

Regulators of male and female sexual development critical for transmission of a malaria parasite

Andrew J C Russell ^{1†}, Theo Sanderson ^{1, 2†}, Ellen Bushell ^{1, 3, 4†}, Arthur M Talman ^{1, 5†}, Burcu Anar ¹, Gareth Girling ¹, Mirjam Hunziker ^{3, 4}, Robyn S Kent ^{6, ‡}, Tom Metcalf ¹, Ruddy Montandon ^{1, §}, Vikash Pandey ^{3, 4}, A Brett Roberts ⁶, Claire Sayers ^{3, 4}, Frank Schwach ¹, Julian C Rayner ⁷, Thierry Voet ^{1, 8, 9}, Katarzyna K Modrzynska ⁶, Andrew P. Waters ^{6*}, Mara K N Lawniczak ^{1*}, Oliver Billker ^{3, 4 *}

[†] These authors contributed equally to this work

[‡] Present address: University of Vermont, Burlington VT 05405, USA

[§] Present address: Wellcome Centre for Human Genetics, Roosevelt Dr, Headington, Oxford OX3 7BN, United Kingdom.

*Corresponding authors. For bulk transcriptomic time course and validation, email: andy.waters@glasgow.ac.uk; single cell analysis and validation, email: mara@sanger.ac.uk; functional screen and validation, email: oliver.billker@umu.se.

Affiliations:

¹Wellcome Sanger Institute, Hinxton CB10 1SA, United Kingdom.

²Francis Crick Institute, 1 Midland Road, London NW1 1AT, United Kingdom.

³Laboratory for Molecular Infection Medicine Sweden, Umeå University, Umeå S-90187, Sweden.

⁴Department of Molecular Biology, Umeå University, Umeå S-90187, Sweden,

⁵MIVEGEC, University of Montpellier, IRD, CNRS, Montpellier, France.

⁶Wellcome Centre for Integrative Parasitology, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, G12 8TA, United Kingdom.

⁷Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, United Kingdom.

⁸Department of Human Genetics, University of Leuven, KU Leuven, Leuven, Belgium

⁹KU Leuven Institute for Single Cell Omics, LISCO, KU Leuven, Leuven, Belgium

Summary

The transmission of malaria parasites from vertebrate host to mosquito vector requires a developmental switch in asexually dividing blood-stage parasites to sexual reproduction. In *Plasmodium berghei* the transcription factor AP2-G is required and sufficient for this switch, but how a particular sex is determined in a haploid parasite remains unknown. Using a global screen of barcoded mutants, we here identify ten genes essential for the formation of either male or female sexual forms and validate their importance for transmission. High-resolution single-cell transcriptomics of wild-type and mutant parasites portrays the developmental bifurcation and reveals a regulatory cascade of putative gene functions in determination and subsequent differentiation of each sex. A male-determining gene with a LOTUS/OST-HTH domain points towards unexpected conservation of molecular mechanisms of gametogenesis in animals and a distantly related eukaryotic parasite.

Introduction

The transmission of malaria parasites to their mosquito vectors requires that a subset of blood-stage parasites switch from repeated asexual replication to sexual development. Epigenetically controlled expression of the transcription factor AP2-G is essential for asexual parasites to commit to sexual development (1–4), but the events that regulate the subsequent differentiation of sexually committed parasites into either sex are not understood. Although induced over-expression of AP2-G is sufficient to reprogram asexual parasites for sexual development experimentally (5, 6), it remains unknown how this single transcription factor generates two diverging gene expression programs that rapidly lead to male and female gametocytes.

Gametocyte sex ratios differ both between infected hosts and during the course of individual infections in ways that affect transmission to the vector and thereby the epidemiology of malaria (7–10). Gametocyte sex determination cannot involve sex chromosomes or inherited mating type loci, because asexual *Plasmodium* blood stages are haploid, and because competent clones retain the ability to produce both male and female gametocytes. Together, these observations suggest an epigenetic, i.e. non-chromosomal, mechanism for the emergence and differentiation of different sexes that is responsive to environmental regulation. Although sex is thought to have evolved only once in the ancestral eukaryote (11), mechanisms for how different sexes are determined evolve rapidly (12). As a result, none of the genes involved in equivalent determinations in other eukaryotes (12) have clear homologues in the Apicomplexa, the phylum of divergent eukaryotes to which malaria parasites belong. The genes involved in sex determination in *Plasmodium* are therefore likely to require employment of a genome-wide screen to be discovered.

In the rodent parasite *P. berghei*, pools of barcoded mutants can now be screened to discover gene functions in an unbiased manner (13–15), and blood-stage data suggest that phenotypes observed in *P. berghei* are largely predictive of those in *P. falciparum* (13, 16). Here, we have used this approach to screen for genes required for the formation of male and female gametocytes. We then used bulk and single-cell transcriptomics to map the differentiation of these lineages at high temporal resolution. We find that parasites express markers indicative of their eventual sex early in the developmental bifurcation and, by disrupting these genes and characterising mutants, we identify essential components of the male and female transcriptional programs.

Results

A systematic screen identifies sexual development genes.

To screen for genes required for sexual development, we mutagenized the *P. berghei* reporter line 820 (17), which expresses green and red fluorescent proteins (GFP, RFP) from promoters specific for male and female gametocytes, respectively (Fig. 1A). This line was transfected with pools of barcoded *PlasmoGEM* vectors (18) targeting 1,302 genes previously determined to be non-essential for asexual erythrocytic growth (relative growth rate at least half of wild type according to (13)). Using expression of fluorescent proteins as proxies for sexual development, parasitised red blood cells from mice infected with each superpool were sorted into male (GFP-positive), female (RFP-positive) and asexual (Hoechst-only) populations ($\sim 10^6$ of each; Fig. 1B).

Barcodes were then counted by deep sequencing of amplicons obtained from the genomic DNA of the sorted parasites. After six replicate screens in large pools, 50 top hits were re-screened in duplicate and with a similar number of control mutants. This smaller pool provided more precise measurements of reporter expression (see Table S1 for all screen data).

96% of parasite genes that could be queried were not required for the expression of sexual reporter proteins because their respective barcodes were equally represented in the sorted populations (Fig. 1B). At the defined significance threshold, 30 mutants were depleted from both sexual populations, another 14 were reduced only in the male and 21 only from the female population (Fig. 1C, Table S1). Reassuringly, hits included the transcriptional activator *ap2-g* (4, 6) and the repressor *ap2-g2* (4, 19), both known to regulate gene expression during gametocyte formation in *P. berghei*. Other genes affecting both sexes often had weaker effects and were also required for normal asexual growth ((13); Table S1), suggesting they contribute to cell survival more broadly, examples including a putative NAD synthase (PBANKA_0827500) and a pre-mRNA splicing factor (PBANKA_0409100). Only a few mutants resembled *ap2-g* in affecting both sexual markers profoundly, while not reducing asexual growth. Most notable in this category were a putative ubiquitin-conjugating enzyme (PBANKA_0806000) and a conserved *Plasmodium* protein of unknown function (PBANKA_0824300). The notion that both sexes require these genes for fertility is consistent with evidence from an earlier screen showing that neither the female nor the male gametocyte can pass either of these disrupted alleles to the oocyst stage, which establishes the infection in the mosquito (>24-fold reduction in oocyst numbers shown by (15); Table S1).

