RESEARCH ARTICLE SUMMARY

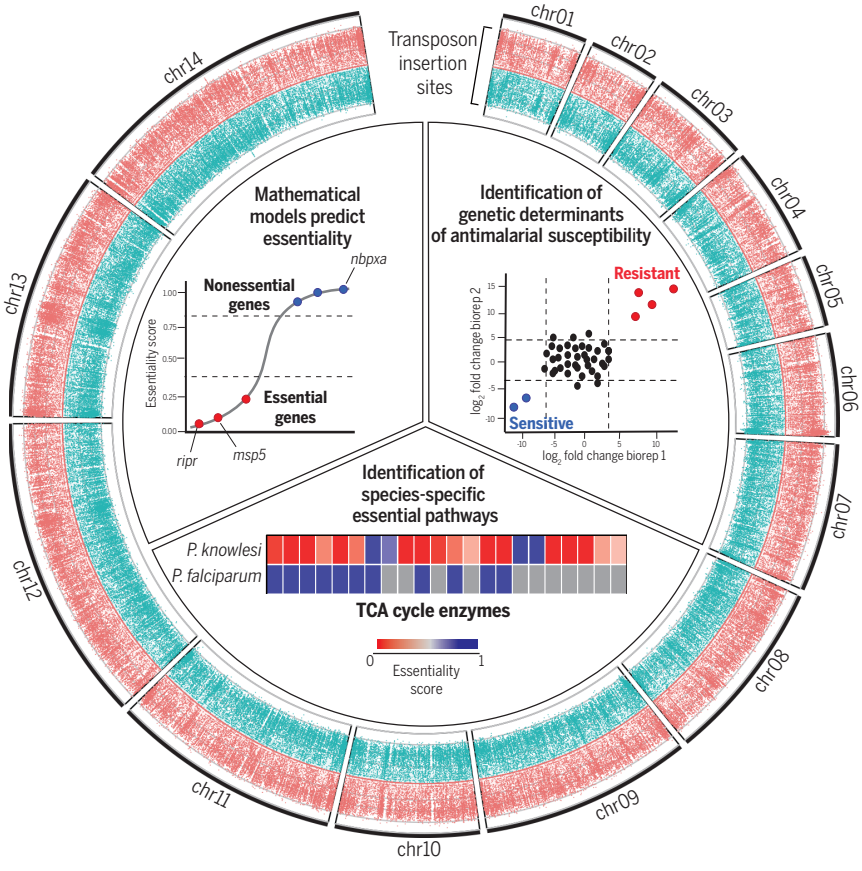
MALARIA

# The essential genome of *Plasmodium knowlesi* reveals determinants of antimalarial susceptibility

**Brendan Elsworth†, Sida Ye†, Sheena Dass†, Jacob A. Tennessen, Qudseen Sultana, Basil T. Thommen, Aditya S. Paul, Usheer Kanjee, Christof Grüring, Marcelo U. Ferreira, Marc-Jan Gubbels, Kourosh Zarringhalam*, Manoj T. Duraisingh***

**INTRODUCTION:** Approximately 249 million cases of malaria, caused by infection with *Plasmodium* species parasites, result in about 608,000 deaths annually. Measures to combat the parasites that cause malaria have become compromised because of reliance on a small arsenal of drugs and emerging drug resistance. Moreover, zoonotic cases of malaria are increasing in prevalence in Southeast Asia, caused by infection with *Plasmodium knowlesi* and *Plasmodium cynomolgi*. These parasites are evolutionarily closely related to *Plasmodium vivax* and serve as powerful in vitro models because *P. vivax* cannot be cultured in vitro.

**RATIONALE:** The genomes of each *Plasmodium* spp. reveal conservation but also substantial differences. Although functional genetic studies in *Plasmodium* spp. parasites have provided information on the essentiality and function of genes throughout the *Plasmodium* life cycle, the lack of high-confidence genome-wide essentiality data limits the ability to prioritize targets for vaccines and therapeutics, particularly for parasites in the *P. vivax* clade, which includes *P. knowlesi*. In this study, we harnessed the development of high-efficiency molecular genetics in *P. knowlesi* to elicit genome-wide piggyBac transposon mutagenesis and provide the most complete determination of gene essentiality for the blood-stage infection in any *Plasmodium* spp. with the resolution to define truncatable genes.




