from oocysts in *P. berghei* (Jeninga et al.,
435 2019; Modrzynska et al., 2017).
436 The timing of ApiAP2-O expression coincides with the down regulation of three genes:
437 P25, H2B and CPR. These genes are known to be i) a zygote surface protein (Dijk et
438 al., 2010), ii) a component of the nucleosome known for packaging chromatin (Bennett
439 et al., 1995), and interestingly iii), an antimalarial resistance-associated gene in
440 sporozoites (Fan et al., 2011), respectively. The timing of expression of these genes
441 raises the intriguing question of whether AP2-O may be involved in negative regulation
442 of their expression. The involvement of AP2 transcription factors in the negative
443 regulation of genes occurs in *Toxoplasma gondii* (Radke et al., 2018), but has only
444 been speculated on for *Plasmodium spp*. In later time points, AP2-O upregulates an
445 array of genes, among them SOAP and WARP, both important for ookinete stages and
446 important in invasion (Kaneko et al., 2015). Taken together, our results highlight the
447 coordinated expression of ApiAP2 transcription factors during the developmental
448 trajectory of *P. falciparum* in the mosquito midgut and beyond.
449 Since parasite numbers decline precipitously during mosquito midgut development, we
450 also sought to elucidate the genetic program involved in this critical bottleneck of the
451 *P. falciparum* life cycle. The parasite is inflicted by oxidative stress produced by; free
452 heme resulting from hemoglobin digestion, ROS, toxic molecules produced by the

midgut microbiome, human-derived immune and inflammatory molecules, and immune cells ingested with the blood meal (Smith et al., 2014). Subsequently, once the mature ookinete traverses the midgut epithelium, the mosquito complement immune response is activated, leading to increased susceptibility to nitric oxide toxicity (Oliveira et al., 2012). All of these factors contribute to the reduction of parasite numbers seeing during the major lifecycle bottleneck in the mosquito midgut (Griffin et al., 2010; Smith et al., 2014). Hence, we investigated the expression of genes involved in the response against ROS and Reactive Nitrogen Species (RNS), cellular stress and cell death. We found a temporal induction of known stress-related genes indicating that different stressors from the mosquito midgut environment are associated with specific stress responses from the parasite. For example, the redox genes nPrx, HSP90 and dTCTP were highly expressed in the intermediate stages (C1, C0) likely to facilitate survival within the harsh midgut conditions. Previous studies have reported the induction of similar genes, like 2-Cys peroxiredoxin and peroxiredoxin-1, due to the environment of the mosquito midgut and their role in protecting the parasite midgut stages from oxidative stress (Peterson & Luckhart, 2006; Turturice et al., 2013).

Previous studies have used I-TASSER (Yang & Zhang, 2015) to elucidate the structure and function of proteins of interest, including *in silico* experimentation for finding novel antimalarials (Pandey et al., 2018). We used I-TASSER to predict protein function of a set of highly expressed, non-annotated genes linked to the mid and late developmental points. Using 3D model predictions, we selected 17 non-annotated and highly expressed genes with a predicted TM domain and/or a signal peptide, indicating their possible membrane localization. These proteins were further indicated to be intrinsically disordered proteins (IDPs), proteins largely disordered in their structure

and reported to be highly immunogenic (Guy et al., 2015), a trait essential for antibody-based therapies.

The *P. falciparum* midgut stages represent a tenuous point in the parasite's life cycle given the large decline in population size and genetic diversity. Such small numbers render emergence of parasites resistant to therapeutics far less likely, as compared to human stages. These properties make the *Plasmodium* mosquito stages a highly strategic point for malaria intervention. Drugs which target this stage will be highly specific against the parasite and can be administered to cure mosquitos and indirectly prevent human infection (Paton et al., 2019). Further, human antibodies are potent, remain viable within the mosquito midgut and are becoming affordable for use in low income settings, indicating that transmission-blocking vaccines targeting the transmission stages of the parasite are possible and merit further exploration (de Jong et al., 2021; Kundu et al., 2018; Sauerwein & Bousema, 2015). Overall, our data provides an important resource for exploring the hitherto largely undescribed transcriptome of the mosquito midgut stages of *Plasmodium falciparum*, and offers untapped potential for the exploration of transmission-blocking therapeutics.

Materials and Methods

In vitro culture of *P. falciparum* parasites and gametocyte induction

Gametocytes from the *P. falciparum* gametocyte-producing NF54 cell line were induced and cultured as previously described in (Delves et al., 2016). In brief, parasites were cultivated on type O positive erythrocytes at 4% hematocrit in standard culture media containing RPMI supplemented with 10% heat inactivated human serum (Karolinska Hospital blood bank, Stockholm, Sweden) using standard culturing techniques. Cultures were gassed with 96% N_2, 1% O_2 and 3%CO_2 and maintained on

a shaking incubator at 37 °C to prevent multiple infections in RBCs. Gametocytes were then induced at 5-6% parasitaemia and maintained, with daily media changes for 14 to 17 days in gametocyte culture medium containing 25 ml of RPMI and 10% heat inactivated human serum. Media were changed on a daily basis after which mature gametocytes were harvested for mosquito infection on days 14 and 17.

Mosquito infections

Gametocyte cultures were harvested on day 17 and diluted to 1% gametocytemia at 50% hematocrit, and delivered directly into water-jacket glass-membrane feeders connected to a 37°C circulating water bath. Twenty *Anopheles gambiae* females (5-7 day old), maintained at 25-27^{o}C and 80% humidity, were allowed to blood-feed for about 30 minutes, after which any unfed mosquitos were isolated and discarded. Between 15 to 20 mosquito midguts were dissected at time points 2, 4, 8, 12 and 20 hours post infection. Mosquito midguts were collected in 1.5 ml Eppendorf tubes with 500µl of RPMI and homogenized using a micro-tube homogenizer (F65000-0000, SP-Bel-Art) at short intervals for 30 seconds. In addition, one batch of mature gametocytes were treated with aphidicolin (Sigma #A0781-1MG) *in vitro* and cultured for 2h at room temperature. Aphidicolin prevents DNA replication in male gametes rendering them immature and thus blocks female fertilization.

Staining, imaging, and isolation of single parasites

After midgut homogenization, 2µl of anti-Pfs25 antibody (1 µg/ml) was added to the sample and incubated at RT for 15 minutes. 1 µl of Alexa Fluor 488 goat anti-mouse IgG (H+L) at 2 µg/ml (Life Technologies, cat#A11001) and 0.5 µl of 10 mg/ml Hoechst 33342 was added to the sample and incubated for 15 minutes at RT in the dark. Cells

were washed two times with sterile DPBS (Dulbecco's Phosphate Buffered Saline, ThermoFisher Scientific, cat#A1285801), parasite pellets were subsequently resuspended in 500 µl of DPBS. 100 µl of stained parasites were placed on BSA coated petri dishes to prevent the adhesion of the parasite to the glass bottom of the petri dish. Targeted individual parasites were visualized and imaged using a Leica DMi8 microscope with K5 camera, 4.2 megapixel (Leica, Germany). Targeted cells were collected using capillary-based micro-manipulation with glass capillaries containing an inner diameter of 8µm (Eppendorf, Hamburg, Germany). Capillaries were pre-coated with a sterile 2% BSA solution to prevent the gametes from sticking to the glass capillary surface. Z-stacks taken of targeted single parasites and were then processed using IMARIS software (Version 9.3). Image processing analysis for each time point was set to a standard and was used to measure the mean fluorescence intensity and the volume of the nucleus against different cell types yielded by clustering analysis. Individual parasites were then captured and replaced in 1x DPBS and transferred into 200 µl thin-walled PCR tubes (Corning, NY) containing 3.5 µl of lysis buffer (0.6% Triton-X100) 2U/µl recombinant RNase inhibitor, 1 µl oligo-dT (10 µM) and 1 µl dNTP mix (10 mM). All samples were immediately stored at -80°C after isolation.

cDNA synthesis and library preparation

cDNA libraries of single parasites were generated using a modified version of the Smart-seq2 protocol (Picelli et al., 2014). In short, cDNA synthesis was performed using *P. falciparum* optimized primers (Real et al., 2021) and PCR amplification was carried out over 24 cycles. cDNA products were subsequently purified using CA Beads (Sigma, Cat. Nr. 81260) for size selection using 8.8% PEG6000 to exclude primer-

dimers and non-specific amplicons with sizes less than 150bp. Combinatorial indexing via tagmentation was carried out in 96-well plates using 200pg (measured in a Qubit) of amplified cDNA, for a final volume of 10 µl/well. cDNA fragmentation using Tn5 transposase was carried out for 20 minutes on ice using the Illumina Nextra XT DNA sample preparation kit. Ligation and amplification of adaptors was carried out over 15 cycles in a final volume of 25 µl/well. Primer indices were used in the reaction from illumina (Nextera index primers-i7 and i5 cat# FC-131-1001). Tagmented and barcoded amplicons were then purified using CA Beads for size selection. Quality control and fragment size distribution of the cDNA libraries were performed on a BioAnalyzer with the Agilent high sensitivity DNA chip cat# 5067-4626. Concentrations of each sample of cDNA libraries were measured on a Picogreen 96-well plate NucleoScan fluorometer using a high sensitivity dsDNA (HS assay kit, cat# Q32851). To perform library dilutions, the average fragment sizes of all cDNA libraries were measured for a final concentration of 2nM in each sample. Finally, cDNA libraries were pooled and sequenced using Illumina NextSeq with 75 bp paired-end reads.

Computational analysis:

scRNA-seq raw data mapping and feature counts

Raw scRNA-seq reads were processed and trimmed for quality and adapter content using FastQC (Version 0.11.5) (Andrews S. 2012) and Trimmomatic (Version 0.36) (M. Bolger 2014) respectively. All quality processed reads were mapped to Human (GCF_000001405.38_GRCh38.p12_genomic.fna), mosquito vector (GCF_000005575.2_AgamP3_genomic.fna) and *P. falciparum* parasite (GCF_000002765.4_ASM276v2_genomic.fna) using FastQ_Screen, and the reads mapping to parasite but not human and vector were retained. All orphan unpaired

reads were discarded and remaining paired reads were mapped to *P. falciparum* genome downloaded from PlasmoDB (PlasmoDB-39_Pfalciparum3D7_Genome.fasta) using STAR (Version 2.5.3a) (Dobin A. 2013). Read quantification was performed using HTseq-count (Anders S. 2015) (parameters: -t exon -i gene_id -r pos -m intersection-nonempty) and custom bash script were used to generate gene count matrix.

Quality control and normalization

For estimation of good quality cells to perform downstream analysis we used the Seurat package (v 4.0.3) on the raw features count matrix for an overview of the distribution of the number of reads and genes detected per cell within each time point. We set a cutoff to filter out cells that had fewer than 600 genes and 2500 reads resulting in 125 cells that were included for the downstream analysis. We then employed a global-scaling “LogNormalize” to the feature expression measurements for each single cell from the total expression multiplied by a scale factor (10,000) with the normalized values stored in the Seurat object.

Identification of highly variable features

In order to delineate cell-to-cell variation across the dataset and highlight the biological signal of highly expressed genes, we used the feature selection method “vst” implemented in the Seurat package (FindVariableFeatures) to find variable features across the single cells’ transcriptomes. The top 2000 variable genes were selected for the downstream analysis.

602 **Dimensionality reduction, cell clustering and projection**

603 To dissect cellular heterogeneity during the development of *P. falciparum* zygotes, we
604 first scaled the expression so that the variance across cells had an equal weight of 1
605 in the downstream analysis. We then performed linear dimensionality reduction
606 (principal component analysis – PCA) using the top highly variable genes (HVGs) on
607 the normalized-scaled transcriptome data to characterize the variation across the
608 dataset. To estimate the number of PCs indicating a true signal of transcriptional
609 variation, we visualized both the cells and features using the DimHeatmap function on
610 the first 20 PCs and then performed the JackStraw test for the P-value
611 estimation. Next, we used the first 10 PCs (most PCs showing the signal variation) to
612 delineate the transcriptome into cell type communities by running kNN graph-based
613 clustering and applying modularity optimization with the Louvain algorithm to relatively
614 group the cells with a resolution = 0.8. Different cell types' transcriptomes were then
615 visualized using UMAP non-linear dimensionality reduction. Detailed steps of the
616 analysis are explained in the data and code availability section.

617

618 **Differential gene expression analysis**

619 To define markers distinguishing each cluster generated by the Louvain algorithm, we
620 set a minimum percentage of cells expressing certain features to be detected at 25%.
621 Finding differentially expressed genes across the cluster was set with Logfc.threshold
622 of 0.25 using the "Wilcoxon rank sum test" with a minimum of 25% cells for cluster
623 specific markers comparing all remaining cells reporting positively and negatively
624 regulated markers to be used in GSEA (Gene set enrichment analysis) for particular
625 biological pathway analyses. The differentially expressed gene dataset is presented in
626 Supplementary Table 3.

627

628 **Cell lineage and Pseudotime analysis**

629 Louvain clusters were ordered along developmental trajectory using Slingshot (v1.8.0)
630 (Street et al., 2018). Non-fertilized female gametocytes (NFFGs) were excluded and
631 the feature expression matrix was normalized and scaled to estimate the total mRNA
632 abundance with the aid of the Seurat object tool along with metadata previously
633 generated from Louvain clustering. To restructure cell lineage and develop a
634 pseudotime inference for the purpose of uncovering the global mRNA structure of the
635 single-cell dataset, we used the MST (minimum spanning tree) algorithm and fit our
636 cells simultaneously against a "principal curve" to establish mRNA distribution in an
637 unsupervised manner. Applying the MST recovers a single lineage of developmental
638 trajectory composed of three global cell communities (Pseudotime regulatory
639 modules). Based on differentially expressed features recovered using the
640 FindAllmarkers function, specific cluster markers were identified and (C3) were
641 assigned as rooting cells or the lineage starting point (Initial cells) of the developmental
642 trajectory while the pseudotime values were estimated simultaneously for each single
643 cell ordered along the fitting curve. To visualize the pseudotime values specifically for
644 the Seurat clustering and the isolated time points, we overlaid the single cells'
645 metadata to the Slingshot lineage to have a better understanding of the relationship
646 between the clusters and the isolated time points (Figure 2B). To find genes that
647 change their expression over the course of the development, we used the Tradeseq
648 package 1.4.0 (Van den Berge et al., 2020) to calculate the relationship between gene
649 expression and pseudotime. In brief, we used a general additive model (GAM) to model
650 the relationship between genes and conduct an association test to estimate the *p*-value
651 of genes significantly expressed over time (Figure 2C and Supplementary Table 4). We

summarized snapshot transcripts of the global structure lineage and notable GO terms from PlasmoDB.org were used to distinguish pseudotime gene co-expression modules (pt0-pt2). To assess genes transcriptionally regulated during the course of development with significant variation in their expression, we employed the global clustering structure forming slingshot single lineage.