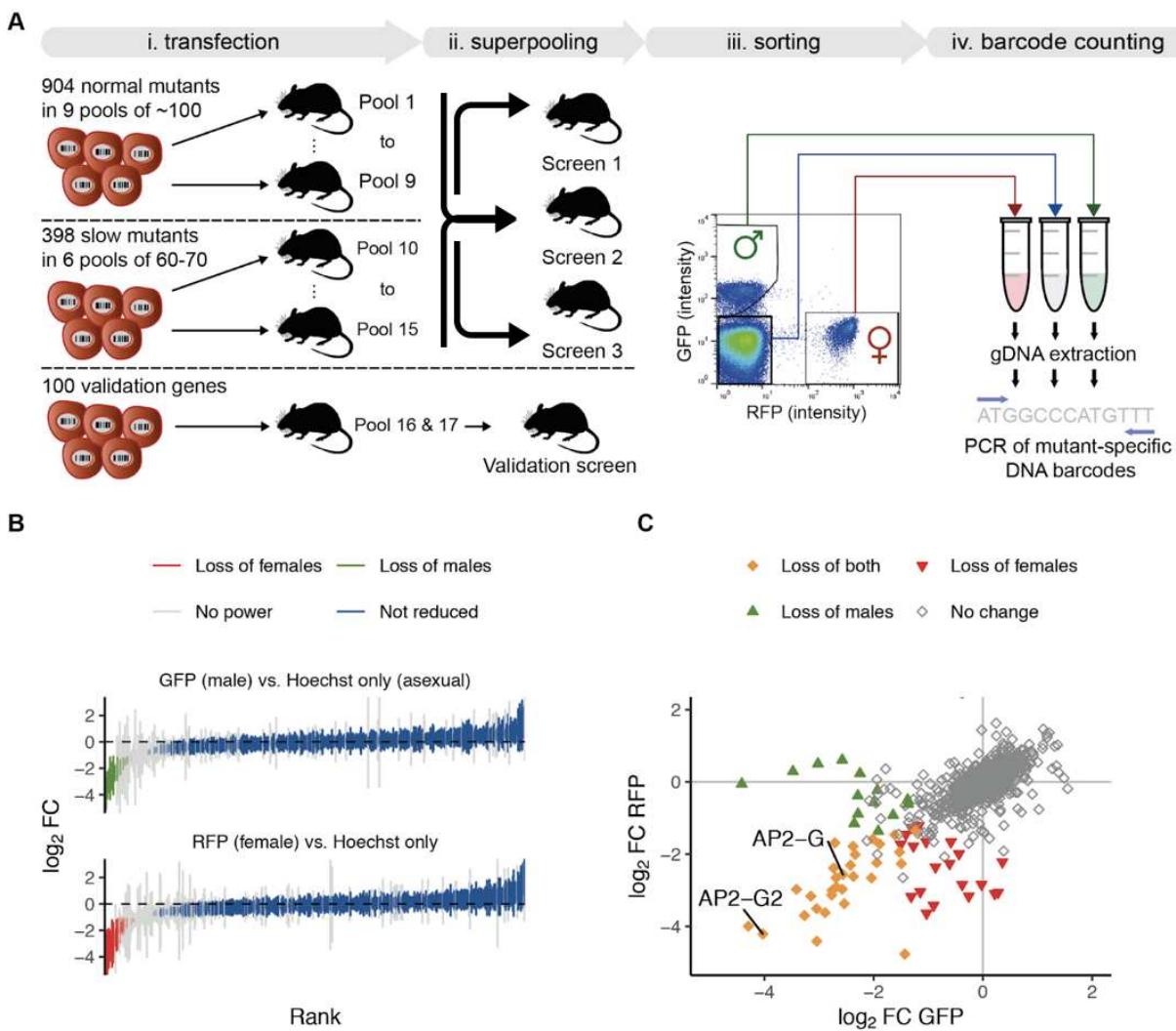


Fig. 1. A barseq screen for sexual reporter expression in *P. berghei*. (A) Schematic overview showing how *PlasmoGEM* vectors were allocated to pools for transfection and the resulting mutants combined into superpools for sorting on reporter expression. Duplicate barcode PCRs were performed for each sorted population and converted into sequencing libraries for barcode counting. (B) All robustly quantified mutants are ranked by the degree to which cells expressing either the male or the female reporter gene were underrepresented. Error bars show standard deviations from at least 4 independent screening experiments. (C) Combined results from both reporters, showing filled symbols where the underrepresentation was significant for either one or both sexes. Two highlighted mutants in *ApiAP2* genes confirm the expected loss of both markers from the population, as published (4).

Genes with functions specific to a single sex were highly represented among a small group of 60 genes that we previously showed respond within six hours of experimentally reprogramming ring stages to sexual development by induced expression of AP2-G (6). Since early response genes may hold a clue to sex determination, we wanted to increase the temporal resolution of the time course experiment. *In vitro* synchronised schizonts were reprogrammed into gametocytes, injected into mice and harvested for bulk RNA-seq analysis at additional time points during the first 6 hours after induction (Fig. S1, Table S2). Co-expression analysis by neural network-based dimensionality reduction now identified an even smaller cluster of 12 co-regulated genes that responded to *ap2-g* induction within 1-2 h and plateaued from 8-12 h (Fig. 2, cluster 32, and Table S2). This group contained no previously known sex markers, and their transcripts increased before the main wave of around 300 canonical male and female specific

genes, whose expression only began to increase detectably from 8-12 h after induction (Fig. 2A). Single sex screen hits accounted for five of the early response genes, a significant enrichment ($p < 10^{-8}$). Cluster 32 included several putative nucleic acid binding proteins of unknown function (Fig. 2B), which we hypothesized could be involved either in determining the sex of gametocytes or in their subsequent sex-specific differentiation. To examine this idea further, we selected all five screen hits from this cluster for further validation. We added to the validation group (shown in Fig. 2C), three genes from other clusters, which also responded rapidly to *ap2-g* overexpression. Cluster 32 included two additional genes encoding putative nucleic acid binding proteins, PBANKA_1302700 and PBANKA_1454800, which had, however, not been covered by the screen because they lacked barcoded *PlasmoGEM* vectors. These genes were selected to complement the hits from the unbiased screen because their domain architecture and expression pattern suggested they may be functionally related. Genes in the validation set are not required by asexual blood stages according to available growth data (13).

Flow cytometry with individual knockouts in the 820 line confirmed the biased expression of fluorescent sex reporters for all screen hits and further showed sex-specific losses of marker expression for the two newly included early response genes (Fig. 2D, Fig. S2). Depending on the affected sex, we refer to the validated genes as “male development” (*md1* to *md5*) or “female development” (*fd1* to *fd4*, Fig. 2B). As expected, mutants showed a complete or nearly complete loss in their ability to form oocysts in mosquitoes, with the exception of *md3*, in which oocysts were merely reduced (Fig. 2E). Genetic crosses using either individual mutants (Fig. 2F) or barcoded single-sex pools (Fig. 2G) showed that for most genes, fertility was sex-specifically affected precisely as predicted by reporter expression. Two notable deviations from the screen results were observed. A cloned mutant in PBANKA_0828000, which by FACS only lacked parasites expressing the female marker, additionally suffered from male infertility (Fig. 2F); due to its broader gametocyte development (*gd*) phenotype we refer to this gene as *gd1*. The second mutant requiring further consideration is *md3*, for which both a transmission experiment (Fig. 2E) and a cross with a cloned line (Fig. 2F) showed that fertile male gametocytes were still present. However, consistent with the screen result, male fertility of all male mutants, including *md3*, was reduced to <1% when measured in competition with fertile mutants (Fig. 2G). We hypothesise that the *md3* mutant produces fertile microgametes at a level that is much reduced, but still sufficient to give rise to a relative abundance of oocysts, probably because optimised laboratory infections of *P. berghei* produce an excess of zygotes, so that oocyst numbers begin to saturate at low input levels (20).