**A high-resolution transposon mutagenesis screen in *P. knowlesi* informs antimalarial discovery and reveals drug resistance determinants.** A near-saturation piggyBac transposon screen in *P. knowlesi* and improved models for calling gene essentiality provide a high-confidence genome-wide map of gene essentiality, allowing for comparative essentiality studies, prioritization of drug and vaccine targets, and identification of genes that alter antimalarial susceptibility. chr, chromosome.

**RESULTS:** We observed 1,456,750 independent transposon insertions, ~38-fold more than was previously achieved in the *Plasmodium falciparum* piggyBac screen. A mathematical simulation was used to determine that we have achieved a 3.6-fold higher number of insertions than the required level of insertions to reach gene-level saturation. Overall, 2037 genes (38.68%) with at least one TTAA site (that serve as targeting sites for the piggyBac transposon) were classified as essential, 2124 (40.34%) were classified as dispensable, and 1105 (20.98%) were indeterminate. The number of essential genes is similar to previous observations in apicomplexans but is higher than in many free-living organisms, which reflects the highly specialized parasitic lifestyle of *Plasmodium* spp. that relies heavily on host metabolism. We further identified 126 genes whose disruption results in reduced fitness, 122 genes that can tolerate 5′ or 3′ truncation, and 53 essential long noncoding RNAs (lncRNAs). We found that the essentiality of the druggable genome is broadly conserved between *Plasmodium* spp.; however, we identified the *P. falciparum* drug target bifunctional farnesyl/geranylgeranyl diphosphate synthase (FPP/GGPPS) as dispensable in *P. knowlesi*. Notably, *P. knowlesi* showed a greater reliance on the tricarboxylic acid (TCA) cycle genes, and both *P. knowlesi* and *P. vivax*—in contrast to *P. falciparum*—are inhibited by the aconitase inhibitor sodium fluoroacetate. Perturbation analyses with the frontline antimalarial artemisinin revealed modulators that both increase and decrease susceptibility.

**CONCLUSION:** The essentiality of the *P. knowlesi* genome will be a useful tool for antimalarial drug and vaccine target prioritization and for comparative essentiality studies across apicomplexan parasites and eukaryotes. To make the data readily accessible to the community, we provide a web-based app that can be used interactively to examine gene essentiality scores (https://umbibio.math.umb.edu/PkEssenDB/); the data are also available through VEuPathDB. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: kourosh.zarringhalam@umb.edu (K.Z.); mduraisi@hsph.harvard.edu (M.T.D.)
†These authors contributed equally to this work.
Cite this article as B. Elsworth *et al.*, *Science* **387**, eadq6241 (2025). DOI: 10.1126/science.adq6241

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adq6241

RESEARCH ARTICLE

MALARIA

# The essential genome of *Plasmodium knowlesi* reveals determinants of antimalarial susceptibility

**Brendan Elsworth[1,2]†, Sida Ye[1,3,4]†, Sheena Dass[1]†, Jacob A. Tennessen[1], Qudseen Sultana[4,5], Basil T. Thommen[1], Aditya S. Paul[1], Usheer Kanjee[1], Christof Grüring[1]‡, Marcelo U. Ferreira[6,7], Marc-Jan Gubbels[8], Kourosh Zarringhalam[3,4]*, Manoj T. Duraisingh[1]***

**Measures to combat the parasites that cause malaria have become compromised because of reliance on a small arsenal of drugs and emerging drug resistance. We conducted a transposon mutagenesis screen in the primate malaria parasite *Plasmodium knowlesi*, producing the most complete classification of gene essentiality in any *Plasmodium* spp. to date, with the resolution to define truncatable genes. We found conservation in the druggable genome between *Plasmodium* spp. and divergences in mitochondrial metabolism. Perturbation analyses with the frontline antimalarial artemisinin revealed modulators that both increase and decrease drug susceptibility. Our findings aid prioritization of drug and vaccine targets for the *Plasmodium vivax* clade and reveal mechanisms of resistance that can inform therapeutic development.**