Motif identification for ApiAP2 regulators

We explored known ApiAP2 target sequence motifs using a Plasmodb search on transcriptionally regulated genes generated from pseudotime analysis (differentially expressed genes along the pseudotime) which returned only the genes with a specific known binding mix-base motif within a 1 kb upstream region of the starting codon for each ApiAP2 transcription factor target genes. Next, we visualized the target gene dynamics of expression in relation to the specific transcription factor showing the regulation of target genes over the course of pseudotime (Kaneko et al., 2015).

Biological pathways and gene set enrichment analysis (GSEA)

To track tests for top functional class enrichment among the global clusters building the pseudotime lineage, we used conservative markers generated for each cluster on PlasmoDB GO analysis tool to conclude the enriched ontology terms as previously mentioned. The GSE analysis was performed on differentially expressed genes over the pseudotime as input in cluster profiler v 3.18.1 and ggupset package v 0.3.1 with a p-value cut off = 0.05, minGSSsize = 3, maxGSSize=800 and scoreType=” pos" to estimate for biological process ontology changes over the pseudotime lineage and developmental progress (Supplementary Figure 3A and Supplementary table 1). The

676 top 20 biological processes were visualized using the Clusterprofiler package (v
677 3.18.1) dot plot function.
678
679 **Protein prediction analyses**
680 The amino acid sequences of significantly upregulated, non-annotated genes of *P.*
681 *falciparum* 3D7 were retrieved from the PlasmoDB website and processed for their
682 primary structures and physico-chemical properties. Briefly, Expasy´s ProtParam tool
683 (Colovos & Yeates, 1993) was utilized to calculate the physico-chemical characteristics
684 and the secondary structural properties including α helix, 310 helix, Pi helix, Beta
685 Bridge, Extended strand, Bend region, Beta turns, Random coil, Ambiguous states and
686 other states using MLRC/SOPMA (Guermeur et al., 1999) I-TASSER (Yang & Zhang,
687 2015) was used for the structure-based functional annotations. Predicted proteins with
688 high confidence prediction annotations were then examined for the presence of signal-
689 peptide or TM domains. Alpha-Fold (protein structure database) was utilized to retrieve
690 the .pdb files of the predicted protein structures (Jumper et al., 2021). I-Tasser
691 Modrefiner was used for structure refinement (Xu & Zhang, 2011) based on atomic-
692 level energy minimization. The accuracy and stereochemical features of the predicted
693 models were then calculated with PROCHECK on the PDBsum server (Laskowski et
694 al., 1996) using “protein structure and model assessment tools”. Three putative
695 membrane protein 3D models were then subjected to druggability assessment using
696 DogSitescorer server for druggable pocket predictions (Supplementary Figure 7).
697
698 **Identification of intrinsically disordered proteins (IDP)**
699 We computationally characterized the 3D homology structures of protein models
700 showing a lack of well-characterized protein segments (IDP) using CSpritz server for

accurate detection of protein disorder (Walsh et al., 2011) in combination with comparative dynamic simulations with python-based docking libraries (py3Dmol, openbabel and nglview). Briefly, .pdb files retrieved from Alpha-fold and imported in Jupyter notebook with perquisites libraries imported and compared with well-structurally annotated proteins. We then identified the characteristics of bulky hydrophobic amino acids sequences with high net charges promoting disorder in form of extended loops regions coupled with folding and binding compared to the core structure resulting in instability and irregularity of the secondary structure (van der Lee et al., 2014).

Data integration of 24-hour Ookinetes (Day 1) Real et al 2021

Expression matrices and supporting metadata files were downloaded from https://zenodo.org/record/4719664/. Single-cell data were subset according to the target stage (Day 1 Ookinete). Variable genes were intersected between the two datasets using Scanpy (v1.5.0) (Wolf FA. 2018). We then used the concatenate and ingest function to integrate data annotations and labels and corrected for batch effect using BBKNN integrated in scanpy workflow.

Data and code availability

In-house bash, R code scripts and data that were implemented in this study are available on GitHub [https://github.com/ANKARKLEVLAB/Single-cell-P.falciparum-midgut]. Expression matrices and meta data are available via [https://zenodo.org/deposit/4683823] and the data is also searchable via [https://mubasher-mohammed.shinyapps.io/shinyapp/].

Acknowledgment

We would like to thank The Microbial Single Cell Genomics (MSCG) facility, SciLifeLab, for assistance with the single-cell sorting and optimization of molecular workflows and library preparations and the National Genomics Infrastructure (NGI) for Computing (SNIC) for data handling and preprocessing of the scRNA-seq raw data for this study, partially funded by the Swedish Research council through grants agreement 2018-05973 to J.A. The computations were performed using resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2020/16-146 and SNIC 2021/22-492.

This work was supported by grants from the Swedish Society for Medical Research (SSMF), The Swedish Research Council (VR-NT), The Jeansson Foundation and SciLifeLab to J.A.; The NIH Distinguished Scholars Program and the Intramural Research Program of the Division of Intramural Research AI001250-01, NIAID, National Institutes of Health; the Strategic Research Area (SFO) program of the Swedish Research Council (VR) through Stockholm University to V.S and M.R.F; the Swedish Research Council (VR) #2021-06602 to J.H.; and NIH grant R01AI031478 to M.J.L..

Authors contributions

J.A. and **J.V.R.** conceived the study; **J.A., J.V.R., M.R.F.** and **M.J.L.** supervised the project; **M.M., J.V.R.** and **J.A.** performed cell culture, infections, dissections, staining, cell imaging and micromanipulation for cell isolation; **N.E.** provided technical support and assistance with mosquito culturing and infections; **M.M.** performed Smart-seq2, library preparations; **M.M.** and **I.B.** performed sequencing; **V.S.** performed scRNA-seq

751 data preprocessing and quality control; **M.M.** performed computational analysis, data
752 integration and interactive web visualization; **B.B.** and **A.J.** performed the iTasser
753 analyses. **J.H. J.A. and M.R.F.** supervised the computational analysis; **J.V.R., M.K.,**
754 **T.S.** performed the imaging analysis; **A.D.** conceptualized and coordinated data
755 visualization, **J.A., M.M., AD** wrote the manuscript with help of **J.V.R., M.J.L.** and **J.H**,
756 all authors edited and provided critical input.
757

758 **Competing interests**

759 The authors declare no competing interest.
760

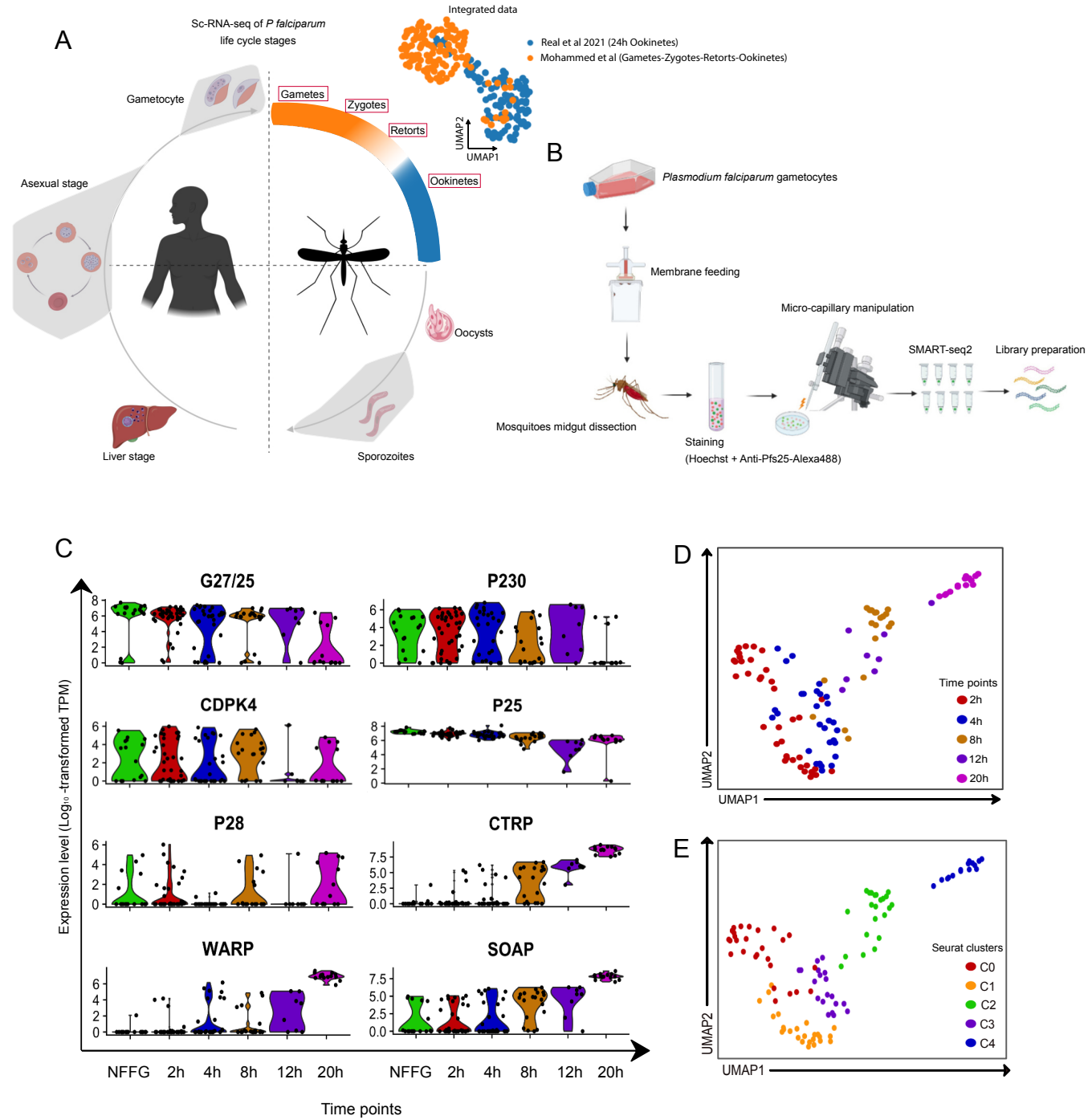
761 **Additional information**

762 **Supplementary information is available for this article.**

763 Correspondence and requests for materials should be addressed to:

764 johan.ankarklev@su.se or joel.vega-rodriguez@nih.gov

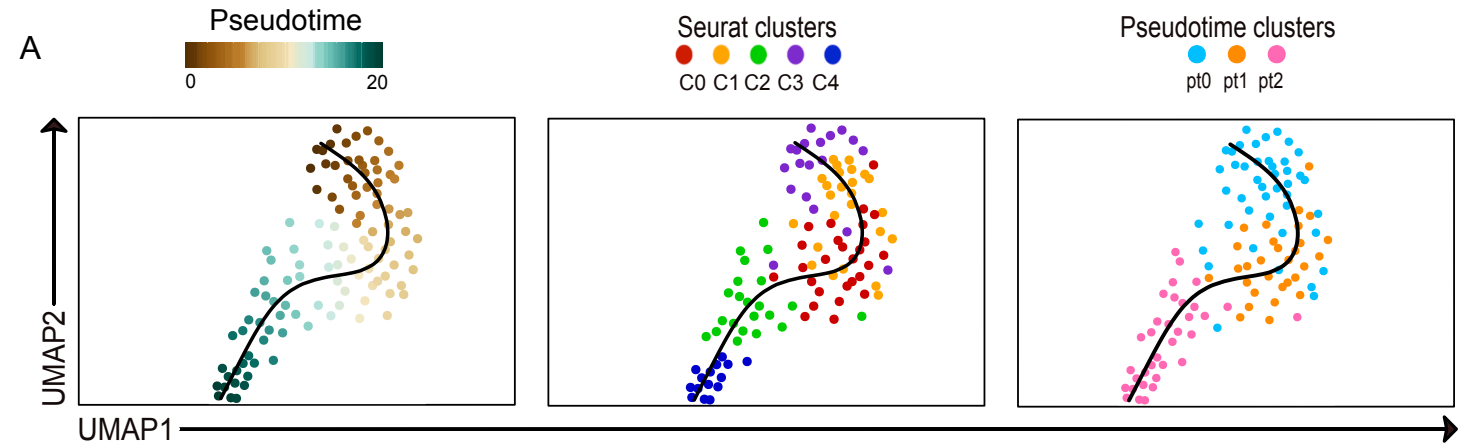
Figure 1

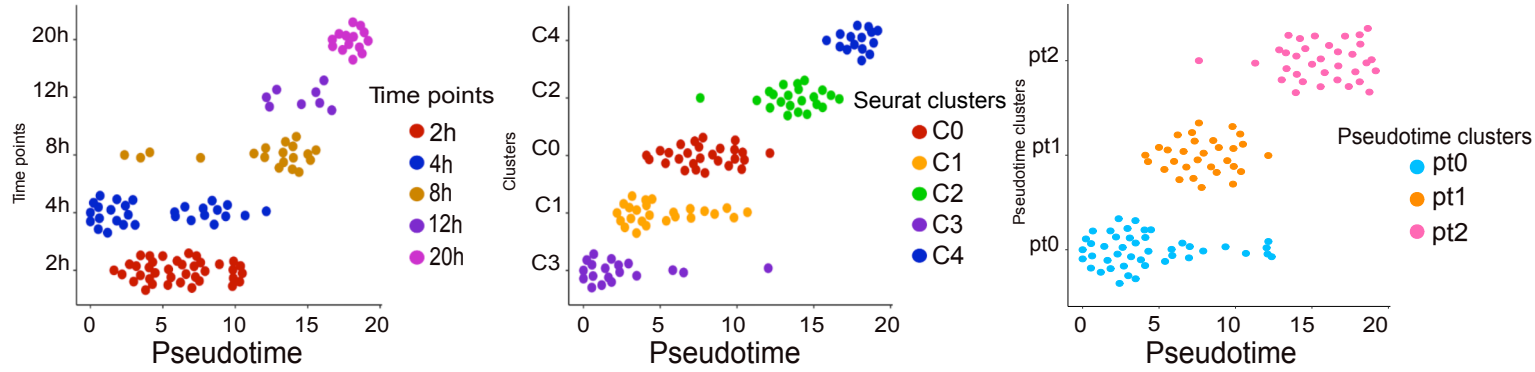

A
Sc-RNA-seq of P falciparum
life cycle stages
Integrated data
Real et al 2021 (24h Ookinetes)
Mohammed et al (Gametes-Zygotes-Retorts-Ookinetes)
UMAP2
UMAP1
Gametocyte
Gametes
Zygotes
Retorts
Ookinetes
Asexual stage
Oocysts
Liver stage
Sporozoites
B
Plasmodium falciparum gametocytes
Membrane feeding
Micro-capillary manipulation
SMART-seq2
Library preparation
Mosquitoes midgut dissection
Staining
(Hoechst + Anti-Pfs25-Alexa488)
C
G27/25
P230
CDPK4
P25
P28
CTRP
WARP
SOAP
Expression level (Log10 -transformed TPM)
NFFG 2h 4h 8h 12h 20h
Time points
D
Time points
2h
4h
8h
12h
20h
UMAP2
UMAP1
E
Seurat clusters
C0
C1
C2
C3
C4
UMAP2
UMAP1