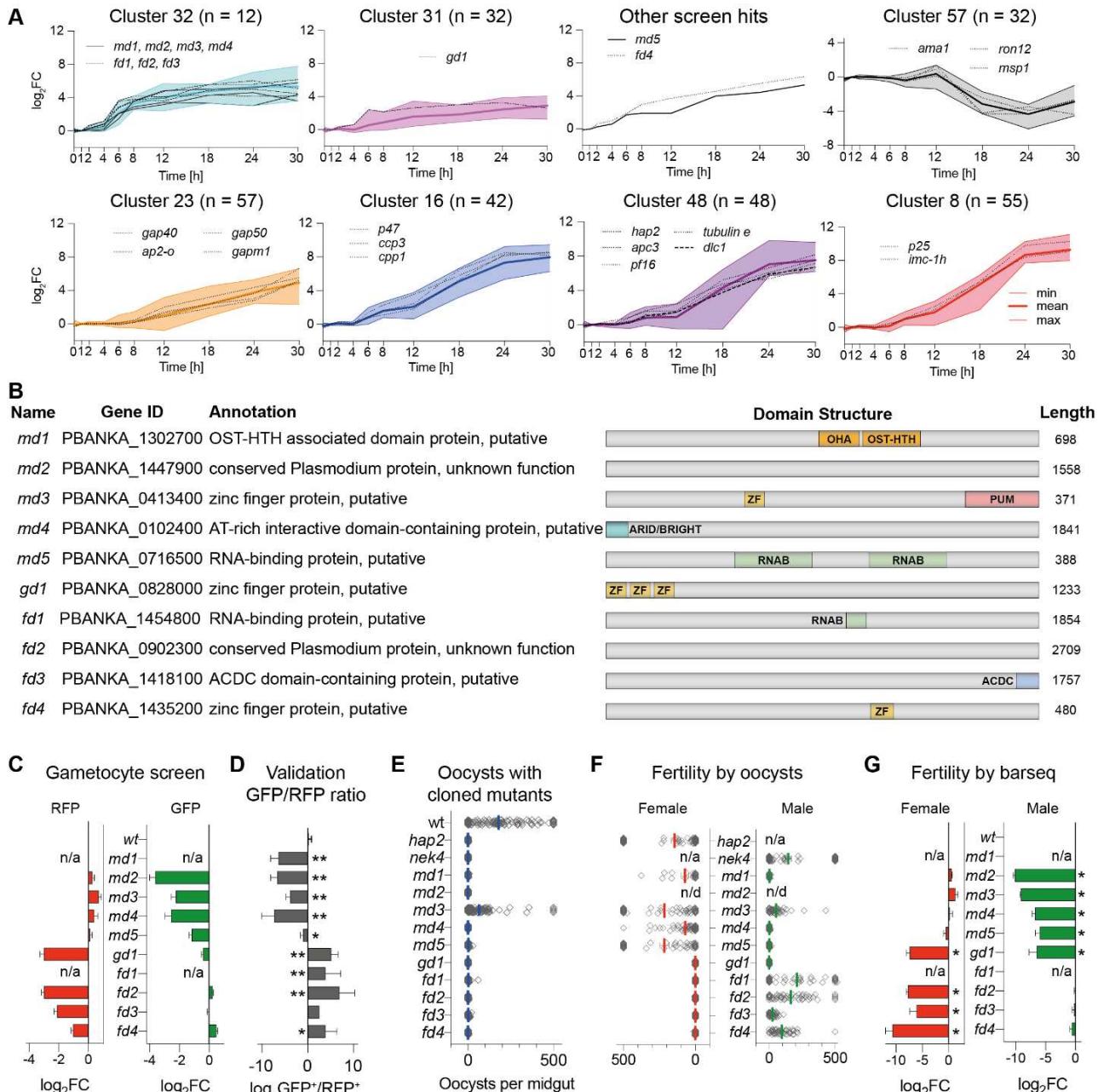


Fig. 2. Selection and validation of ten *P. berghei* genes with sex-specific roles in gametocytogenesis. (A) Selected gene expression clusters from a bulk RNA-seq timecourse of induced sexual development. Ring stage parasites were reprogrammed at $t = 0$ h by inducing *ap2-g* (6). Relative transcript abundances are given as log2-fold change relative to uninduced, asexually developing parasites. Shown are selected clusters [number of genes] with screen hits designated *md*, *fd* or *gd*. Selected well-characterised marker genes of male, female and asexual development are also shown. (B) Schematic illustration of genes with validated roles in sexual development. OST-HTH = oskar-TDRD5/TDRD7 winged helix-turn-helix domain; OHA = OST-HTH associated domain; ARID/BRIGHT = AT-rich interaction domain; ZN = C3H1 zinc finger; PUM = Pumilio RNA-binding repeat profile; RNAB = RNA-binding domain; ACDC = apetala 2 domain-coincident C-terminal domain; PH-like = PH domain like. (C) Change in reporter-positive cells in the barseq screen. Error bars show standard deviations. $p < 10^{-10}$ for the affected sex. (D) Sex ratio in individual mutants determined by flow cytometry. Error bars show standard deviations from 2-4 biological replicates with cloned mutants, except for *fd3*, where the uncloned population is shown. * $p < 0.05$; ** $p < 0.01$ in unpaired T test. (E) Transmission efficiency of mutant clones in vivo determined by counting oocysts on midguts 10 days after an infectious blood meal. (F) Male and female fertility as determined by the ability of mutant clones to give rise to oocysts in mosquitoes when crossed to *nek4* and *hap2* mutants, which provide fertile male or female gametes, respectively. Oocysts counts show combined data from 25-80 dissected mosquitoes from 2-3 independent experiments. n/d = not done. n/a = not applicable. (G) Change in female and male fertility determined by

barseq of infected midguts following mutagenesis of female-only or male-only lines, respectively. Error bars show standard deviations from four biological replicates. * $p < 0.001$.

Males and females differentiate from a shared sexual branch

To characterize the developmental block in each mutant more precisely, we performed a series of single-cell RNA sequencing (scRNA-seq) experiments. Using the plate-based Smart-seq2 method, we generated single cell transcriptomes from 2028 red blood cells infected with mutant parasites, and 689 wild-type controls for comparison. These parasites had all been cultured for 24 h to allow any atypically developing gametocytes to survive without being cleared by the spleen (Tables S1 and S3, Fig. S3A and B). We combined this Smart-seq2 data with a high-resolution map of gametocyte development created from 6191 droplet-based single cell transcriptomes covering the asexual cycle, sexual commitment, and the bifurcation into either sex. The two datasets were integrated to generate a combined uniform manifold approximation and projection (UMAP) plot (Fig. 3A and S4). Branching and pseudotime analysis on the combined data showed male and female gametocytes initially follow a common transcriptional trajectory after branching from the asexual cycle, before they assume distinct sexual identities (Fig. 3A, S5 and S6).

We clustered the 10x transcriptomes to resolve the branch points of sexual development (Fig. 3B). Transcripts from most sexual development genes first became detectable at the joint root of both sexes (Fig. 3C, Table S5 & Fig. S6B, bipotential cluster), where *ap2-g* was also upregulated, but before transcripts of canonical sex genes such as a dynein heavy chain (*mg1*, male) and *ccp2* (female) became detectable. *md* and *fd* genes were generally upregulated most strongly along the specific sexual trajectory affected by their disruption and thus serve as early markers downstream of *ap2-g* (Fig. 3B). Having assigned cells to male, female and asexual lineages in pseudotime (Fig. 3C), we identified co-expression gene modules in the single-cell data that delineate the sex-specific developmental programs (Fig. 3D). Mapping all knockout phenotypes from the screen onto these gene modules shows a significant enrichment for gametocytogenesis phenotypes among the first wave of sex specific genes in each branch ($p < 10^{-3}$), (Fig. 3D, Table S4), providing further validation for the screen and independent confirmation of the early response genes first identified in the bulk transcriptomes from the reprogramming time course (Fig. 2A and S1).