Approximately 249 million cases of malaria, caused by infection with *Plasmodium* spp. parasites, result in about 608,000 deaths annually (*1*). Efforts to reduce the global burden of malaria have yielded diminishing returns in recent years (*1*). Five *Plasmodium* spp. are responsible for most malaria cases in humans. Although vaccines against *Plasmodium falciparum* are becoming available, treatment of the disease has become increasingly compromised as a result of resistance emerging to antimalarial drugs (*2*, *3*). Moreover, zoonotic cases of malaria are increasing in prevalence in Southeast Asia, caused by infection with *P. knowlesi* and *Plasmodium cynomolgi*, naturally found in cynomolgus macaques (*Macaca fascicularis*). *P. falciparum* (and *Laverania* spp.) are evolutionarily distinct from the majority of *Plasmodium* spp. that infect primates, whereas *P. knowlesi* and *P. cynomolgi* sit within the *P. vivax* clade that are more representative of the majority of *Plasmodium* spp. and serve as valuable in vitro models because *P. vivax* cannot be cultured in vitro (*4*–*8*). The genomes of each *Plasmodium* spp. reveal conservation but also substantial differences. Although functional genetic studies in *Plasmodium* spp. parasites have provided information on the essentiality and function of genes throughout the *Plasmodium* life cycle, the lack of high-confidence genome-wide essentiality data limits the ability to prioritize targets for vaccines and therapeutics, particularly for parasites in the *P. vivax* clade, which includes *P. knowlesi*. In this work, we harnessed the development of high-efficiency molecular genetics in *P. knowlesi* to elicit genome-wide transposon mutagenesis to provide the most complete determination of gene essentiality for the blood-stage infection in any *Plasmodium* spp.

[1]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [2]Laboratory of Emerging Pathogens, Division of Emerging and Transfusion Transmitted Diseases, Office of Blood Research and Review, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA. [3]Department of Mathematics, University of Massachusetts Boston, Boston, MA, USA. [4]Center for Personalized Cancer Therapy, University of Massachusetts Boston, Boston, MA, USA. [5]Department of Computer Science, University of Massachusetts Boston, Boston, MA, USA. [6]Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil. [7]Global Health and Tropical Medicine, Institute of Hygiene and Tropical Medicine, Nova University of Lisbon, Lisbon, Portugal. [8]Department of Biology, Boston College, Chestnut Hill, MA, USA.

***Corresponding author. Email: kourosh.zarringhalam@umb.edu (K.Z.); mduraisi@hsph.harvard.edu (M.T.D.)**

†These authors contributed equally to this work.

‡Present address: Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Allschwil, Switzerland.