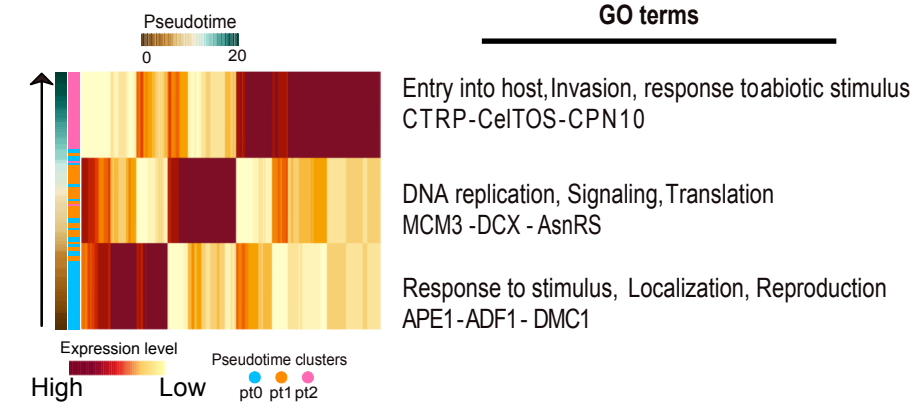
765 **Figure 1. Transcriptomic analysis of single *P. falciparum* parasites isolated from**
766 **the midgut blood bolus of *An. gambiae* mosquitoes.**

767 **A**. Schematic diagram of the *P. falciparum* lifecycle, highlighting the coverage of single-
768 cell transcriptome data across the lifecycle that makes up the current malaria atlas.
769 The UMAP shows an overlay of the scRNA-seq data of 24h ookinetes from Real *et al*.,
770 2021 (blue) and the current study (orange, top right) (Illustrations created using
771 www.biorender.com). **B.** Schematic diagram of the experimental pipeline: *in vitro*
772 cultured gametocytes were fed to *An. gambiae* female mosquitoes by standard
773 membrane feeding. Mosquito midguts were dissected at six different time points,
774 homogenized, stained and the homogenate was placed in an inverted fluorescence
775 microscope. Parasites were selected based on their positivity to Hoechst and Anti-
776 Pfs25-Alexa-488 staining by fluorescence microscopy and isolated by micro-
777 manipulation. cDNA libraries were prepared for each cell using a modified version of
778 the Smart-seq2 protocol. **C**. Violin plots of the average expression level (Log_{10}
779 transformed TPM) of marker genes for *P. falciparum* midgut stages. The marker genes
780 were selected from Bennink *et al*. 2016 for validation of our single-cell transcriptome
781 dataset. **D** and **E**. UMAP of the single-cell transcriptome data overlaid with the
782 collection time points at which they were isolated (D) or the respective Seurat clusters
783 (E). These plots show the two-dimensional projection of five isolated time points. The
784 global transcriptome similarities and differences were assessed using a k-nearest
785 neighbors (kNN) force-directed graph on the first 10 Principal Components with true
786 signal variation from our single-cell transcriptome dataset based on Elbow plot.
787
788
789

Figure 2







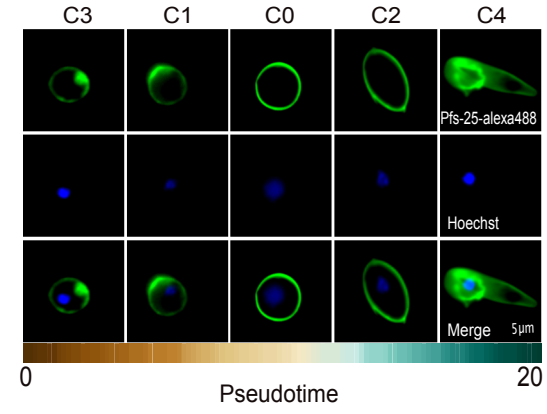


Figure 2. Ordering cells based on the developmental trajectory

A. Ordering of cells along their pseudotime developmental trajectory using Slingshot. Left Panel: Pseudotime alignment of cells used in this study, where each dot represents individual parasites and their positioning is determined by the total relative expression in comparison to the other parasites included in the analysis. The color coding represents their predicted development along the pseudotime axis. Middle panel: The pseudotime developmental trajectory overlaid with the assigned Seurat clusters, indicating that C3 is at the root of the trajectory and C4 represents the terminal state. Right panel: The pseudotime line overlaid with three unique mRNA patterns based on overarching similarity among major cell communities formed during the process of lineage reconstruction. **B**. Slingshot ordering of cells plotted over a pseudotime axis. Left panel: cells from each collection timepoint plotted on the pseudotime axis. Middle panel: cells representing each of the five Seurat clusters plotted on the pseudotime axis. Right panel: pseudotime clusters plotted on the pseudotime axis. **C**. Heatmap showing the average transcript abundance across the pseudotime (pt) axis, based on the Slingshot clustering (pt0, pt1 and pt2). Notable GO terms are indicated for each of the three pt clusters. **D**. Representative Z-stack images of developmental stages based on the ordering of the Seurat clusters (C0-C4) over the pseudotime axis. Images depict the surface marker P25-Alexa488 (green, top panel), nuclear Hoechst stain (blue, middle panel), and merged P25-Alexa488+Hoechst images (bottom panel).

Figure 3

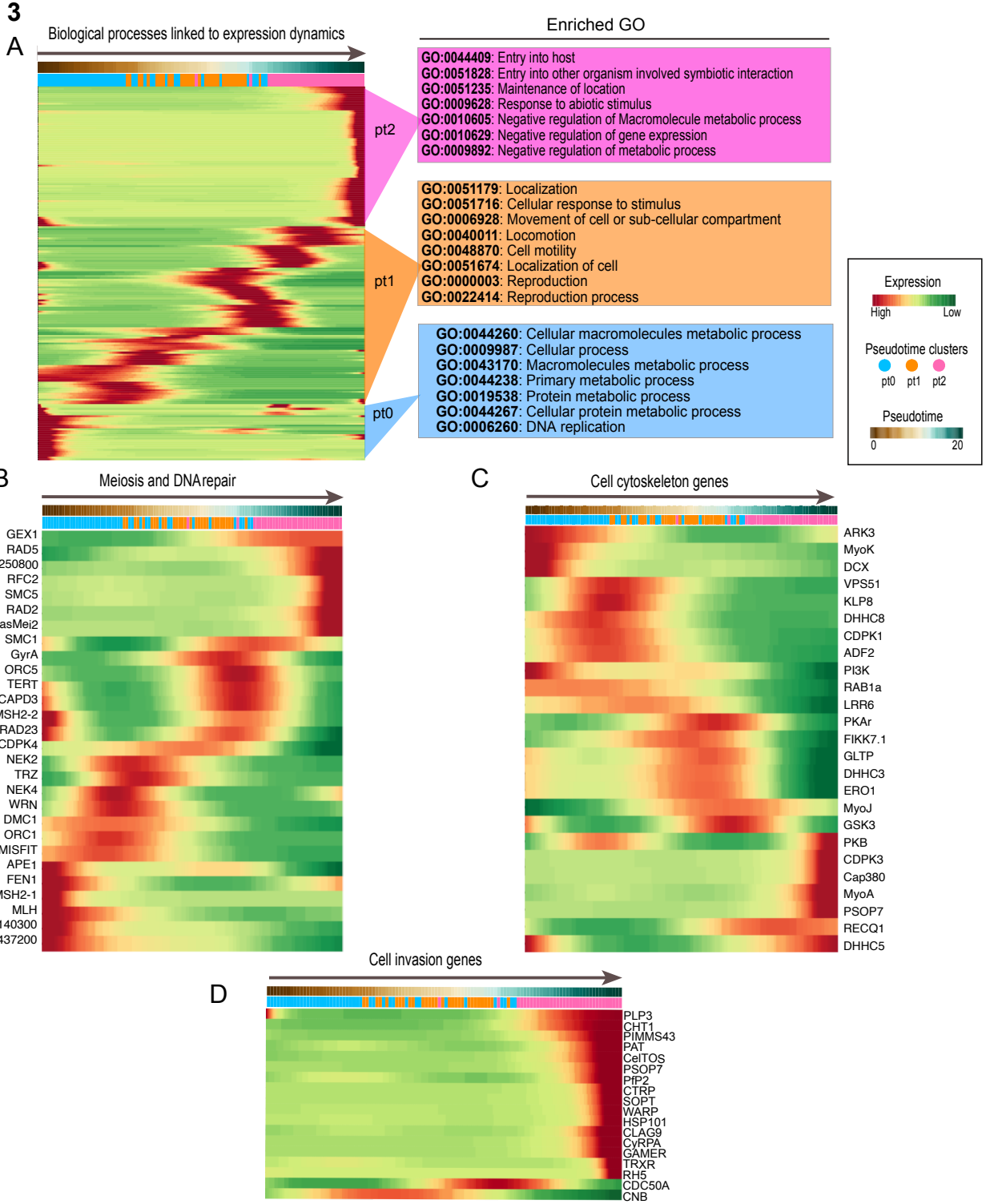
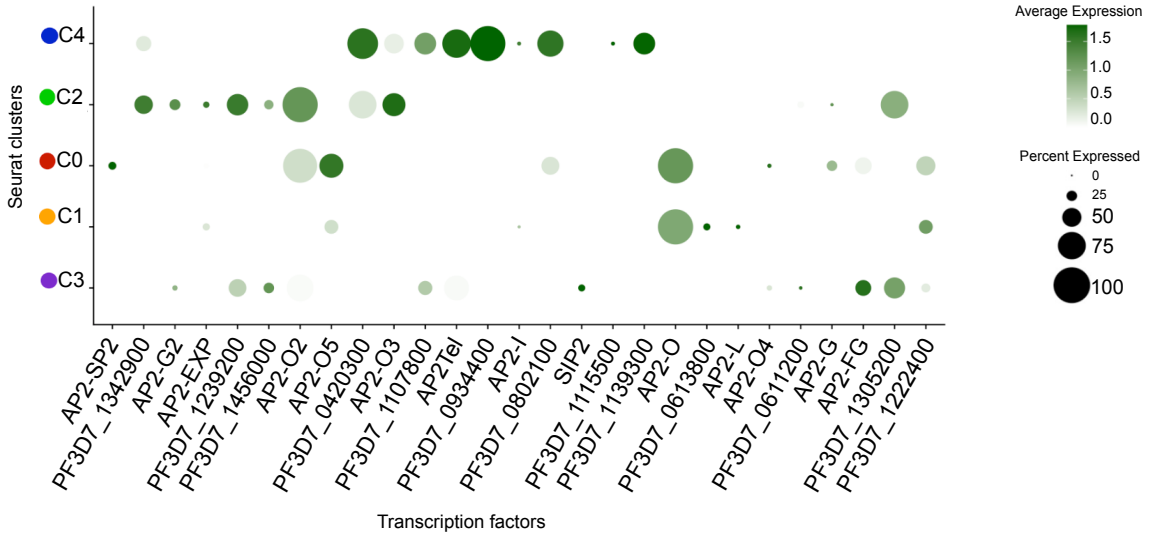

A
Biological processes linked to expression dynamics
Enriched GO
pt2
pt1
pt0
GO:0044409: Entry into host
GO:0051828: Entry into other organism involved symbiotic interaction
GO:0051235: Maintenance of location
GO:0009628: Response to abiotic stimulus
GO:0010605: Negative regulation of Macromolecule metabolic process
GO:0010629: Negative regulation of gene expression
GO:0009892: Negative regulation of metabolic process
GO:0051179: Localization
GO:0051716: Cellular response to stimulus
GO:0006928: Movement of cell or sub-cellular compartment
GO:0040011: Locomotion
GO:0048870: Cell motility
GO:0051674: Localization of cell
GO:0000003: Reproduction
GO:0022414: Reproduction process
GO:0044260: Cellular macromolecules metabolic process
GO:0009987: Cellular process
GO:0043170: Macromolecules metabolic process
GO:0044238: Primary metabolic process
GO:0019538: Protein metabolic process
GO:0044267: Cellular protein metabolic process
GO:0006260: DNA replication
Expression
High
Low
Pseudotime clusters
pt0 pt1 pt2
Pseudotime
0
20
B
Meiosis and DNArepair
GEX1
RAD5
PF3D7_1250800
RFC2
SMC5
RAD2
PlasMei2
SMC1
GyrA
ORC5
TERT
CAPD3
MSH2-2
RAD23
CDPK4
NEK2
TRZ
NEK4
WRN
DMC1
ORC1
MISFIT
APE1
FEN1
MSH2-1
MLH
PF3D7_1140300
PF3D7_1437200
C
Cell cytoskeleton genes
ARK3
MyoK
DCX
VPS51
KLP8
DHHC8
CDPK1
ADF2
PI3K
RAB1a
LRR6
PKAr
FIKK7.1
GLTP
DHHC3
ERO1
MyoJ
GSK3
PKB
CDPK3
Cap380
MyoA
PSOP7
RECQ1
DHHC5
D
Cell invasion genes
PLP3
CHT1
PIMMS43
PAT
CelTOS
PSOP7
PfP2
CTRP
SOPT
WARP
HSP101
CLAG9
CyRPA
GAMER
TRXR
RH5
CDC50A
CNB

Figure 3. Top differentially expressed genes across the pseudotime axis

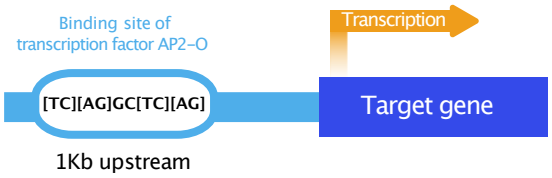
A. Hierarchical clustering of the top 200 expressed genes across the Slingshot pseudotime axis. The genes were grouped according to their ontology terms and analyzed for the top GO terms corresponding to each pseudotime cluster is indicated to the right of the heatmap. **B-D**. Heatmap of genes directly or indirectly involved in meiosis and DNA replication (B), cytoskeleton remodeling (C) and cell invasion (D) across the Slingshot pseudotime axis. For all heat maps, genes were selected based on their biological profile and annotation in Plasmodb and clustering was based on the Euclidean distance of scaled rows.