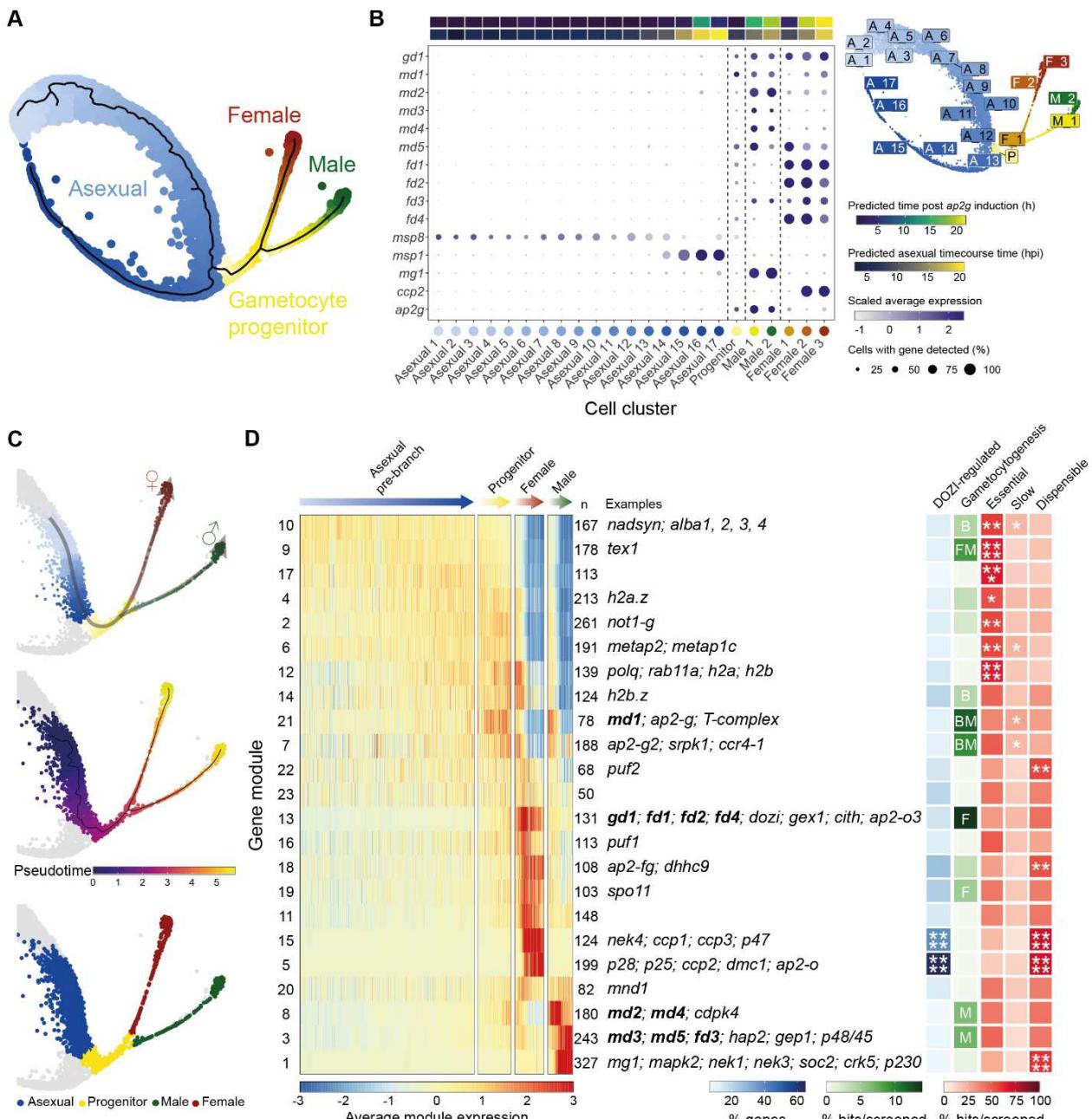


Fig. 3. Combined analysis of wild-type 10x and Smart-seq2 transcriptomes from *P. berghei* infected mouse red blood cells. (A) UMAP plot of all 6880 wild-type cells (Smart-seq2 and 10x), coloured by their assigned sex designation and progression along development. (B) Dot plot showing expression of known marker genes alongside the candidate genes in 23 cell clusters. The estimated average time point of each cluster is annotated by correlating single-cell data to bulk time course data (this study and (21)). Cluster locations are shown on the UMAP plot (top right). (C) Wild-type cells were sub-clustered and 10x-only cells were selected in order to re-calculate pseudotime for the branches of interest and construct modules of genes that are co-expressed over pseudotime. (D) Heatmap showing the scaled average expression of gene modules in cells shown in C. n = the number of genes per module. DOZI-regulated = % of DOZI-regulated genes within each cluster according to (22). B, F, and M, represent significant ($P \leq 0.05$) enrichment of both, female, or male hits in the screen, respectively. Essential, Slow, Dispensable = % genes per cluster with asexual blood stage phenotype according to (13). Examples include previously characterised *Plasmodium* genes with: female-specific expression (*ap2-o3* (23), *ap2-fg* (24), *nek4* (25, 26), *dhhc9* (27), *ccp2* (25), *ccp1* (28, 29), *ccp3* (29, 30), *p25* (31–33), *p28* (32)); functions in meiosis (*spo11* (34), *spo11-2* (34), *dmc1* (35), *gex1* (36)); male-specific expression (*cdpk4* (37), *hap2* (38), *mg1* (26), *soc2* (39), *crk5* (39)); female-biased RNA-binding proteins (*puf1* (40), *puf2* (41), *dozi* (42), *cith* (17)); All 8 T-complex proteins, except *cct4* (30). * = $P \leq 0.05$; ** = $P \leq 0.01$; * = $P \leq 0.001$; **** = $P \leq 0.0001$.**

Single cell transcriptomes distinguish differentiation genes from putative regulators of sex ratio determination.

Using the wild-type data to stage parasites (Fig. S7), we found that mutants fell into two classes: those where cells of the infertile sex were undetectable and those where the infertile sex was still present but had an atypical transcriptional signature (Fig. 4A, B). The latter category of mutants still expressed some of the core marker genes of the infertile sex but cells often clustered separately from wild type (Fig. S8, S9 and Table S6), and more cells resembled earlier points in pseudotime (Fig. 4B). This analysis identified *md4*, *md5*, *fd2*, *fd3* and *fd4* as differentiation mutants, because they become committed to a sex but fail to develop its complete transcriptional signature. Whilst the loss of *gd1* results in near complete absence of females, it also perturbs differentiation in males (Fig. 4A and B). Differential gene expression analysis shows differentiation mutants to be defined by unique transcriptional states in the infertile sex (Fig. 4C, S9 and Table S6), suggesting each gene affects transcription through a unique mechanism.

When the sex ratios of mutants were re-assessed using Smart-seq2 data for staging (Fig. 4B), four candidate genes for sex ratio determination emerged. *md1* and *md2* are potential maleness-determining genes since their disruption leads to a complete loss of cells expressing male markers. *md3* also showed a marked sex ratio shift towards females. However, consistent with the reduced male fertility retained by this mutant (Fig. 2E and F), we found enough *md3* males for enrichment by flow sorting. The transcriptomes of these cells were essentially indistinguishable from wild type (Fig. S10), suggesting deletion of *md3* does not affect male differentiation, but sex ratio, albeit with less complete penetrance than *md1* and *md2*. *gd1* behaved like a sex determining gene with respect to the female pathway, because cells with a female transcriptional signature were absent; the inability of a disrupted *gd1* locus to transmit to oocysts (Fig. 2E) may thus involve different mechanisms in each sex.

md1 is the most noticeable AP2-G responsive gene to be upregulated at the base of the sexual branch (Fig. 3C, D). It encodes a protein with a putative LOTUS/OST-HTH domain (43, 44). OST-HTH domain proteins exist in pro- and eukaryotes, but only those of animals have been studied in detail and all are involved in gametogenesis (45). Examples include Oskar, an important germline determining factor in *Drosophila* embryos (46) and the tudor domain contain proteins (TDRD5/TDRD7) with sex-specific fertility functions in mice (47). LOTUS domains organise ribonucleoprotein (RNP) complexes and it may therefore be significant that the zinc finger gene *md3* contains a weak homology with a pumilio RNA-binding domain, while *md2* lacks clear homologs outside of *Plasmodium*. Together, these considerations raise the intriguing possibility that creating the male lineage in *P. berghei* shares elements with germline definition in multicellular organisms.

Male differentiation mutants are *gd1*, *md4* and *md5* in order of decreasing severity with respect to the presence of core male transcripts (Fig. S9). Disrupting *gd1* and *md4* has profound but distinct effects on male gene expression. *md4* encodes a conserved *Plasmodium* protein of unknown function characterized by a putative N-terminal ARID/BRIGHT DNA binding domain, which in other eukaryotes targets developmental transcription factors to AT-rich DNA sequences (48), suggesting it may regulate transcription downstream of the initial commitment

to the male developmental trajectory. Transcript abundance of *md4* decreases in mature males but it stays high in *gd1* mutants (Fig. S11), illustrating the early developmental arrest of *gd1* males and also indicating that *gd1* is not required for the expression of *md4*. Deletion of *md5* has a less severe effect on core male transcripts and likely operates through a different mechanism because the protein contains putative RNA binding motifs.