## High-density transposon mutagenesis of the *P. knowlesi* genome

To generate the transposon mutant library for *P. knowlesi*, which like all *Plasmodium* spp. is haploid during blood-stage proliferation, we used a hyperactive piggyBac transposase (*9*). *P. knowlesi* transfectants were generated in rhesus macaque red blood cells (RBCs)—which are more physiologically relevant than human RBCs for *P. knowlesi* and support higher transfection efficiency—and insertions were mapped using a modified quantitative insertion-site sequencing (QIseq) approach (fig. S1A) (*10*). As expected, many presumably essential genes (e.g., ribosomal genes) contained a low number of reads, likely owing to a low level of continued piggyBac transposition or the presence of dead parasites. We therefore used a curated set of background correction genes, the orthologs of which are essential in multiple apicomplexan parasites, to subtract background signal from all potential sites (fig. S1B and data S1). There are 158,734 TTAA sites in the *P. knowlesi* nuclear genome that serve as targeting sites for the piggyBac transposon. All models described in this work use either the number of reads per site or the number of independent insertion events (IIEs), defined as the presence or absence of insertions in a site, to provide a measure of relative abundance of mutants in the library. TTAA sites are palindromic, and orientation of insertions can be differentiated, resulting in up to two IIEs per site per transfection pool (with a maximum of 20 IIEs per site across 10 unique transfection pools) (fig. S1C). We observed 1,456,750 IIEs in our screen after background correction, covering 80.1% of all TTAA sites and 99.3% of protein coding genes with at least one TTAA (Fig. 1, A and B, and fig. S1, D to H). Summary statistics can be found in data S2. This is ~38-fold more insertions than was previously achieved in the *P. falciparum* piggyBac screen (*11*). Quantitative polymerase chain reaction (qPCR) was used to determine that the average number of transposon insertions per parasite genome was 1.76 (fig. S2A). Of sites that contained an insertion, 89.0% contained multiple independent insertions, which suggests that near-saturation mutagenesis of all TTAA sites was achieved (Fig. 1B and fig. S1, D to F). Sites in exons and introns contained fewer insertions compared with intergenic sites because mutants with insertions in essential genes will not proliferate, as seen for *ripr*, a known essential gene in *P. knowlesi* (Fig. 1, C and D) (*12*). Within individual genes, the number of reads mapped to introns and exons were weakly correlated to one another (fig. S2B). Because there are some discrepancies between insertion in introns and exons, we limited further analysis to insertions within coding sequences (CDS) that are most likely to disrupt gene function to increase the confidence in our calls, albeit with a loss of TTAA sites. Moreover, we examined sequence-dependent insertional bias using a hidden Markov model (HMM). The HMM was trained to identify sites with fewer-than-expected insertions (coldspots; 1.9% of all sites) in regions of the genome that are expected to be nonessential based on high insertions in surrounding sites. Motif analysis in the flanking regions revealed that although most TTAA sites in the genome have a poly-T upstream, coldspots have a poly-T downstream (Fig. 1E and fig. S3, A to C). The coldspot motif was observed in <1% of all TTAA sites and therefore was not further considered when determining gene essentiality (fig. S3, D and E). To assess whether we had achieved the number of insertions required to reach gene-level saturation, we performed a mathematical simulation, which indicated that we have achieved 3.6- or 4.7-fold higher number of insertions than the required level of insertions, assuming that 0% (upper bound) or 40%