Figure 4

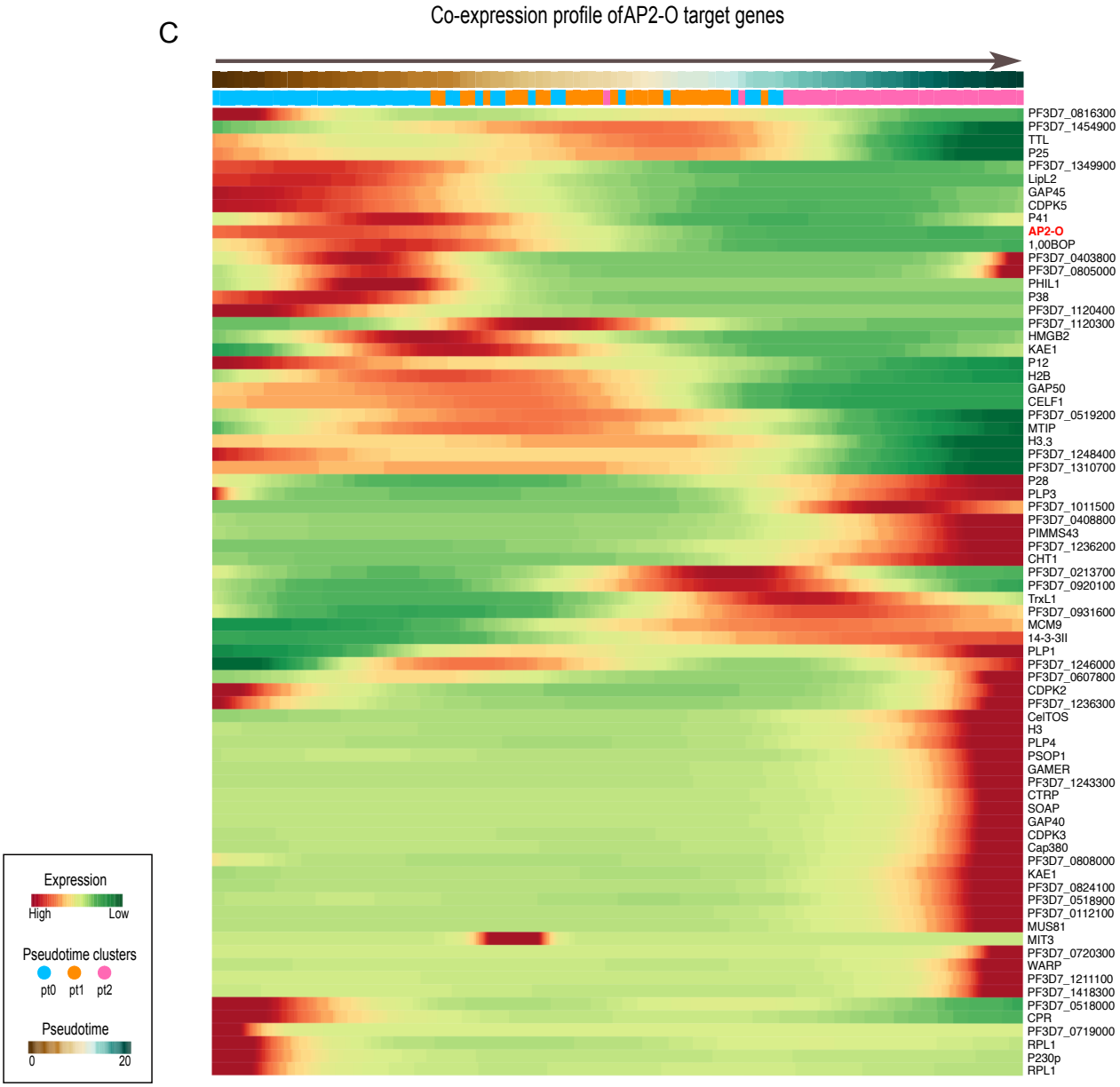
A



B



C



826 **Figure 4. Gene expression patterns of the ApiAP2 transcription factors**
827 **throughout the *P. falciparum* development in the mosquito midgut**

828 **A**. Dot plot showing the average expression and the proportion of cells that express
829 each of the 27 ApiAP2 transcription factors across the Seurat pseudotime clusters. **B**.
830 Schematic illustration of the process of validating genes harboring an ApiAP2-O
831 binding site, including a motif screen of the 1 KB upstream region from the transcription
832 start site of select genes. **C**. Pseudotime heatmap demonstrating the timing of
833 expression of the ApiAP2-O transcription factor and its putative cis-target genes.
834 Scaled expression values of the genes were ordered based on high (red) or low (green)
835 differential gene expression (DGE) across the pseudotime axis.

836

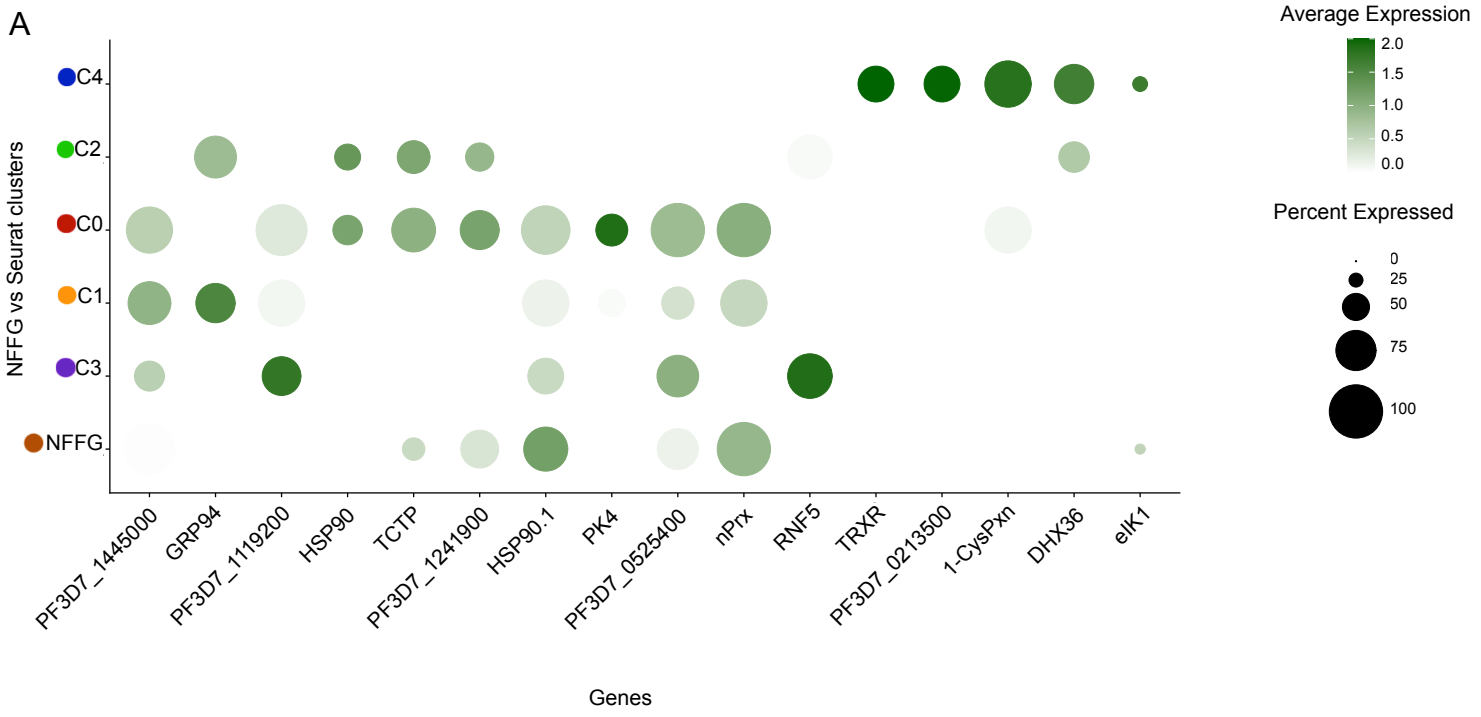
837

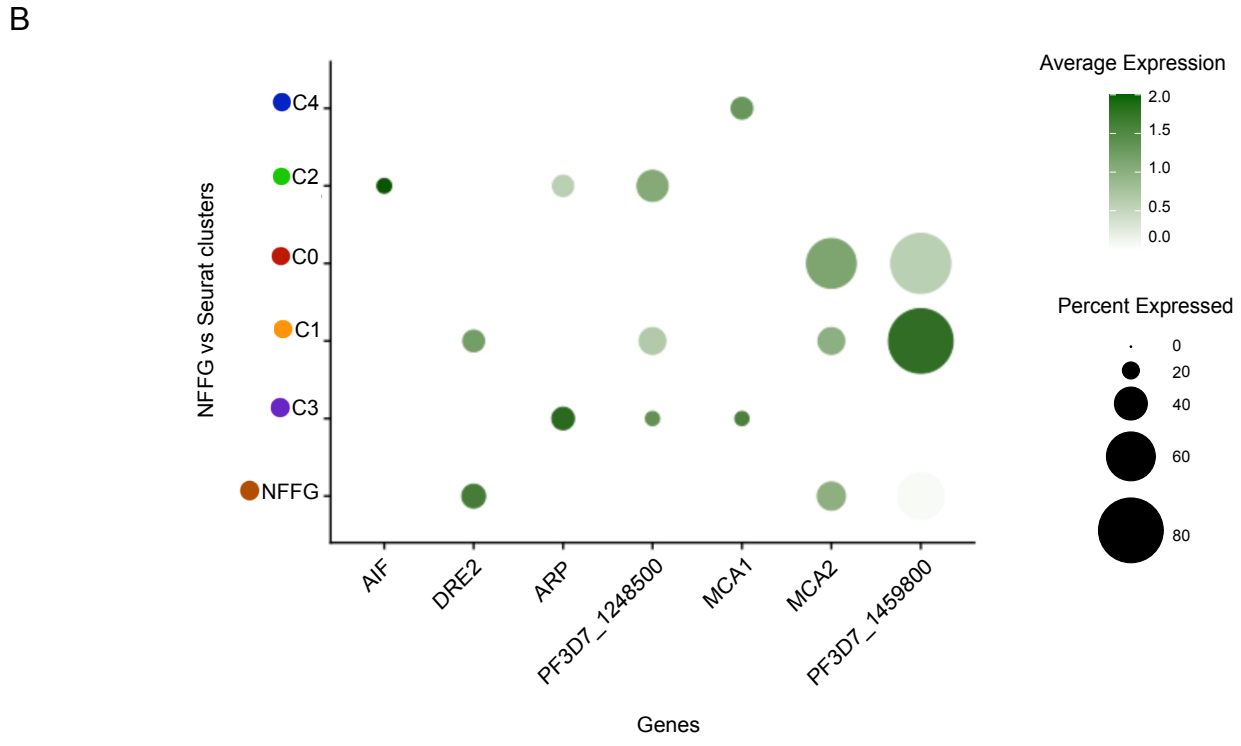
838

839

840

Figure 5





841 **Figure 5. Timing of expression of genes linked to parasite cellular stress and**
842 **apoptosis**

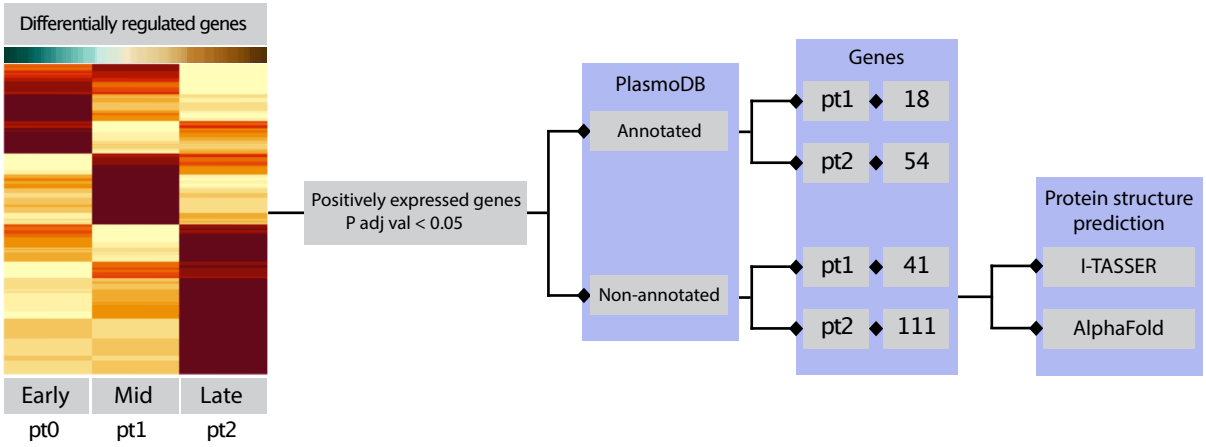
843 **A**. Dot plot comparing the average expression of stress response genes among cell
844 clusters (post-infection) versus non-fertilized female gametes (NFFG) isolated directly
845 from cell cultures. **B**. Dot plot comparing the average expression of apoptosis-related
846 genes between post-infection cell communities inside the mosquito and NFFGs.