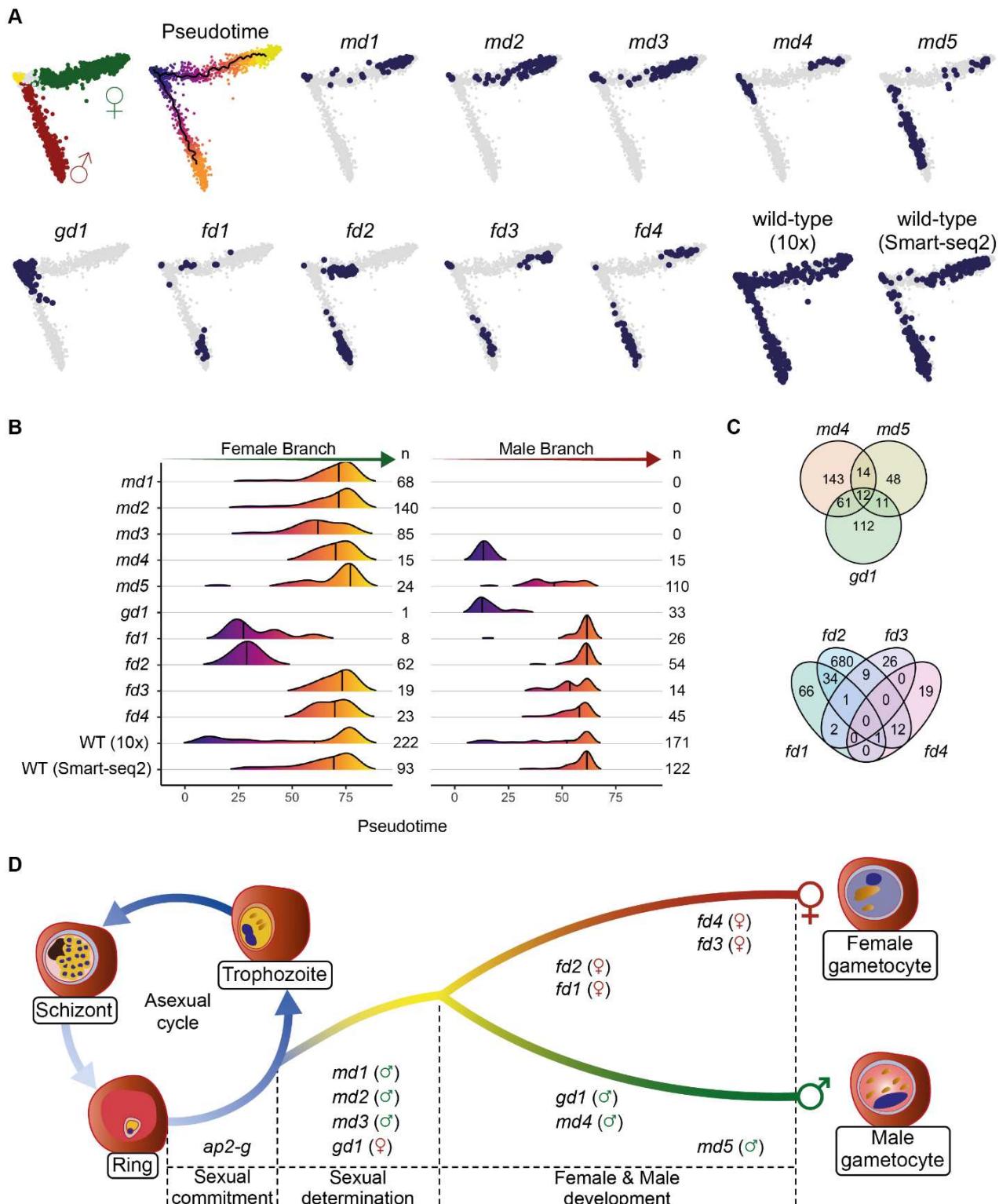


Fig. 4. Smart-seq2 analysis of *in vitro*-matured wild-type and mutant parasites. (A) Principal Component Analysis (PCA) plots of 3012 transcriptomes from single parasitised RBCs of the sexual

branch that were obtained by merging all wild-type and mutant cells and subsetting the branch of interest (Fig. S7) **(B)** Density plots showing the distribution of assigned female (left) and male (right) cells along each pseudotime trajectory and grouped by genotype. n = the number of cells for each condition as shown in panel A. WT = wild-type. The line indicates the median. **(C)** Venn diagrams showing the numbers of genes with differences in transcript abundance in female and male gametocytes, respectively, relative to wild-type. Genes that result in the absence of a sex cannot be evaluated. **(D)** Model of how sexual determination and differentiation is affected by a cascade dominated by putative nucleic acid binding proteins identified in this study. In addition to its role in commitment, *ap2g* probably has sex specific roles in differentiation (90), together with sex-specific transcription factors, such as *ap2-fg* (24).

Among the female differentiation genes, *fd1* encodes a putative RNA-binding protein, and *fd2* encodes a conserved *Plasmodium* protein. Both have profoundly perturbed female transcriptomes, and *fd1* mutant females expressed only some female markers. In contrast, *fd3* and *fd4* have more subtle roles for the formation of transcriptionally normal females. *fd3* is characterized by an AP2-coincident C-terminal domain of unknown function (ACDC). This type of domain was first identified in a number of *P. falciparum* AP2 domain-containing proteins (49), but in *fd3* it is not associated with a known DNA-binding domain. *fd4* encodes another putative zinc finger protein, and the mutant only shows a moderate downregulation of a few late female transcripts (Fig. 4C, S11 and Table S6). This gene is downregulated in *fd1* and *fd2* mutants, suggesting it may operate downstream of both.

Discussion

Our transcriptomic data suggest a model in which male and female gametocytes differentiate from a common sexual precursor in ways that rely on a cascade of nucleic acid binding proteins which are co-expressed downstream of AP2-G and fulfil distinct functions in a hierarchy of regulatory events (Fig. 4D). In this hierarchy the triple zinc finger protein GD1 is the strongest candidate for a top-level factor for female determination or differentiation upstream of the female-specific transcription factor AP2-FG (24). *Fd1* to *fd4* probably contribute downstream and with less impact (Fig. 4D).

Additional work is required to identify how sex ratio is determined in *P. berghei*. One view of how the sexes are formed in *P. falciparum* is that commitment to a particular sex co-incides with or even precedes the induction of sexual development by AP2-G (50). In that case, gene expression in response to AP2-G would be expected to follow a sex-specific pattern from the start. Importantly, neither our global transcriptomic analysis of single cells, nor the expression of candidate genes from the functional screen produced evidence that a sex-specific transcriptomic signature precedes the commitment to sexual development. Instead, we observed a common branch of sexual precursors and discovered key roles for AP2-G early response genes, suggesting sex may be determined downstream of *ap2-g* induction. Alternatively or additionally, non-transcriptional mechanisms may operate upstream of *md1-3* and *gd1*, involving for instance chromatin marks, differential splicing or phosphorylation states that determine the emergence of sex-specific transcriptional signatures downstream of AP2-G. Notwithstanding its early role in the switch to sexual development, AP2-G also binds to the upstream sequences of many male and female specific genes later during gametocytogenesis.

Genes that respond early during the induction of sexual development have also been identified in *P. falciparum*. All ten *P. berghei* genes validated here have orthologs in *P. falciparum* (Table

S7), where their transcripts are all upregulated during sexual development, with seven transcripts peaking during early (stage I-II) gametocytogenesis (51). Furthermore, chromatin immunoprecipitation of AP2-G in synchronously developing parasites identified *P. falciparum* orthologs of *gd1*, *fd1*, *fd2*, *md3*, and *md4* as likely direct targets for AP2-G binding in stage I gametocytes, but not in sexually committed ring stages or asexual schizonts (52). Taken together these data point at a high degree of conservation in the first steps of gametocyte differentiation following the initial commitment to sexual development in malaria parasites.

In summary, our functional screen, in combination with single-cell transcriptomes of cloned mutants, has identified a diverse group of proteins that are co-expressed downstream of AP2-G and whose deletion affects either the determination of, or differentiation along, a male or female cellular trajectory. Further analysis of these genes will shed light on the precise molecular mechanisms of sex ratio determination and sexual development. The screen revealed an abundance of putative RNA binding proteins, including the maleness inducing factor MD1 and the early response genes MD4, MD5 and FD1, whose molecular targets now need to be identified. The LOTUS/OST-HTH domain gene *md1* raises the intriguing possibility that the function of this domain in gametogenesis is conserved beyond animals, possibly to the origin of sex in the ancestral eukaryote. Sex determination mechanisms evolve rapidly but often involve RNA-dependent regulation, for instance through differential splicing or translational repression (53–55) and our data suggest similar principles operate in *P. berghei*. Discovering the molecular targets and interactors of the proteins of GD1, MD1 and FD1 now provide a route to establishing these mechanisms in more detail.