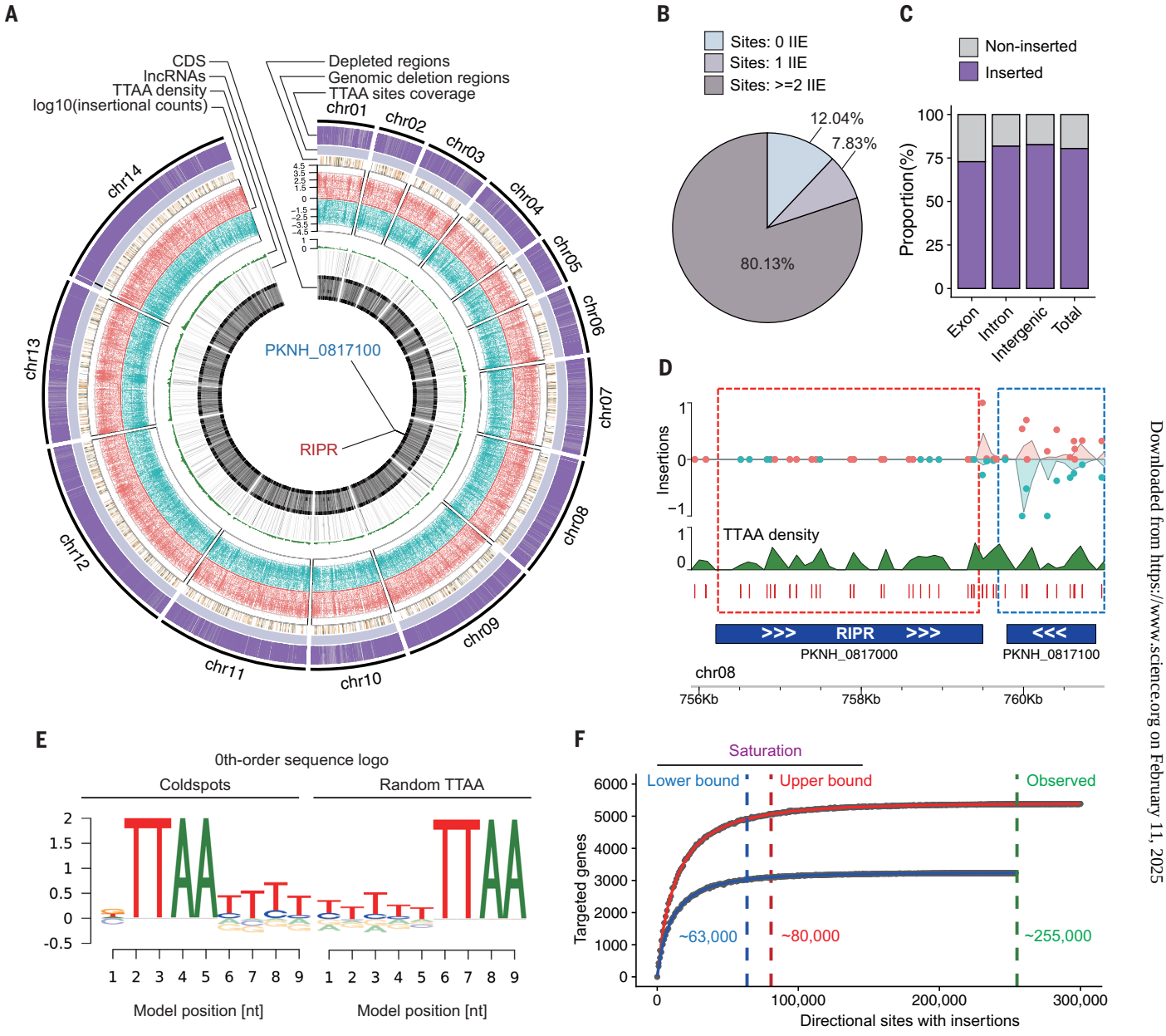



**Fig. 1. Generation of a near–site-level saturation transposon mutant library in *P. knowlesi*.** (**A**) Circos plot showing even distribution of transposon insertions throughout the genome. Track colors are as follows: TTAA sites with an observed transposon insertion across the genome (purple), genomic deletions in YH1 strain compared with the reference H strain (black bar on light purple background), insertionally depleted regions with ≥5 consecutive TTAA sites with reduced insertions identified by a HMM overlapping with coding regions (light blue), read counts mapped to each TTAA site on the + strand (red) or at the same site on the − strand (teal), the TTAA density (potential insertion sites) across the genome (dark green), lncRNAs identified in this study (black), and CDS (black). Genomic locations of two representative genes are marked (blue and red text). (**B**) Pie chart showing number of TTAA sites that contain 0, 1, or ≥2 IIEs (fig. S1). (**C**) Background-corrected proportion of TTAA sites in exons, introns, and intergenic regions that have at least one insertion (inserted; purple) or no insertions (noninserted; gray). (**D**) Zoomed-in region from (B) showing that the known essential gene in *P. knowlesi*, *ripr* (red box), has very few read counts mapping to transposon insertions (*y* axis), whereas the intergenic region and neighboring gene (blue box; PKNH_0817100, which is nonessential in *P. falciparum* and *P. berghei*) have insertions throughout the gene. Insertions (read counts reflecting relative abundance of mutants) and TTAA density have been normalized in the window shown to range from 0 to 1. Red lines indicate TTAA sites in the genome that serve as potential insertion sites. (**E**) Identification of a motif associated with reduced transposon insertion throughout the genome (coldspots) (see also fig. S2). nt, nucleotide. (**F**) Mathematical simulation showing the number of TTAA sites in the genome that contain at least one insertion [observed (OB)] is significantly higher than that required to reach gene-level saturation using the lower bound (LB) (40% essential genes) and upper bound (UB) (0% essential genes). For greater accuracy in measuring saturation, both orientations of each TTAA site are counted separately in this model. The total number of sites is 317,468.