847

848

Figure 6

A



B

Gene ID	PlasmoDB annotation	TM domain-Count	Signal peptide
PF3D7_0827400	conserved Plasmodium protein, unknown function		
PF3D7_0904300	conserved protein, unkonwn function		
PF3D7_0925100	conserved Plasmodium protein, unknown function		
PF3D7_1109100	conserved Plasmodium protein, unknown function		
PF3D7_1316700	conserved protein, unkonwn function		

Present Absent

C

Gene ID	PlasmoDB annotation	TM domain-Count	Signal peptide
PF3D7_0408800	conserved Plasmodium protein, unknown function		
PF3D7_0419400	conserved Plasmodium protein, unknown function		
PF3D7_0614900	conserved Plasmodium membrane protein, unknown function		
PF3D7_0926600	conserved Plasmodium membrane protein, unknown function		
PF3D7_0929300	conserved Plasmodium protein, unknown function		
PF3D7_1024000	conserved Plasmodium protein, unknown function		
PF3D7_1025400	conserved Plasmodium membrane protein, unknown function		
PF3D7_1106500	conserved Plasmodium protein, unknown function		
PF3D7_1117200	conserved Plasmodium protein, unknown function		
PF3D7_1136200	conserved Plasmodium protein, unknown function		
PF3D7_1227300	conserved Plasmodium protein, unknown function		
PF3D7_1306200	conserved protein, unkonwn function		

Present Absent

D

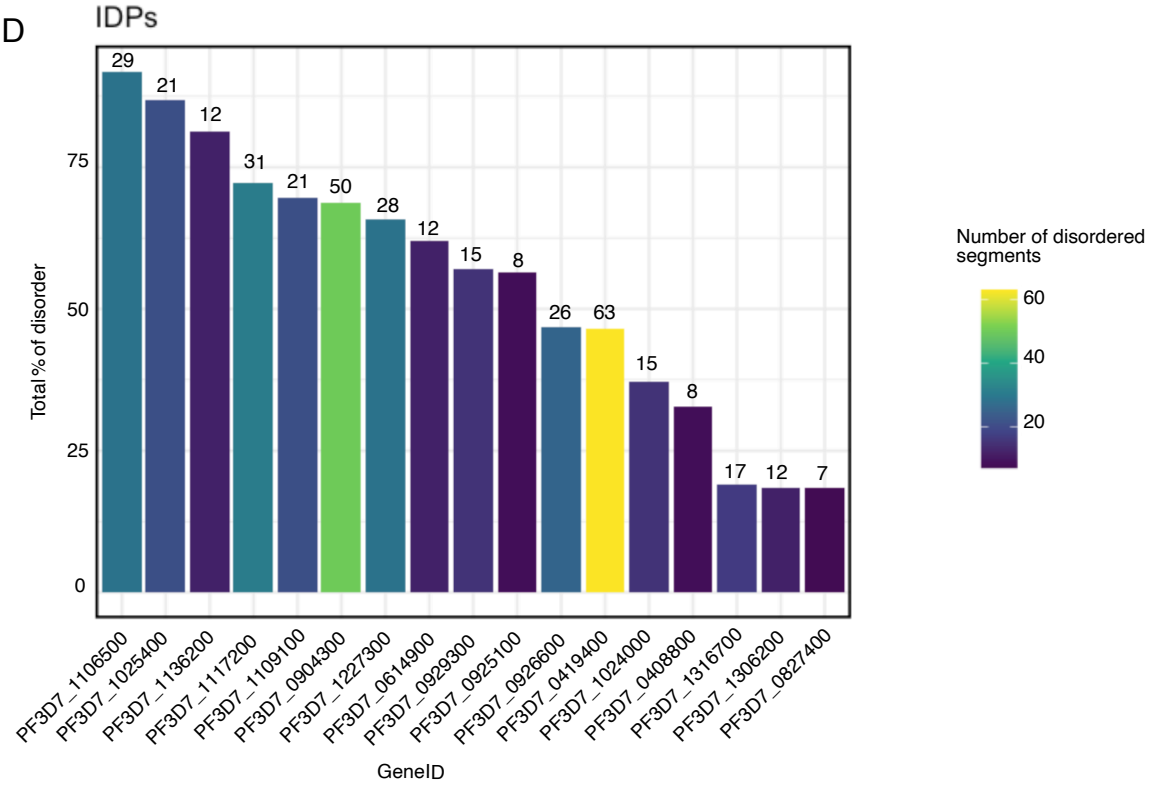
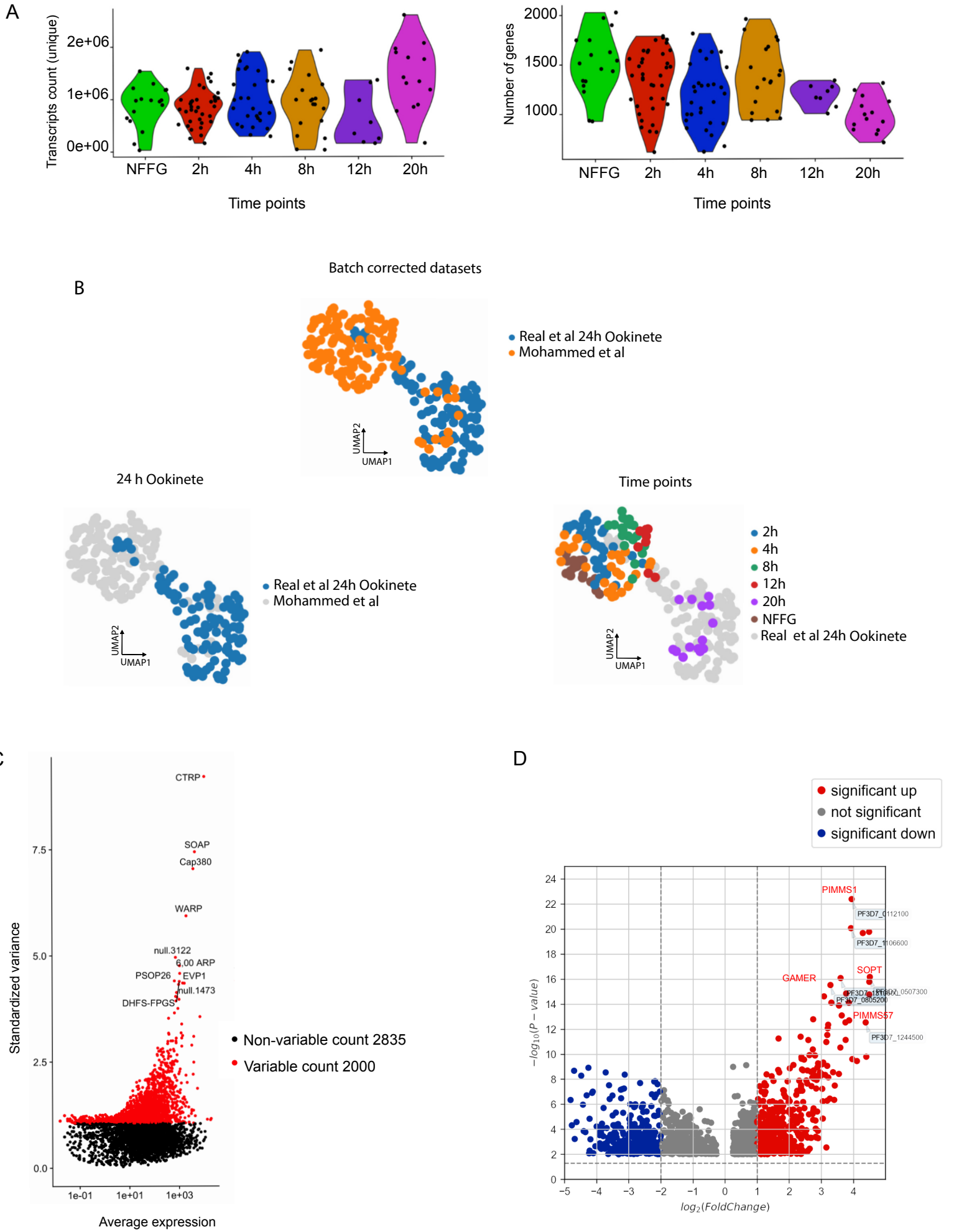


Figure 6. Structural predictions of putative therapeutic targets

A. Flow diagram depicting the selection process of highly expressed, non-annotated genes. Genes with *P* adj values of < 0.05 were selected in pt1 and pt2, and further analyzed using PlasmoDB, I-TASSER and Alpha-fold, which indicate a significant likelihood that certain genes encode membrane proteins and signal peptides. **B**. Table of genes from (A) linked to the pt1 cluster, showing the PlasmoDB annotation and the presence (green) or absence (pink) of TM domains and/or signal peptides. **C.** Table of genes from (A) linked to the pt2 cluster, showing the PlasmoDB annotation and the presence (green) or absence (pink) of a TM domain and/or signal peptides. **D.** Bar plot depicting the total percentage of disordered residues (Y-axis) in the predicted protein structures of the highly expressed, non-annotated genes (X-axis). The color scale indicates the number of disordered segments present in each protein. The 3D structures were analyzed using the CSpritz server.

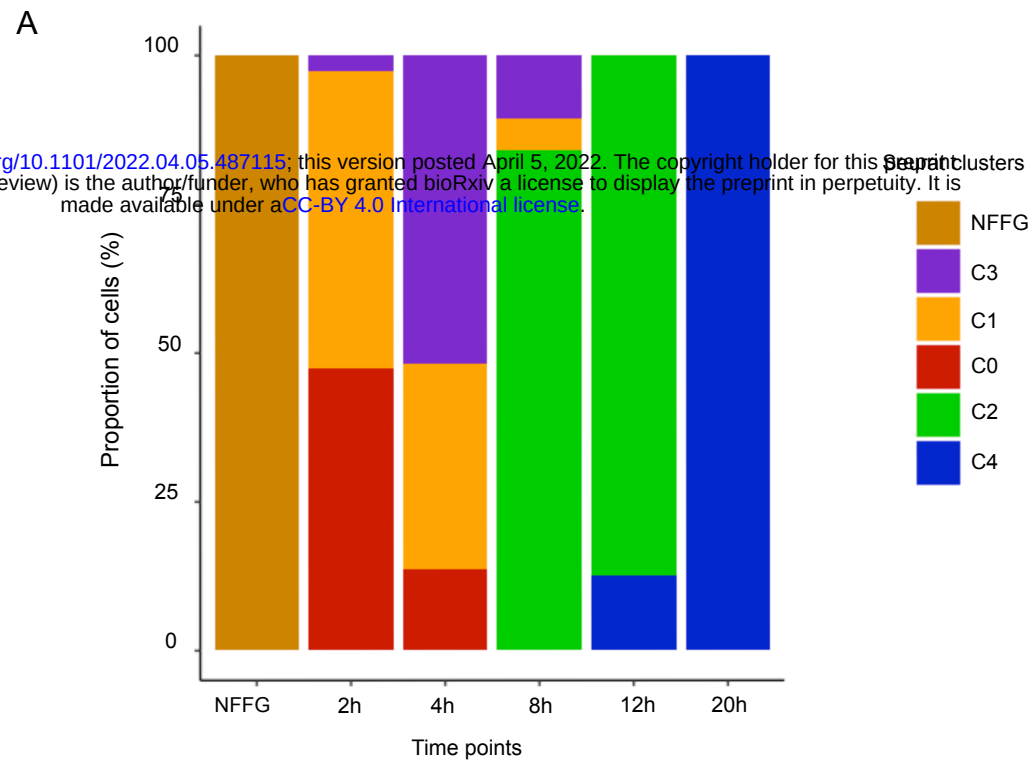
Supplementary Fig 1

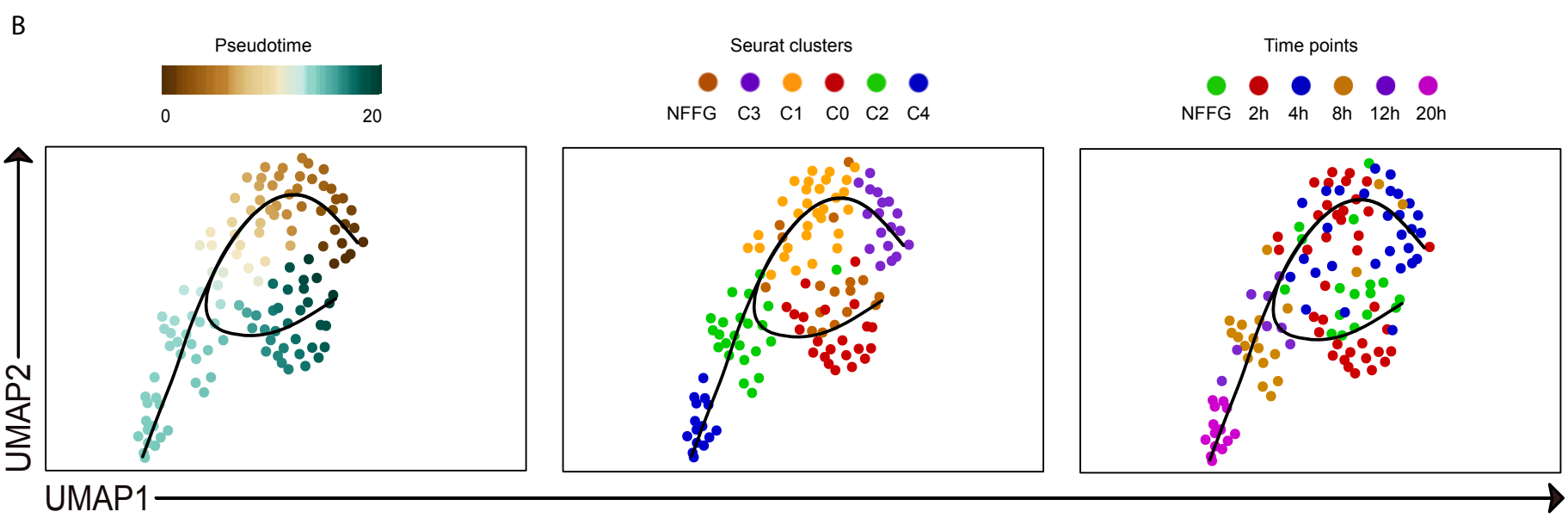


Supplementary Fig 1. Quality assessment of Smart-seq2 data and data integration with Real *et al*. 2021

A. Violin plots showing the number of unique transcripts detected per cell (Left panel) and the number of genes detected per cell (Right panel) across the six collection timepoints. **B**. Data integration of Real *et al.* 2021 day 1 (24h) single-cell transcriptomes with time points determined in this work non-fertilized female gametes (NFFG) to 20 h single-cell transcriptomes. Upper panel: shows integrated datasets. Bottom panel left: integrated datasets highlighting (blue) 24h ookinetes from Real *et al*. . Bottom panel right: integrated datasets highlighting the isolated time points determined in this work. After batch correction, more single cells from this work represent early development of *P. falciparum* in mosquito midgut stages. **C**. Volcano plot depicting the most highly variable genes, with the top 10 genes labelled. A total of 2,835 features were identified as non-variable while 2,000 features were variable and selected for downstream analysis. **D**. Volcano plot showing differentially expressed genes, based on Seurat cluster comparison, color-coded as red for significantly up-regulated, blue for significantly down-regulated, or Grey for not significant. The x-axis represents the fold change while the y-axis represents the *P*-value. A total of 1,576 differentially expressed genes were identified