References

1. N. M. B. Brancucci, N. L. Bertschi, L. Zhu, I. Niederwieser, W. H. Chin, R. Wampfler, C. Freymond, M. Rottmann, I. Felger, Z. Bozdech, T. S. Voss, Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe*. **16**, 165–176 (2014).
2. M. Filarsky, S. A. Fraschka, I. Niederwieser, N. M. B. Brancucci, E. Carrington, E. Carrió, S. Moes, P. Jenoe, R. Bártfai, T. S. Voss, GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science*. **359**, 1259–1263 (2018).
3. B. F. C. Kafsack, N. Rovira-Graells, T. G. Clark, C. Bancells, V. M. Crowley, S. G. Campino, A. E. Williams, L. G. Drought, D. P. Kwiatkowski, D. A. Baker, A. Cortés, M. Llinás, A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*. **507**, 248–252 (2014).
4. A. Sinha, K. R. Hughes, K. K. Modrzynska, T. D. Otto, C. Pfander, N. J. Dickens, A. A. Religa, E. Bushell, A. L. Graham, R. Cameron, B. F. C. Kafsack, A. E. Williams, M. Llinás, M. Berriman, O. Billker, A. P. Waters, A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature*. **507**, 253–257 (2014).
5. O. Llorà-Batlle, L. Michel-Todó, K. Witmer, H. Toda, C. Fernández-Becerra, J. Baum, A. Cortés, Conditional expression of PfAP2-G for controlled massive sexual conversion in. *Sci Adv*. **6**, eaaz5057 (2020).
6. R. S. Kent, K. K. Modrzynska, R. Cameron, N. Philip, O. Billker, A. P. Waters, Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite *Plasmodium berghei*. *Nat Microbiol*. **3**, 1206–1213 (2018).
7. V. Robert, A. F. Read, J. Essong, T. Tchuinkam, B. Mulder, J. P. Verhave, P. Carnevale, Effect of gametocyte sex ratio on infectivity of *Plasmodium falciparum* to *Anopheles gambiae*.

Trans. R. Soc. Trop. Med. Hyg. **90**, 621–624 (1996).

- 8. S. E. Reece, D. R. Drew, A. Gardner, Sex ratio adjustment and kin discrimination in malaria parasites. *Nature*. **453**, 609–614 (2008).
- 9. R. E. Paul, T. N. Coulson, A. Raibaud, P. T. Brey, Sex Determination in Malaria Parasites. *Science*. **287**, 128–131 (2000).
- 10. F. G. Tadesse, L. Meerstein-Kessel, B. P. Gonçalves, C. Drakeley, L. Ranford-Cartwright, T. Bousema, Gametocyte Sex Ratio: The Key to Understanding *Plasmodium falciparum* Transmission? *Trends Parasitol.* **35**, 226–238 (2019).
- 11. U. Goodenough, J. Heitman, Origins of eukaryotic sexual reproduction. *Cold Spring Harb. Perspect. Biol.* **6** (2014), doi:10.1101/cshperspect.a016154.
- 12. D. Bachtrog, J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, Tree of Sex Consortium, Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899 (2014).
- 13. E. Bushell, A. R. Gomes, T. Sanderson, B. Anar, G. Girling, C. Herd, T. Metcalf, K. Modrzynska, F. Schwach, R. E. Martin, M. W. Mather, G. I. McFadden, L. Parts, G. G. Rutledge, A. B. Vaidya, K. Wengelnik, J. C. Rayner, O. Billker, Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell*. **170**, 260–272.e8 (2017).
- 14. A. R. Gomes, E. Bushell, F. Schwach, G. Girling, B. Anar, M. A. Quail, C. Herd, C. Pfander, K. Modrzynska, J. C. Rayner, O. Billker, A Genome-Scale Vector Resource Enables High-Throughput Reverse Genetic Screening in a Malaria Parasite. *Cell Host & Microbe*. **17** (2015), pp. 404–413.
- 15. R. R. Stanway, E. Bushell, A. Chiappino-Pepe, M. Roques, T. Sanderson, B. Franke-Fayard, R. Caldelari, M. Golomangi, M. Nyonda, V. Pandey, F. Schwach, S. Chevalley, J. Ramesar, T. Metcalf, C. Herd, P.-C. Burda, J. C. Rayner, D. Soldati-Favre, C. J. Janse, V. Hatzimanikatis, O. Billker, V. T. Heussler, Genome-Scale Identification of Essential Metabolic Processes for Targeting the *Plasmodium* Liver Stage. *Cell*. **179**, 1112–1128.e26 (2019).
- 16. M. Zhang, C. Wang, T. D. Otto, J. Oberstaller, X. Liao, S. R. Adapa, K. Udenze, I. F. Bronner, D. Casandra, M. Mayho, J. Brown, S. Li, J. Swanson, J. C. Rayner, R. H. Y. Jiang, J. H. Adams, Uncovering the essential genes of the human malaria parasite by saturation mutagenesis. *Science*. **360** (2018), doi:10.1126/science.aap7847.
- 17. G. R. Mair, E. Lasonder, L. S. Garver, B. M. D. Franke-Fayard, C. K. Carret, J. C. A. G. Wiegant, R. W. Dirks, G. Dimopoulos, C. J. Janse, A. P. Waters, Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog.* **6**, e1000767 (2010).
- 18. F. Schwach, E. Bushell, A. R. Gomes, B. Anar, G. Girling, C. Herd, J. C. Rayner, O. Billker, PlasmoGEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. *Nucleic Acids Res.* **43**, D1176–82 (2015).
- 19. M. Yuda, S. Iwanaga, I. Kaneko, T. Kato, Global transcriptional repression: An initial and essential step for *Plasmodium* sexual development. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12824–12829 (2015).
- 20. R. E. Sinden, E. J. Dawes, Y. Alavi, J. Waldock, O. Finney, J. Mendoza, G. A. Butcher, L. Andrews, A. V. Hill, S. C. Gilbert, M.-G. Basáñez, Progression of *Plasmodium berghei* through *Anopheles stephensi* is density-dependent. *PLoS Pathog.* **3**, e195 (2007).

21. R. Hoo, L. Zhu, A. Amaladoss, S. Mok, O. Natalang, S. A. Lapp, G. Hu, K. Liew, M. R. Galinski, Z. Bozdech, P. R. Preiser, Integrated analysis of the *Plasmodium* species transcriptome. *EBioMedicine*. **7**, 255–266 (2016).
22. A. Guerreiro, E. Deligianni, J. M. Santos, P. A. G. C. Silva, C. Louis, A. Pain, C. J. Janse, B. Franke-Fayard, C. K. Carret, I. Sidén-Kiamos, G. R. Mair, Genome-wide RIP-Chip analysis of translational repressor-bound mRNAs in the *Plasmodium* gametocyte. *Genome Biol.* **15**, 493 (2014).
23. Z. Li, H. Cui, J. Guan, C. Liu, Z. Yang, J. Yuan, Plasmodium transcription repressor AP2-O3 regulates sex-specific identity of gene expression in female gametocytes. *EMBO Rep.*, e51660 (2021).
24. M. Yuda, I. Kaneko, S. Iwanaga, Y. Murata, T. Kato, Female-specific gene regulation in malaria parasites by an AP2-family transcription factor. *Mol. Microbiol.* **113**, 40–51 (2020).
25. S. M. Khan, B. Franke-Fayard, G. R. Mair, E. Lasonder, C. J. Janse, M. Mann, A. P. Waters, Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell.* **121**, 675–687 (2005).
26. M. Ponzi, I. Sidén-Kiamos, L. Bertuccini, C. Currà, H. Kroeze, G. Camarda, T. Pace, B. Franke-Fayard, E. C. Laurentino, C. Louis, A. P. Waters, C. J. Janse, P. Alano, Egress of *Plasmodium berghei* gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. *Cell. Microbiol.* **11**, 1272–1288 (2009).
27. J. Miao, Z. Chen, Z. Wang, S. Shrestha, X. Li, R. Li, L. Cui, Sex-Specific Biology of the Human Malaria Parasite Revealed from the Proteomes of Mature Male and Female Gametocytes. *Mol. Cell. Proteomics*. **16**, 537–551 (2017).
28. E. Lasonder, S. R. Rijpma, B. C. L. van Schaijk, W. A. M. Hoeijmakers, P. R. Kensche, M. S. Gresnigt, A. Italiaander, M. W. Vos, R. Woestenenk, T. Bousema, G. R. Mair, S. M. Khan, C. J. Janse, R. Bárta, R. W. Sauerwein, Integrated transcriptomic and proteomic analyses of *P. falciparum* gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res.* **44**, 6087–6101 (2016).
29. S. M. Scholz, N. Simon, C. Lavazec, M.-A. Dude, T. J. Templeton, G. Pradel, PfCCP proteins of *Plasmodium falciparum*: gametocyte-specific expression and role in complement-mediated inhibition of exflagellation. *Int. J. Parasitol.* **38**, 327–340 (2008).
30. M. A. Olshina, H. Baumann, K. R. Willison, J. Baum, Plasmodium actin is incompletely folded by heterologous protein-folding machinery and likely requires the native *Plasmodium* chaperonin complex to enter a mature functional state. *FASEB J.* **30**, 405–416 (2016).
31. M. del Carmen Rodriguez, P. Gerold, J. Dessens, K. Kurtenbach, R. T. Schwartz, R. E. Sinden, G. Margos, Characterisation and expression of pbs25, a sexual and sporogonic stage specific protein of *Plasmodium berghei*. *Mol. Biochem. Parasitol.* **110**, 147–159 (2000).
32. A. M. Tomas, G. Margos, G. Dimopoulos, L. H. van Lin, T. F. de Koning-Ward, R. Sinha, P. Lupetti, A. L. Beetsma, M. C. Rodriguez, M. Karras, A. Hager, J. Mendoza, G. A. Butcher, F. Kafatos, C. J. Janse, A. P. Waters, R. E. Sinden, P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. *EMBO J.* **20**, 3975–3983 (2001).
33. P. Schneider, S. E. Reece, B. C. L. van Schaijk, T. Bousema, K. H. W. Lanke, C. S. J. Meaden, A. Gadalla, L. C. Ranford-Cartwright, H. A. Babiker, Quantification of female and male *Plasmodium falciparum* gametocytes by reverse transcriptase quantitative PCR. *Mol. Biochem. Parasitol.* **199**, 29–33 (2015).