(lower bound; based on previous essentiality data in *Plasmodium* spp.) of genes are essential, respectively (Fig. 1F).

### Improved models to identify gene essentiality in *P. knowlesi*

To make high-confidence calls of gene essentiality at the genome-wide level, we considered, for each gene, the total number of TTAA sites, their spatial distribution on the gene, variations in insertions across sites within the same gene, and baseline background insertions (fig. S4A). The proportion of genes with read counts above background increases with the number of TTAA sites per gene, with no significant change observed with increase in TTAA sites beyond five [5 consecutive TTAA sites on average covers 758 base pairs (bp)], indicating that essentiality can be called with high confidence for these 4021 genes (fig. S4, B and C). To develop an improved gene essentiality model, the hybrid model score (HMS), we combined a Bayesian mutagenesis score (BMS) with a modified version (MMIS) of the mutagenesis index score (MIS) (*11*). The BMS accounts for the factors listed above for high-confidence gene essentiality calls with ≥5 TTAA, whereas the MMIS is more sensitive for short genes with <5 TTAA sites (figs. S4A and S5). The HMS model softly transitions between the BMS and MMIS based on the number of TTAA sites within a gene to provide the most accurate score for a given gene (fig. S4A). To make essentiality calls using HMS, we identified two cutoff points, corresponding to the transition points ("knee-points") of the HMS curve in terms of ranked-order genes. The cutoffs are HMS < 0.26 for essential genes and HMS > 0.88 for dispensable genes (Fig. 2, A and B, and fig. S6A). Genes in between cannot be confidently called by HMS. We curated a set of gold plus essential and dispensable genes based on conserved essentiality calls from previous studies in *P. falciparum*, *Plasmodium berghei*, and *Toxoplasma gondii* using conservative criteria defined in the Materials and methods section and data S1 (*11*, *13*, *14*). We examined the classification of the gold plus essential genes and found these genes to be called with 86.4% and 80.5% accuracy, respectively, which provides support for the HMS model with these thresholds (fig. S6, B and C, and data S1). We developed a second model, occupancy index score (OIS), which leverages information from independent transfection pools (based on IIEs rather than total read counts per a site across transfections used in HMS—i.e., the number of independent insertions across transfections rather than the relative abundance of all samples combined) that is particularly useful for identifying essential genes with few TTAA sites or genes that are truncatable (figs. S5 and S6D). OIS classified 94.7% and 70.1% of predicted essential and dispensable genes accurately, respectively (fig. S6, E and F). Though genes classified as dispensable by OIS are done so with high confidence, it also classifies a higher number of dispensable genes falsely as essential (fig. S6, D to F). By contrast, HMS provides a conservative score where genes with insufficient data receive an intermediate score to prevent false calls; we therefore used HMS for large-scale genome-wide analyses (fig. S6, A to C). Of 14 genes that have previously been genetically validated in *P. knowlesi*, 12 were classified as expected by the HMS model, 1 gene was classified as intermediate (ARP, PKNH_0515300, HMS = 0.264, expected to be essential), and 1 gene (GAMA, PKNH_1322900, HMS = 0.21) was classified as essential in our screen but dispensable in a previous study using a different *P. knowlesi* strain and may reflect strain-specific invasion ligand dependencies (fig. S6G and data S1) (*5*, *12*, *15*–*20*). As expected, all but one of the background genes, with high-confidence essentiality in other apicomplexans, were considered essential by HMS (the exception is NEK1, PKNH_1447700, with a HMS near the threshold, 0.3) (fig. S6H). A comparison of each model can be found in figs. S5 and S6.