Supplementary Fig 2





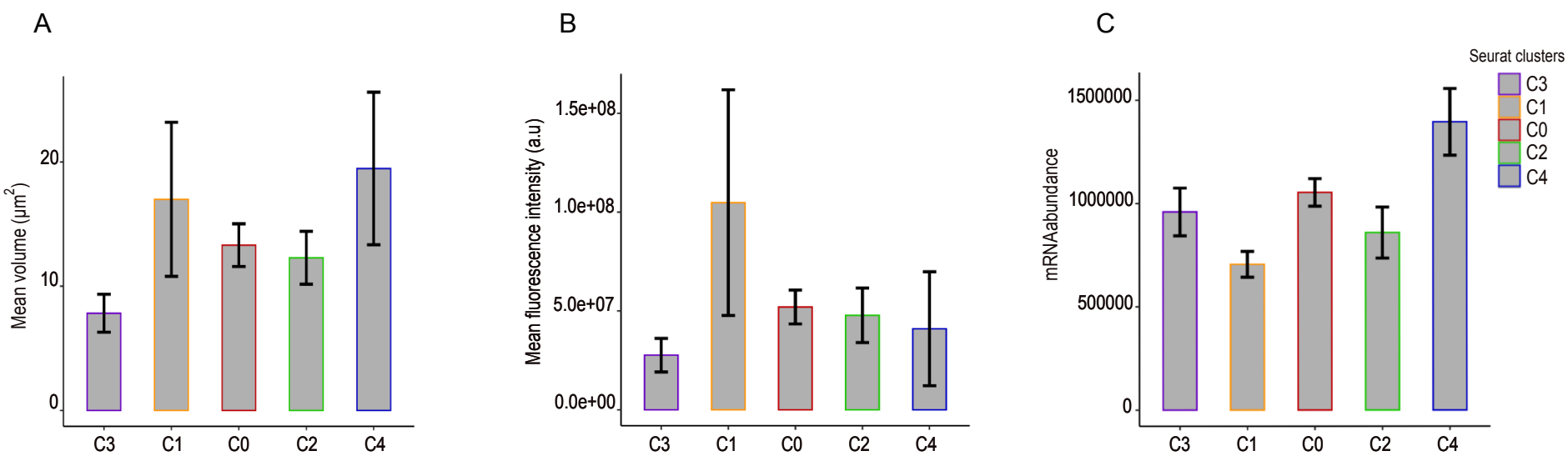
884 **Supplementary Fig 2. Comparison of the Slingshot trajectory analysis in the**
885 **presence or absence of the non-fertilized female gametes (NFFG)**

886 **A**. Bar plot showing the ratio of Seurat clusters identified within each collection time
887 point. **B.** The UMAPs indicate a modified trend of the Slingshot pseudotime trajectory
888 when NFFGs are included in the analysis as compared with the trajectories shown in
889 Figure 2A, where the inclusion of the single-cell transcriptomes from NFFGs appears
890 to distort the global lineage reconstruction. Left panel: The alternative pseudotime
891 trajectory where the color coding represents the cells predicted development along the
892 pseudotime axis. Middle panel: The pseudotime developmental trajectory overlaid with
893 the assigned Seurat clusters and where NFFGs are separately color coded in brown.
894 Right panel: The pseudotime developmental trajectory overlaid with the cell isolation
895 timepoints and including NFFG.

896

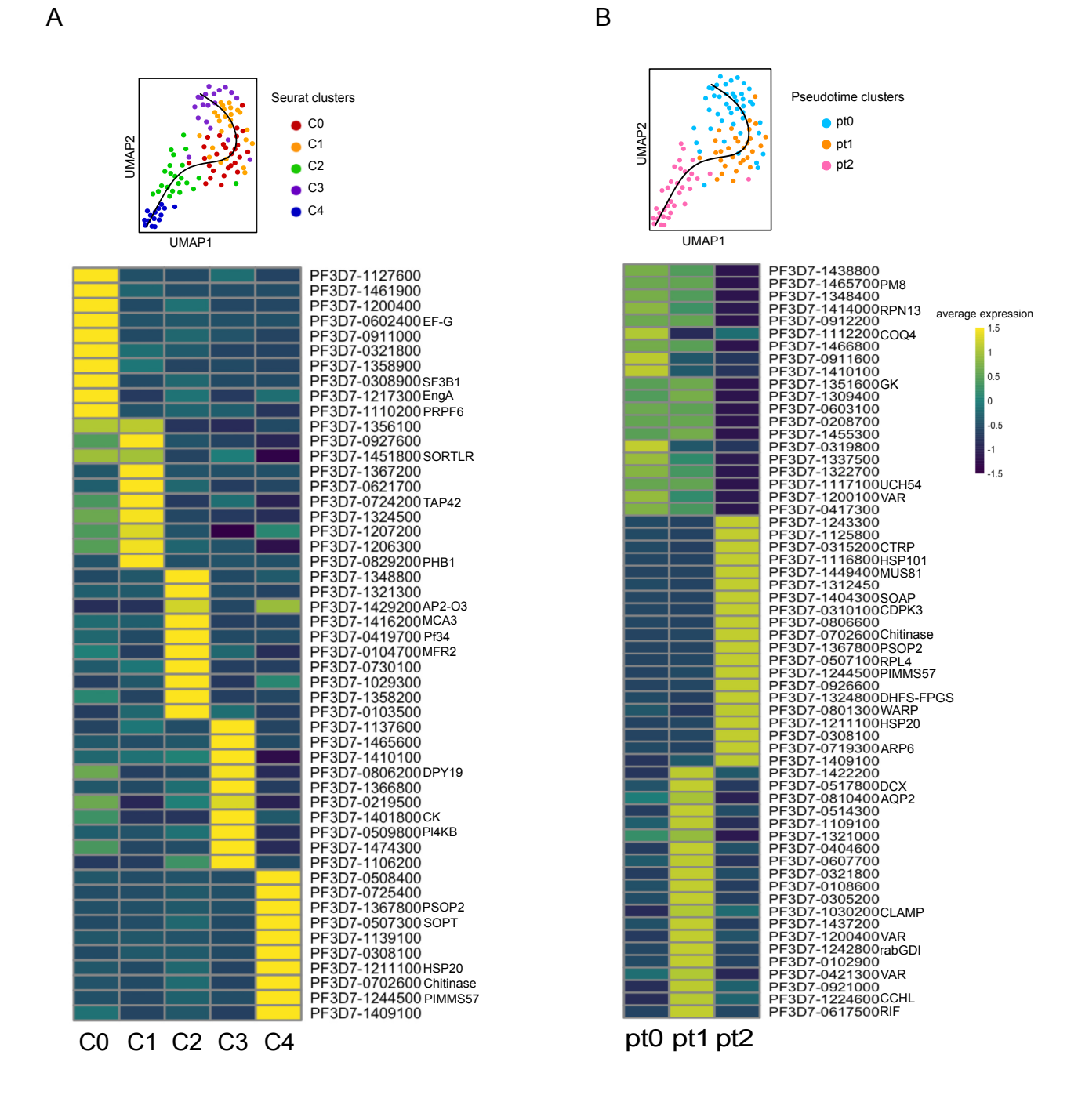
897

Supplementary Fig 3



898 **Supplementary Fig 3. Imaging parameters of ordered Seurat clusters leveraging**
899 **Pseudotime in characterizing nuclear phenotype features.**
900 **A**-**C**. Bar graphs showing the mean volume of the nucleus with standard errors (A),
901 mean fluorescence intensity (B) and mean value of total mRNA abundance (C), among
902 ordered Seurat clusters.
903
904

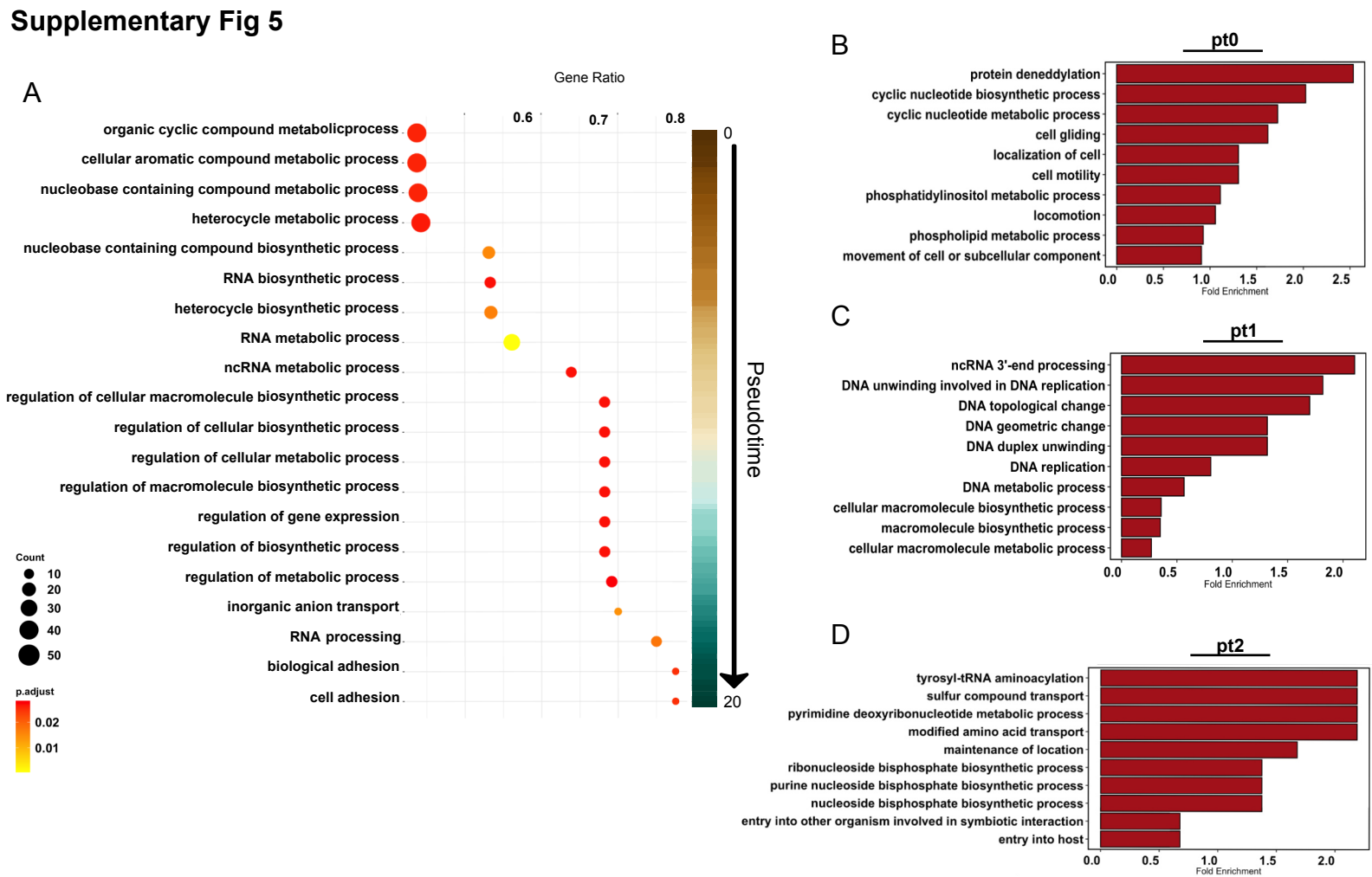
Supplementary Fig 4



Supplementary Fig 4. Comparison of top differentially expressed genes between the Seurat and Slingshot clusters along the pseudotime trajectory

A. The Slingshot pseudotime developmental trajectory overlaid with the Seurat clusters (top) with a hierarchical clustering heatmap showing the top 10 differentially expressed genes for each individual cluster (bottom). **B**. Slingshot pseudotime developmental trajectory overlaid with the Slingshot pseudotime clusters (pt0, pt1, pt2) (top) with a hierarchical clustering heatmap of the top 20 differentially expressed genes for each individual cluster (bottom). The Wilcoxon rank test was used in the differential gene expression analysis (for both A and B) of genes significantly expressed in at least 25% of cells and logFC threshold of 0.25.

Supplementary Fig 5



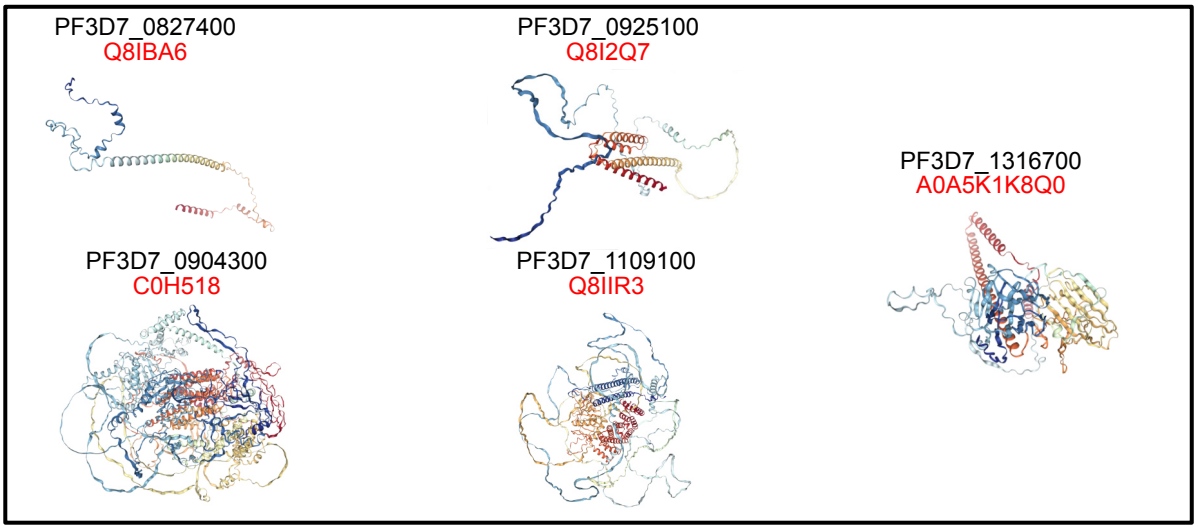
Supplementary Fig 5. Gene ontology enrichment by Slingshot clustering

A. Dot plot of a gene set enrichment analysis from the top differentially expressed genes over the pseudotime. Genes of interest were grouped according to the top 20 GO terms identified in the three pseudotime clusters and statistically analyzed by gene ratio with more significantly (red) and less significantly (yellow) expressed genes, color coded and organized by gene count (dot size). The X-axis indicates gene ratio, and the Y-axis indicates GO terms. **B**-**D**. Bar plots of top GO terms showing the over-represented GO terms for three (pt0, pt1, pt2) cell communities determined using Slingshot pseudotime curve reconstruction. Top 10 GO terms were selected per cluster.