34. S.-B. Malik, M. A. Ramesh, A. M. Hulstrand, J. M. Logsdon Jr, Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Mol. Biol. Evol.* **24**, 2827–2841 (2007).
35. G. Mlambo, I. Coppens, N. Kumar, Aberrant sporogonic development of Dmc1 (a meiotic recombinase) deficient *Plasmodium berghei* parasites. *PLoS One.* **7**, e52480 (2012).
36. J. Ning, T. D. Otto, C. Pfander, F. Schwach, M. Brochet, E. Bushell, D. Goulding, M. Sanders, P. A. Lefebvre, J. Pei, N. V. Grishin, G. Vanderlaan, O. Billker, W. J. Snell, Comparative genomics in *Chlamydomonas* and *Plasmodium* identifies an ancient nuclear envelope protein family essential for sexual reproduction in protists, fungi, plants, and vertebrates. *Genes Dev.* **27**, 1198–1215 (2013).
37. O. Billker, S. Dechamps, R. Tewari, G. Wenig, B. Franke-Fayard, V. Brinkmann, Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. *Cell.* **117**, 503–514 (2004).
38. Y. Liu, R. Tewari, J. Ning, A. M. Blagborough, S. Garbom, J. Pei, N. V. Grishin, R. E. Steele, R. E. Sinden, W. J. Snell, O. Billker, The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev.* **22**, 1051–1068 (2008).
39. A. C. Balestra, M. Zeeshan, E. Rea, C. Pasquarello, L. Brusini, T. Mourier, A. K. Subudhi, N. Klages, P. Arboit, R. Pandey, D. Brady, S. Vaughan, A. A. Holder, A. Pain, D. J. Ferguson, A. Hainard, R. Tewari, M. Brochet, A divergent cyclin/cyclin-dependent kinase complex controls the atypical replication of a malaria parasite during gametogony and transmission. *eLife.* **9** (2020), doi:10.7554/eLife.56474.
40. S. Shrestha, X. Li, G. Ning, J. Miao, L. Cui, The RNA-binding protein Puf1 functions in the maintenance of gametocytes in *Plasmodium falciparum*. *J. Cell Sci.* **129**, 3144–3152 (2016).
41. J. Miao, J. Li, Q. Fan, X. Li, X. Li, L. Cui, The Puf-family RNA-binding protein PfPuf2 regulates sexual development and sex differentiation in the malaria parasite *Plasmodium falciparum*. *J. Cell Sci.* **123**, 1039–1049 (2010).
42. G. R. Mair, J. A. M. Braks, L. S. Garver, J. C. A. G. Wiegant, N. Hall, R. W. Dirks, S. M. Khan, G. Dimopoulos, C. J. Janse, A. P. Waters, Regulation of sexual development of *Plasmodium* by translational repression. *Science.* **313**, 667–669 (2006).
43. V. Anantharaman, D. Zhang, L. Aravind, OST-HTH: a novel predicted RNA-binding domain. *Biol. Direct.* **5**, 13 (2010).
44. I. Callebaut, J.-P. Mornon, LOTUS, a new domain associated with small RNA pathways in the germline. *Bioinformatics.* **26** (2010), pp. 1140–1144.
45. J. Kubíková, R. Reinig, H. K. Salgania, M. Jeske, LOTUS-domain proteins - developmental effectors from a molecular perspective. *Biol. Chem.* **402**, 7–23 (2020).
46. A. Ephrussi, R. Lehmann, Induction of germ cell formation by oskar. *Nature.* **358**, 387–392 (1992).
47. T. Tanaka, M. Hosokawa, V. V. Vagin, M. Reuter, E. Hayashi, A. L. Mochizuki, K. Kitamura, H. Yamanaka, G. Kondoh, K. Okawa, S. Kuramochi-Miyagawa, T. Nakano, R. Sachidanandam, G. J. Hannon, R. S. Pillai, N. Nakatsuji, S. Chuma, Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10579–10584 (2011).
48. D. Wilsker, A. Patsialou, P. B. Dallas, E. Moran, ARID proteins: a diverse family of DNA

binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ.* **13**, 95–106 (2002).

49. S. C. Oehring, B. J. Woodcroft, S. Moes, J. Wetzel, O. Dietz, A. Pulfer, C. Dekiwadia, P. Maeser, C. Flueck, K. Witmer, N. M. B. Brancucci, I. Niederwieser, P. Jenoe, S. A. Ralph, T. S. Voss, Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol.* **13**, R108 (2012).
50. F. Silvestrini, P. Alano, J. L. Williams, Commitment to the production of male and female gametocytes in the human malaria parasite *Plasmodium falciparum*. *Parasitology*. **121 Pt 5**, 465–471 (2000).
51. R. van Biljon, R. van Wyk, H. J. Painter, L. Orchard, J. Reader, J. Niemand, M. Llinás, L.-M. Birkholtz, Hierarchical transcriptional control regulates *Plasmodium falciparum* sexual differentiation. *BMC Genomics*. **20**, 920 (2019).
52. G. A. Josling, T. J. Russell, J. Venezia, L. Orchard, R. van Biljon, H. J. Painter, M. Llinás, Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nature Communications*. **11** (2020), , doi:10.1038/s41467-020-15026-0.
53. L. O. F. Penalva, L. Sánchez, *Microbiol. Mol. Biol. Rev.*, in press.
54. J. Mapes, J.-T. Chen, J.-S. Yu, D. Xue, Somatic sex determination in *Caenorhabditis elegans* is modulated by SUP-26 repression of tra-2 translation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18022–18027 (2010).
55. S. Cho, Z. Y. Huang, J. Zhang, Sex-specific splicing of the honeybee doublesex gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. *Genetics*. **177**, 1733–1741 (2007).
56. C. Pfander, B. Anar, F. Schwach, T. D. Otto, M. Brochet, K. Volkmann, M. A. Quail, A. Pain, B. Rosen, W. Skarnes, J. C. Rayner, O. Billker, A scalable pipeline for highly effective genetic modification of a malaria parasite. *Nat. Methods*. **8**, 1078–1082 (2011).
57. N. Hall, M. Karras, J. D. Raine, J. M. Carlton, T. W. A. Kooij, M. Berriman, L. Florens, C. S. Janssen, A. Pain, G. K. Christophides, K. James, K. Rutherford, B. Harris, D. Harris, C. Churcher, M. A. Quail, D. Ormond, J. Doggett, H. E. Trueman, J. Mendoza, S. L. Bidwell, M.-A. Rajandream, D. J. Carucci, J. R. Yates 3rd, F. C. Kafatos, C. J. Janse, B. Barrell, C. M. R. Turner, A. P. Waters, R. E. Sinden, A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science*. **307**, 82–86 (2005).
58. P.-C. Burda, M. A. Roelli, M. Schaffner, S. M. Khan, C. J. Janse, V. T. Heussler, A *Plasmodium* phospholipase is involved in disruption of the liver stage parasitophorous vacuole membrane. *PLoS Pathog.* **11**, e1004760 (2015).
59. C. J. Janse, J. Ramesar, A. P. Waters, High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite *Plasmodium berghei*. *Nat. Protoc.* **1**, 346–356 (2006).
60. L. Reininger, O. Billker, R. Tewari, A. Mukhopadhyay, C. Fennell, D. Dorin-Semblat, C. Doerig, D. Goldring, L. Harmse, L. Ranford-Cartwright, J. Packer, C. Doerig, A NIMA-related protein kinase is essential for completion of the sexual cycle of malaria parasites. *J. Biol. Chem.* **280**, 31957–31964 (2005).
61. L. Chappell, P. Ross, L. Orchard, T. J. Russell, T. D. Otto, M. Berriman, J. C. Rayner, M. Llinás, Refining the transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-free RNA-seq. *BMC Genomics*. **21**, 395 (2020).