Supplementary Fig 6

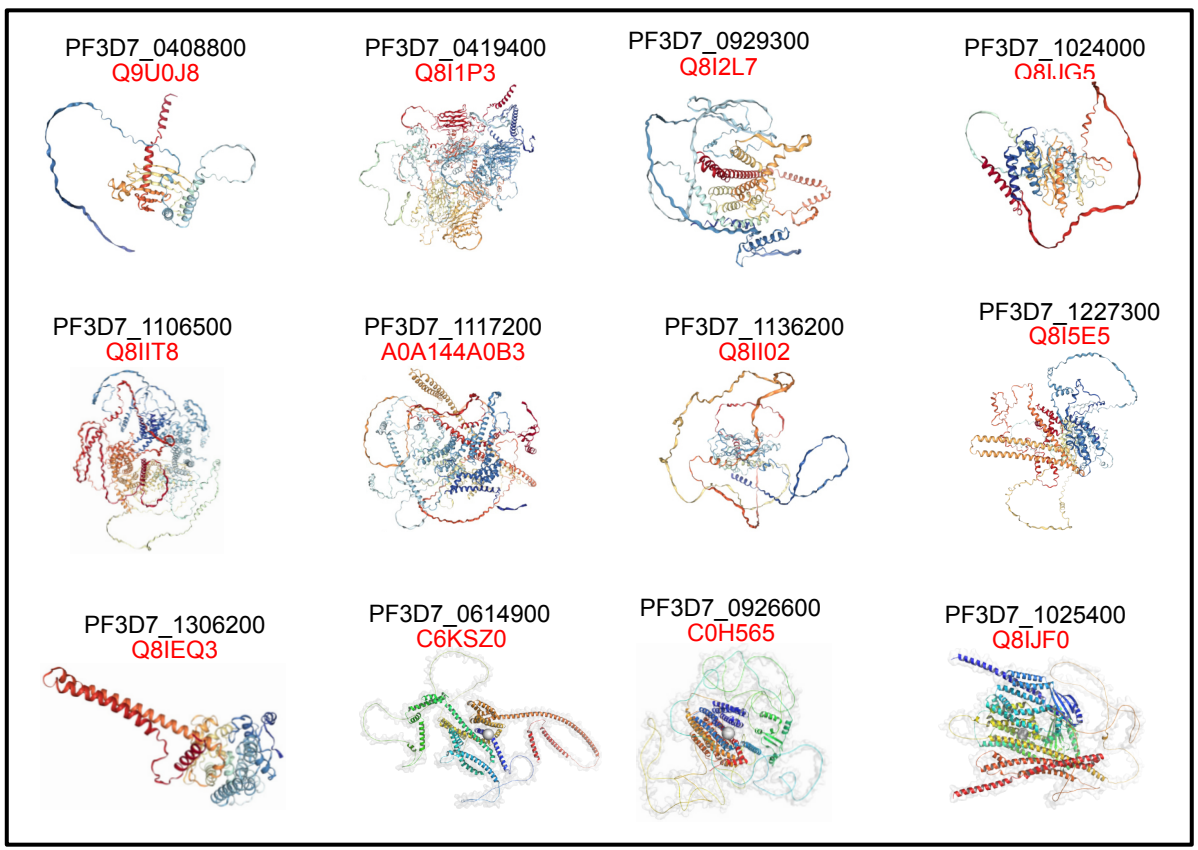
A

pt1 (table b)



B

pt2 (table c)



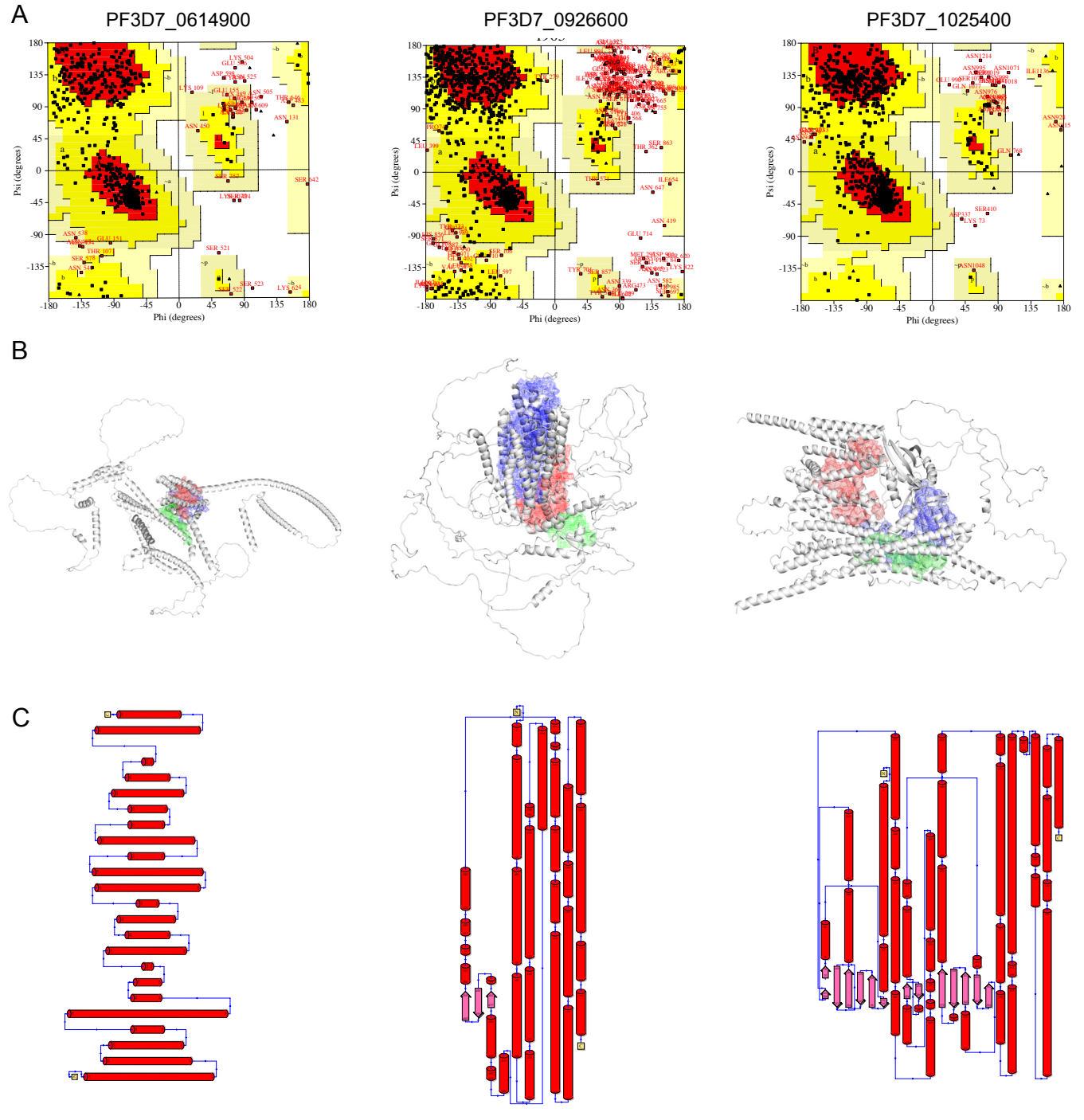
929 **Supplementary Fig 6. 3D models of proteins of the highly expressed genes from**
930 **pt1 and pt2.**

931 Visualization of 3D structure models of intrinsically disordered proteins showing the
932 cartoon topology. The molecular structures were retrieved from the Alpha-fold
933 database and subjected to structural validation using the CSpritz and molecular
934 dynamics simulation tools. **A**. Predicted 3D structures of the non-annotated genes from
935 the pt1 cluster, including their PlasmoDB accession number and UniProt code. **B**.
936 Predicted 3D structures of the non-annotated genes from the pt2 cluster, including their
937 PlasmoDB accession number and UniProt code.

938

939

Supplementary Fig 7



940 **Supplementary Fig 7. Protein structure and topology to predict the druggability**
941 **of three putative membrane proteins**

942 **A**. Ramachandran plot analysis for the three predicted proteins described as putative
943 membrane proteins in PlasmoDB. Red regions indicate the most favored amino acid
944 residues, yellow regions indicate allowed regions, light yellow generously allowed
945 regions and white indicating disallowed region. Phi and Psi scales indicate the
946 predicted torsion angles. **B**. Graphical structures and surface topography of the three
947 predicted protein models depicting the largest three binding pockets predicted by
948 DoGSitescorer and simulated using pyMol. The binding pockets are colored blue, red
949 and green. **C**. Topographies for the three predicted membrane protein models from the
950 N terminus to the C terminus, the amino acid residue positions are visualized within
951 the peptide chain.
952

References:

Akinosoglou, K. A., Bushell, E. S. C., Ukegbu, C. V., Schlegelmilch, T., Cho, J.-S., Redmond, S., Sala, K., Christophides, G. K., & Vlachou, D. (2015). Characterization of Plasmodium developmental transcriptomes in Anopheles gambiae midgut reveals novel regulators of malaria transmission. *Cellular Microbiology*, *17*(2), 254–268. https://doi.org/10.1111/cmi.12363

Bennett, B. J., Thompson, J., & Coppel, R. L. (1995). Identification of Plasmodium falciparum histone 2B and histone 3 genes. *Molecular and Biochemical Parasitology*, *70*(1–2), 231–233. https://doi.org/10.1016/0166-6851(95)00030-5

Bennink, S., Kiesow, M. J., & Pradel, G. (2016). The development of malaria parasites in the mosquito midgut. *Cellular Microbiology*, *18*(7), 905–918. https://doi.org/10.1111/cmi.12604

Billker, O., Shaw, M. K., Margos, G., & Sinden, R. E. (1997). The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of Plasmodium berghei in vitro. *Parasitology*, *115 (Pt 1)*, 1–7. https://doi.org/10.1017/s0031182097008895

Böhme, U., Otto, T. D., Sanders, M., Newbold, C. I., & Berriman, M. (2019). Progression of the canonical reference malaria parasite genome from 2002–2019. *Wellcome Open Research*, *4*, 58. https://doi.org/10.12688/wellcomeopenres.15194.2

Bushell, E. S. C., Ecker, A., Schlegelmilch, T., Goulding, D., Dougan, G., Sinden, R. E., Christophides, G. K., Kafatos, F. C., & Vlachou, D. (2009). Paternal Effect of the Nuclear Formin-like Protein MISFIT on Plasmodium Development in the Mosquito Vector. *PLOS Pathogens*, *5*(8), e1000539. https://doi.org/10.1371/journal.ppat.1000539

Claessens, A., Harris, L. M., Stanojcic, S., Chappell, L., Stanton, A., Kuk, N., Veneziano-Broccia, P., Sterkers, Y., Rayner, J. C., & Merrick, C. J. (2018). RecQ helicases in the malaria parasite Plasmodium falciparum affect genome stability, gene expression patterns and DNA replication dynamics. *PLOS Genetics*, *14*(7), e1007490. https://doi.org/10.1371/journal.pgen.1007490

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Dankwa, D. A., Davis, M. J., Kappe, S. H. I., & Vaughan, A. M. (2016). A Plasmodium yoelii Mei2-Like RNA Binding Protein Is Essential for Completion of Liver Stage Schizogony. *Infection and Immunity*, *84*(5), 1336–1345. https://doi.org/10.1128/IAI.01417-15

de Jong, R. M., Meerstein-Kessel, L., Da, D. F., Nsango, S., Challenger, J. D., van de Vegte-Bolmer, M., van Gemert, G.-J., Duarte, E., Teyssier, N., Sauerwein, R. W., Churcher, T. S., Dabire, R. K., Morlais, I., Locke, E., Huynen, M. A., Bousema, T., & Jore, M. M. (2021). Monoclonal antibodies block transmission of genetically diverse Plasmodium falciparum strains to mosquitoes. *Npj Vaccines*, *6*(1), 1–9. https://doi.org/10.1038/s41541-021-00366-9

Delves, M. J., Straschil, U., Ruecker, A., Miguel-Blanco, C., Marques, S., Dufour, A. C., Baum, J., & Sinden, R. E. (2016). Routine in vitro culture of P. falciparum gametocytes to evaluate novel transmission-blocking interventions. *Nature Protocols*, *11*(9), 1668–1680. https://doi.org/10.1038/nprot.2016.096

Dijk, M. R. van, Schaijk, B. C. L. van, Khan, S. M., Dooren, M. W. van, Ramesar, J., Kaczanowski, S., Gemert, G.-J. van, Kroeze, H., Stunnenberg, H. G., Eling, W. M., Sauerwein, R. W., Waters, A. P., & Janse, C. J. (2010). Three Members of the 6-cys Protein Family of Plasmodium Play a Role in Gamete Fertility. *PLOS Pathogens*, *6*(4), e1000853. https://doi.org/10.1371/journal.ppat.1000853

Dimopoulos, G., Müller, H.-M., Levashina, E. A., & Kafatos, F. C. (2001). Innate immune defense against malaria infection in the mosquito. *Current Opinion in Immunology*, *13*(1), 79–88. https://doi.org/10.1016/S0952-7915(00)00186-2

Duran-Bedolla, J., Téllez-Sosa, J., Valdovinos-Torres, H., Pavón, N., Buelna-Chontal, M., Tello-López, A. T., Argotte-Ramos, R., Rodríguez, M. H., & Rodríguez, M. C. (2017). Cellular stress associated with the differentiation of Plasmodium berghei ookinetes. *Biochemistry and Cell Biology*, *95*(2), 310–317. https://doi.org/10.1139/bcb-2016-0028

Ecker, A., Bushell, E. S. C., Tewari, R., & Sinden, R. E. (2008). Reverse genetics screen identifies six proteins important for malaria development in the mosquito. *Molecular Microbiology*, *70*(1), 209–220. https://doi.org/10.1111/j.1365-2958.2008.06407.x

Fan, Z., Zhang, L., Yan, G., Wu, Q., Gan, X., Zhong, S., & Lin, G. (2011). Bioinformatics analysis for structure and function of CPR of Plasmodium falciparum. *Asian Pacific Journal of Tropical Medicine*, *4*(2), 85–87. https://doi.org/10.1016/S1995-7645(11)60042-4

Graumans, W., Jacobs, E., Bousema, T., & Sinnis, P. (2020). When Is a Plasmodium-Infected Mosquito an Infectious Mosquito? *Trends in Parasitology*, *36*(8), 705–716. https://doi.org/10.1016/j.pt.2020.05.011

Griffin, J. T., Hollingsworth, T. D., Okell, L. C., Churcher, T. S., White, M., Hinsley, W., Bousema, T., Drakeley, C. J., Ferguson, N. M., Basáñez, M.-G., & Ghani, A. C. (2010). Reducing Plasmodium falciparum malaria transmission in Africa: A model-based evaluation of intervention strategies. *PLoS Medicine*, *7*(8), e1000324. https://doi.org/10.1371/journal.pmed.1000324

Guermeur, Y., Geourjon, C., Gallinari, P., & Deléage, G. (1999). Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics (Oxford, England)*, *15*(5), 413–421. https://doi.org/10.1093/bioinformatics/15.5.413

Guttery, D. S., Poulin, B., Ferguson, D. J. P., Szöőr, B., Wickstead, B., Carroll, P. L., Ramakrishnan, C., Brady, D., Patzewitz, E.-M., Straschil, U., Solyakov, L., Green, J. L., Sinden, R. E., Tobin, A. B., Holder, A. A., & Tewari, R. (2012). A Unique Protein Phosphatase with Kelch-Like Domains (PPKL) in Plasmodium Modulates Ookinete Differentiation, Motility and Invasion. *PLOS Pathogens*, *8*(9), e1002948. https://doi.org/10.1371/journal.ppat.1002948

Guy, A. J., Irani, V., MacRaild, C. A., Anders, R. F., Norton, R. S., Beeson, J. G., Richards, J. S., & Ramsland, P. A. (2015). Insights into the Immunological Properties of Intrinsically Disordered Malaria Proteins Using Proteome Scale Predictions. *PLOS ONE*, *10*(10), e0141729. https://doi.org/10.1371/journal.pone.0141729

Hirano, T. (2016). Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell*, *164*(5), 847–857. https://doi.org/10.1016/j.cell.2016.01.033

Howick, V. M., Russell, A. J. C., Andrews, T., Heaton, H., Reid, A. J., Natarajan, K., Butungi, H., Metcalf, T., Verzier, L. H., Rayner, J. C., Berriman, M., Herren, J. K., Billker, O., Hemberg, M., Talman, A. M., & Lawniczak, M. K. N. (2019). The

Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. *Science*. https://doi.org/10.1126/science.aaw2619

Janse, C. J., Van der Klooster, P. F., Van der Kaay, H. J., Van der Ploeg, M., & Overdulve, J. P. (1986). Rapid repeated DNA replication during microgametogenesis and DNA synthesis in young zygotes of Plasmodium berghei. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *80*(1), 154–157. https://doi.org/10.1016/0035-9203(86)90219-1

Jeninga, M. D., Quinn, J. E., & Petter, M. (2019). ApiAP2 Transcription Factors in Apicomplexan Parasites. *Pathogens*, *8*(2), 47. https://doi.org/10.3390/pathogens8020047

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kalitsis, P., Zhang, T., Marshall, K. M., Nielsen, C. F., & Hudson, D. F. (2017). Condensin, master organizer of the genome. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *25*(1), 61–76. https://doi.org/10.1007/s10577-017-9553-0

Kaneko, I., Iwanaga, S., Kato, T., Kobayashi, I., & Yuda, M. (2015). Genome-Wide Identification of the Target Genes of AP2-O, a Plasmodium AP2-Family Transcription Factor. *PLOS Pathogens*, *11*(5), e1004905. https://doi.org/10.1371/journal.ppat.1004905

Kirk, K., & Horner, H. A. (1995). In search of a selective inhibitor of the induced transport of small solutes in Plasmodium falciparum-infected erythrocytes: Effects of arylaminobenzoates. *Biochemical Journal*, *311*(3), 761–768. https://doi.org/10.1042/bj3110761

Kundu, P., Semesi, A., Jore, M. M., Morin, M. J., Price, V. L., Liang, A., Li, J., Miura, K., Sauerwein, R. W., King, C. R., & Julien, J.-P. (2018). Structural delineation of potent transmission-blocking epitope I on malaria antigen Pfs48/45. *Nature Communications*, *9*(1), 4458. https://doi.org/10.1038/s41467-018-06742-9

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, *8*(4), 477–486. https://doi.org/10.1007/BF00228148

Lee, A. H., Symington, L. S., & Fidock, D. A. (2014). DNA Repair Mechanisms and Their Biological Roles in the Malaria Parasite Plasmodium falciparum. *Microbiology and Molecular Biology Reviews : MMBR*, *78*(3), 469–486. https://doi.org/10.1128/MMBR.00059-13

Li, Z., Cui, H., Guan, J., Liu, C., Yang, Z., & Yuan, J. (2021). Plasmodium transcription repressor AP2-O3 regulates sex-specific identity of gene expression in female gametocytes. *EMBO Reports*, *22*(5), e51660. https://doi.org/10.15252/embr.202051660

López-Barragán, M. J., Lemieux, J., Quiñones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., & Su, X. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC Genomics*, *12*(1), 587. https://doi.org/10.1186/1471-2164-12-587

Madura, K., & Prakash, S. (1990). Transcript levels of the Saccharomyes cerevisiae DNA repair gene RAD23 increase in response to UV light and in meiosis but remain constant in the mitotic cell cycle. *Nucleic Acids Research*, *18*(16), 4737–4742. https://doi.org/10.1093/nar/18.16.4737

Modrzynska, K., Pfander, C., Chappell, L., Yu, L., Suarez, C., Dundas, K., Gomes, A. R., Goulding, D., Rayner, J. C., Choudhary, J., & Billker, O. (2017). A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the Plasmodium Life Cycle. *Cell Host & Microbe*, *21*(1), 11–22. https://doi.org/10.1016/j.chom.2016.12.003

Ngara, M., Palmkvist, M., Sagasser, S., Hjelmqvist, D., Björklund, Å. K., Wahlgren, M., Ankarklev, J., & Sandberg, R. (2018). Exploring parasite heterogeneity using single-cell RNA-seq reveals a gene signature among sexual stage Plasmodium falciparum parasites. *Experimental Cell Research*, *371*(1), 130–138. https://doi.org/10.1016/j.yexcr.2018.08.003

Oliveira, G. de A., Lieberman, J., & Barillas-Mury, C. (2012). Epithelial Nitration by a Peroxidase/NOX5 System Mediates Mosquito Antiplasmodial Immunity. *Science*, *335*(6070), 856–859. https://doi.org/10.1126/science.1209678

Ortiz-Bazán, M. Á., Gallo-Fernández, M., Saugar, I., Jiménez-Martín, A., Vázquez, M. V., & Tercero, J. A. (2014). Rad5 plays a major role in the cellular response to DNA damage during chromosome replication. *Cell Reports*, *9*(2), 460–468. https://doi.org/10.1016/j.celrep.2014.09.005

Padrón, A., Molina-Cruz, A., Quinones, M., Ribeiro, J. M., Ramphul, U., Rodrigues, J., Shen, K., Haile, A., Ramirez, J. L., & Barillas-Mury, C. (2014). In depth annotation of the Anopheles gambiae mosquito midgut transcriptome. *BMC Genomics*, *15*(1), 636. https://doi.org/10.1186/1471-2164-15-636

Pandey, R., Kumar, R., Gupta, P., Mohmmed, A., Tewari, R., Malhotra, P., & Gupta, D. (2018). High throughput in silico identification and characterization of Plasmodium falciparum PRL phosphatase inhibitors. *Journal of Biomolecular Structure and Dynamics*, *36*(13), 3531–3540. https://doi.org/10.1080/07391102.2017.1392365

Paton, D. G., Childs, L. M., Itoe, M. A., Holmdahl, I. E., Buckee, C. O., & Catteruccia, F. (2019). Exposing Anopheles mosquitoes to antimalarials blocks Plasmodium parasite transmission. *Nature*, *567*(7747), 239–243. https://doi.org/10.1038/s41586-019-0973-1

Peterson, T. M. L., & Luckhart, S. (2006). A mosquito 2-Cys peroxiredoxin protects against nitrosative and oxidative stresses associated with malaria parasite infection. *Free Radical Biology & Medicine*, *40*(6), 1067–1082. https://doi.org/10.1016/j.freeradbiomed.2005.10.059

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, *9*(1), 171–181. https://doi.org/10.1038/nprot.2014.006

Poran, A., Nötzel, C., Aly, O., Mencia-Trinchant, N., Harris, C. T., Guzman, M. L., Hassane, D. C., Elemento, O., & Kafsack, B. F. C. (2017). Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature*, *551*(7678), 95–99. https://doi.org/10.1038/nature24280

Radke, J. B., Worth, D., Hong, D., Huang, S., Jr, W. J. S., Wilson, E. H., & White, M. W. (2018). Transcriptional repression by ApiAP2 factors is central to chronic toxoplasmosis. *PLOS Pathogens*, *14*(5), e1007035. https://doi.org/10.1371/journal.ppat.1007035

Ralph, S. A., & Cortés, A. (2019). Plasmodium sexual differentiation: How to make a female. *Molecular Microbiology*, *112*(6), 1627–1631. https://doi.org/10.1111/mmi.14340

Rawat, M., Srivastava, A., Johri, S., Gupta, I., & Karmodiya, K. (2021). Single-Cell RNA Sequencing Reveals Cellular Heterogeneity and Stage Transition under Temperature Stress in Synchronized Plasmodium falciparum Cells. *Microbiology Spectrum*, *9*(1), e00008-21. https://doi.org/10.1128/Spectrum.00008-21

Real, E., Howick, V. M., Dahalan, F. A., Witmer, K., Cudini, J., Andradi-Brown, C., Blight, J., Davidson, M. S., Dogga, S. K., Reid, A. J., Baum, J., & Lawniczak, M. K. N. (2021). A single-cell atlas of Plasmodium falciparum transmission through the mosquito. *Nature Communications*, *12*(1), 3196. https://doi.org/10.1038/s41467-021-23434-z

Reid, A. J., Talman, A. M., Bennett, H. M., Gomes, A. R., Sanders, M. J., Illingworth, C. J. R., Billker, O., Berriman, M., & Lawniczak, M. K. (2018). Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *ELife*, *7*, e33105. https://doi.org/10.7554/eLife.33105

Richard, D., Bartfai, R., Volz, J., Ralph, S. A., Muller, S., Stunnenberg, H. G., & Cowman, A. F. (2011). A Genome-wide Chromatin-associated Nuclear Peroxiredoxin from the Malaria Parasite Plasmodium falciparum*. *Journal of Biological Chemistry*, *286*(13), 11746–11755. https://doi.org/10.1074/jbc.M110.198499

Sanders, P. R., Gilson, P. R., Cantin, G. T., Greenbaum, D. C., Nebl, T., Carucci, D. J., McConville, M. J., Schofield, L., Hodder, A. N., Yates, J. R., & Crabb, B. S. (2005). Distinct Protein Classes Including Novel Merozoite Surface Antigens in Raft-like Membranes of Plasmodium falciparum*. *Journal of Biological Chemistry*, *280*(48), 40169–40176. https://doi.org/10.1074/jbc.M509631200

Sauerwein, R. W., & Bousema, T. (2015). Transmission blocking malaria vaccines: Assays and candidates in clinical development. *Vaccine*, *33*(52), 7476–7482. https://doi.org/10.1016/j.vaccine.2015.08.073

Schureck, M. A., Darling, J. E., Merk, A., Shao, J., Daggupati, G., Srinivasan, P., Olinares, P. D. B., Rout, M. P., Chait, B. T., Wollenberg, K., Subramaniam, S., & Desai, S. A. (2021). Malaria parasites use a soluble RhopH complex for erythrocyte invasion and an integral form for nutrient uptake. *ELife*, *10*, e65282. https://doi.org/10.7554/eLife.65282

Sierra-Miranda, M., Vembar, S.-S., Delgadillo, D. M., Ávila-López, P. A., Herrera-Solorio, A.-M., Lozano Amado, D., Vargas, M., & Hernandez-Rivas, R. (2017). PfAP2Tel, harbouring a non-canonical DNA-binding AP2 domain, binds to Plasmodium falciparum telomeres. *Cellular Microbiology*, *19*(9), e12742. https://doi.org/10.1111/cmi.12742

Smith, R. C., Vega-Rodríguez, J., & Jacobs-Lorena, M. (2014). The Plasmodium bottleneck: Malaria parasite losses in the mosquito vector. *Memorias Do Instituto Oswaldo Cruz*, *109*(5), 644–661.

Srinivasan, P., Fujioka, H., & Jacobs-Lorena, M. (2008). PbCap380, a novel oocyst capsule protein, is essential for malaria parasite survival in the mosquito. *Cellular Microbiology*, *10*(6), 1304–1312. https://doi.org/10.1111/j.1462-5822.2008.01127.x

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell

transcriptomics. *BMC Genomics*, *19*(1), 477. https://doi.org/10.1186/s12864-018-4772-0

Turturice, B. A., Lamm, M. A., Tasch, J. J., Zalewski, A., Kooistra, R., Schroeter, E. H., Sharma, S., Kawazu, S.-I., & Kanzok, S. M. (2013). Expression of cytosolic peroxiredoxins in Plasmodium berghei ookinetes is regulated by environmental factors in the mosquito bloodmeal. *PLoS Pathogens*, *9*(1), e1003136. https://doi.org/10.1371/journal.ppat.1003136

Ukegbu, C. V., Giorgalli, M., Tapanelli, S., Rona, L. D. P., Jaye, A., Wyer, C., Angrisano, F., Blagborough, A. M., Christophides, G. K., & Vlachou, D. (2020). PIMMS43 is required for malaria parasite immune evasion and sporogonic development in the mosquito vector. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(13), 7363–7373. https://doi.org/10.1073/pnas.1919709117

Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications*, *11*(1), 1201. https://doi.org/10.1038/s41467-020-14766-3

van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sánchez de Groot, N., Huynen, M. A., Matouschek, A., Fuxreiter, M., & Babu, M. M. (2014). Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution. *Cell Reports*, *8*(6), 1832–1844. https://doi.org/10.1016/j.celrep.2014.07.055

Vaughan, A. M., Pinapati, R. S., Cheeseman, I. H., Camargo, N., Fishbaugher, M., Checkley, L. A., Nair, S., Hutyra, C. A., Nosten, F. H., Anderson, T. J. C., Ferdig, M. T., & Kappe, S. H. I. (2015). Plasmodium falciparum genetic crosses in a humanized mouse model. *Nature Methods*, *12*(7), 631–633. https://doi.org/10.1038/nmeth.3432

Walsh, I., Martin, A. J. M., Di Domenico, T., Vullo, A., Pollastri, G., & Tosatto, S. C. E. (2011). CSpritz: Accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs. *Nucleic Acids Research*, *39*(Web Server issue), W190-196. https://doi.org/10.1093/nar/gkr411

World malaria report 2021. (n.d.). Retrieved March 3, 2022, from https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021

Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, *101*(10), 2525–2534. https://doi.org/10.1016/j.bpj.2011.10.024

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1), W174–W181. https://doi.org/10.1093/nar/gkv342

Zhang, C., Li, Z., Cui, H., Jiang, Y., Yang, Z., Wang, X., Gao, H., Liu, C., Zhang, S., Su, X.-Z., & Yuan, J. (2017). Systematic CRISPR-Cas9-Mediated Modifications of Plasmodium yoelii ApiAP2 Genes Reveal Functional Insights into Parasite Development. *MBio*, *8*(6), e01986-17. https://doi.org/10.1128/mBio.01986-17