62. A. J. Reid, A. M. Talman, H. M. Bennett, A. R. Gomes, M. J. Sanders, C. J. R. Illingworth, O. Billker, M. Berriman, M. K. Lawniczak, Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *Elife*. **7** (2018), doi:10.7554/eLife.33105.
63. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
64. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
65. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67. R. Wehrens, L. M. C. Buydens, Self- and Super-organizing Maps in R: The Kohonen Package. *Journal of Statistical Software*. **21** (2007), , doi:10.18637/jss.v021.i05.
68. G. Tischler, S. Leonard, biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine*. **9** (2014), , doi:10.1186/1751-0473-9-13.
69. FelixKrueger, FelixKrueger/TrimGalore. *GitHub*, (available at <https://github.com/FelixKrueger/TrimGalore>).
70. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*. **12**, 357–360 (2015).
71. A. Fougère, A. P. Jackson, D. P. Bechtel, J. A. M. Braks, T. Annoura, J. Fonager, R. Spaccapelo, J. Ramesar, S. Chevalley-Maurel, O. Klop, A. M. A. van der Laan, H. J. Tanke, C. H. M. Kocken, E. M. Pasini, S. M. Khan, U. Böhme, C. van Ooij, T. D. Otto, C. J. Janse, B. Franke-Fayard, Variant Exported Blood-Stage Proteins Encoded by Plasmodium Multigene Families Are Expressed in Liver Stages Where They Are Exported into the Parasitophorous Vacuole. *PLoS Pathog.* **12**, e1005917 (2016).
72. F. J. Logan-Klumpler, N. De Silva, U. Boehme, M. B. Rogers, G. Velarde, J. A. McQuillan, T. Carver, M. Aslett, C. Olsen, S. Subramanian, I. Phan, C. Farris, S. Mitra, G. Ramasamy, H. Wang, A. Tivey, A. Jackson, R. Houston, J. Parkhill, M. Holden, O. S. Harb, B. P. Brunk, P. J. Myler, D. Roos, M. Carrington, D. F. Smith, C. Hertz-Fowler, M. Berriman, GeneDB—an annotation database for pathogens. *Nucleic Acids Res.* **40**, D98–108 (2012).
73. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
74. R Core Team, R: A Language and Environment for Statistical Computing (2020), (available at <https://www.R-project.org/>).
75. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888–1902.e21 (2019).
76. T. Benaglia, D. Chauveau, D. R. Hunter, D. Young, mixtools: An R Package for Analyzing Finite

Mixture Models. *Journal of Statistical Software*. **32** (2009), , doi:10.18637/jss.v032.i06.

77. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
78. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018), (available at <http://arxiv.org/abs/1802.03426>).
79. R. Rotta, A. Noack, Multilevel local search algorithms for modularity clustering. *ACM Journal of Experimental Algorithms*. **16** (2011), , doi:10.1145/1963190.1970376.
80. X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*. **14**, 979–982 (2017).
81. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. **566**, 496–502 (2019).
82. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
83. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. Juliana McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology*. **16** (2015), , doi:10.1186/s13059-015-0844-5.
84. T. F. de Koning-Ward, D. R. Drew, J. M. Chesson, J. G. Beeson, B. S. Crabb, Truncation of Plasmodium berghei merozoite surface protein 8 does not affect in vivo blood-stage development. *Mol. Biochem. Parasitol.* **159**, 69–72 (2008).
85. C. G. Black, T. Wu, L. Wang, A. E. Topolska, R. L. Coppel, MSP8 is a non-essential merozoite surface protein in Plasmodium falciparum. *Mol. Biochem. Parasitol.* **144**, 27–35 (2005).
86. S. Das, N. Hertrich, A. J. Perrin, C. Withers-Martinez, C. R. Collins, M. L. Jones, J. M. Watermeyer, E. T. Fobes, S. R. Martin, H. R. Saibil, G. J. Wright, M. Treeck, C. Epp, M. J. Blackman, Processing of Plasmodium falciparum Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs. *Cell Host Microbe*. **18**, 433–444 (2015).
87. J. D. Dvorin, D. C. Martyn, S. D. Patel, J. S. Grimley, C. R. Collins, C. S. Hopp, A. T. Bright, S. Westenberger, E. Winzeler, M. J. Blackman, D. A. Baker, T. J. Wandless, M. T. Duraisingham, A plant-like kinase in Plasmodium falciparum regulates parasite egress from erythrocytes. *Science*. **328**, 910–912 (2010).
88. A. Gupta, P. Mehra, R. Nitharwal, A. Sharma, A. K. Biswas, S. K. Dhar, Analogous expression pattern of Plasmodium falciparum replication initiation proteins PfMCM4 and PfORC1 during the asexual and sexual stages of intraerythrocytic developmental cycle. *FEMS Microbiology Letters*. **261** (2006), pp. 12–18.
89. V. M. Howick, A. J. C. Russell, T. Andrews, H. Heaton, A. J. Reid, K. Natarajan, H. Butungi, T. Metcalf, L. H. Verzier, J. C. Rayner, M. Berriman, J. K. Herren, O. Billker, M. Hemberg, A. M. Talman, M. K. N. Lawniczak, The Malaria Cell Atlas: Single parasite transcriptomes across the complete life cycle. *Science*. **365** (2019), doi:10.1126/science.aaw2619.
90. M. Yuda, I. Kaneko, Y. Murata, S. Iwanaga, T. Nishi, Mechanisms of triggering malaria gametocytogenesis by AP2-G. *Parasitol. Int.* **84**, 102403 (2021).

Acknowledgements

The authors would like to thank the staff of the Illumina Bespoke Sequencing and Core Cytometry teams at the Wellcome Sanger Institute for their contribution. **Funding:** Work at the Wellcome Sanger Institute was funded by Wellcome core grant 206194/Z/17/Z awarded to OB, M.L. and A.J.C.R. Work at Umeå University received funding from the Knut and Alice Wallenberg Foundation and the European Research Council (Grant agreement No. 788516). Work at the University of Glasgow (APW, ABR, RSK) was funded by the Wellcome Trust (Refs: 083811, 104111 & 107046 to APW and the BBSRC (Ref BB/J013854/1 to RSK). RSK is supported by BBSRC (Ref BB/J013854/1). KKM is supported by Wellcome Trust and Royal Society (202600/Z/16/Z). MH is supported by SNF (P2SKP3_187635) and HFSP (LT000131/2020-L). **Competing Interests:** The authors declare they have no competing interests. **Data and materials availability:** The raw scRNA-seq data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB44892) (<https://www.ebi.ac.uk/ena/browser/view/PRJEB44892>) and raw bulk RNA-seq data are available from the NCBI Gene Expression Omnibus (GEO) (accession number GSE110201 and GSE168817). Supporting files and code are available on Github at <https://github.com/andyrussell/Gametocytogenesis>. Screen data can be visualised on the *PlasmoGEM* website https://plasmogem.shinyapps.io/Gametocytes_Shiny/. Single-cell RNA-seq data can be searched and visualised on the Malaria Cell Atlas website www.malariaatlas.org. Mutant scRNA-seq data can be interacted with on the cellxgene instance <http://obilab.molbiol.umu.se/gcsko>.

Supplemental information

Materials and Methods

Supplementary Tables S1 – S8

Supplementary Figures S1 – S14