Overall, 2037 genes (38.68%) with at least one TTAA site were classified as essential, 2124 (40.34%) as dispensable, and 1105 (20.98%) as indeterminate owing to an intermediate HMS (Fig. 2C and data S3). In total, 138 genes (2.62%) could not be classified owing to a lack of TTAA sites, and 579 essential genes with <5 TTAA sites in the gene were classified with reduced confidence (data S3). We investigated whether genome amplifications were present that may produce false dispensable calls. Using whole-genome sequencing and qPCR, we identified two amplified regions, including 34 genes that were all classified as dispensable and were excluded from further analysis (fig. S7 and data S3). Taken together, the number of essential genes is similar to previous observations in apicomplexans. The proportion of the genome classified as essential is higher than many free-living organisms, reflecting the highly specialized parasitic lifestyle of *Plasmodium* spp., which relies heavily on host metabolism (*13*). To make the data readily accessible to the community, we provide a web-based app that can be used interactively to examine HMS and OIS scores associated with gene essentiality (https://umbibio.math.umb.edu/PkEssenDB/); this information is also available through VEuPathDB (*21*).

To further identify noncoding regions of the genome that are intractable to mutagenesis, we trained a second HMM to identify stretches of ≥5 TTAAs with reduced insertions (Fig. 1A). The majority of these insertion-depleted regions overlapped with genic regions (87.7% total); however, we also identified 13.3% of depleted regions with no genic overlap. To determine whether the remaining depleted regions were attributable to essential long noncoding RNAs (lncRNAs), which are increasingly being recognized as important for *Plasmodium* spp. biology, we performed total RNA sequencing (RNA-seq) on asynchronous blood-stage *P. knowlesi* and applied the HMS model to determine their essentiality (*22*). This resulted in the identification of 864 lncRNAs, 794 of which have at least one TTAA and 55.8% of which have ≥5 TTAAs in their cognate DNA sequences, of which 53 (6.6%) were essential by HMS (Fig. 2D, fig. S8, and data S4). Only 13 essential lncRNAs overlap with CDS, and therefore lncRNAs do not significantly affect gene essentiality calls on a global scale (fig. S8A and data S4). The RNA-seq data were also used to identify potential discrepancies in *P. knowlesi* gene models or alternative isoforms. We identified 892 genes that putatively have alternative isoforms, including 90 genes with a discrepancy in essentiality between the annotated gene and the alternative isoform or gene model, which suggests either biological differences or incorrect gene models, that will require further validation (data S4). We considered whether bias against insertion in untranslated regions (UTRs) of essential genes contributed to insertionally depleted regions and found reduced insertions, which suggests that insertions proximal to coding sequences disrupt expression (fig. S9, A and B). As seen in mammalian cells, there is a slight increase in insertions observed in the 5′UTR of dispensable genes (fig. S9, A and B) (*23*, *24*). The remaining 2.4% of insertionally depleted regions that do not overlap with genic regions, UTRs (1000 bp proximal to CDS), or lncRNAs could be due to heterochromatin, overlap with other ncRNAs, or missing gene annotations.

### Identification of low-fitness and truncation mutants explains many intermediate genes

We reasoned that additional features relating to the amount and distribution of the TTAA sites within a gene could contribute to genes having an intermediate HMS. Indeed, these genes on average contained fewer TTAA sites than genes with HMS < 0.26 or HMS > 0.88 (fig. S9C). Determination of uneven insertion across genes can be used to identify dispensable domains within larger essential genes, including multidomain proteins. To this end, a truncation model was developed to identify uneven insertion across the gene. This was achieved by fitting a step function to the binarized mutability of sites across the gene (BMS model), with goodness of the fit calculated by mean squared error (MSE) (figs. S4A and S10). The majority of unevenly high insertions occurred in the first and last 10% of the gene (Fig. 2E and fig. S10, A and B; normalized CDS 0 to 0.1 and 0.9 to 1). To increase the likelihood of genes classified as truncatable being biologically informative, we limited further analysis to genes with >10% of gene length disrupted from the 5′ or 3′ end. In total, 122 truncatable genes were identified

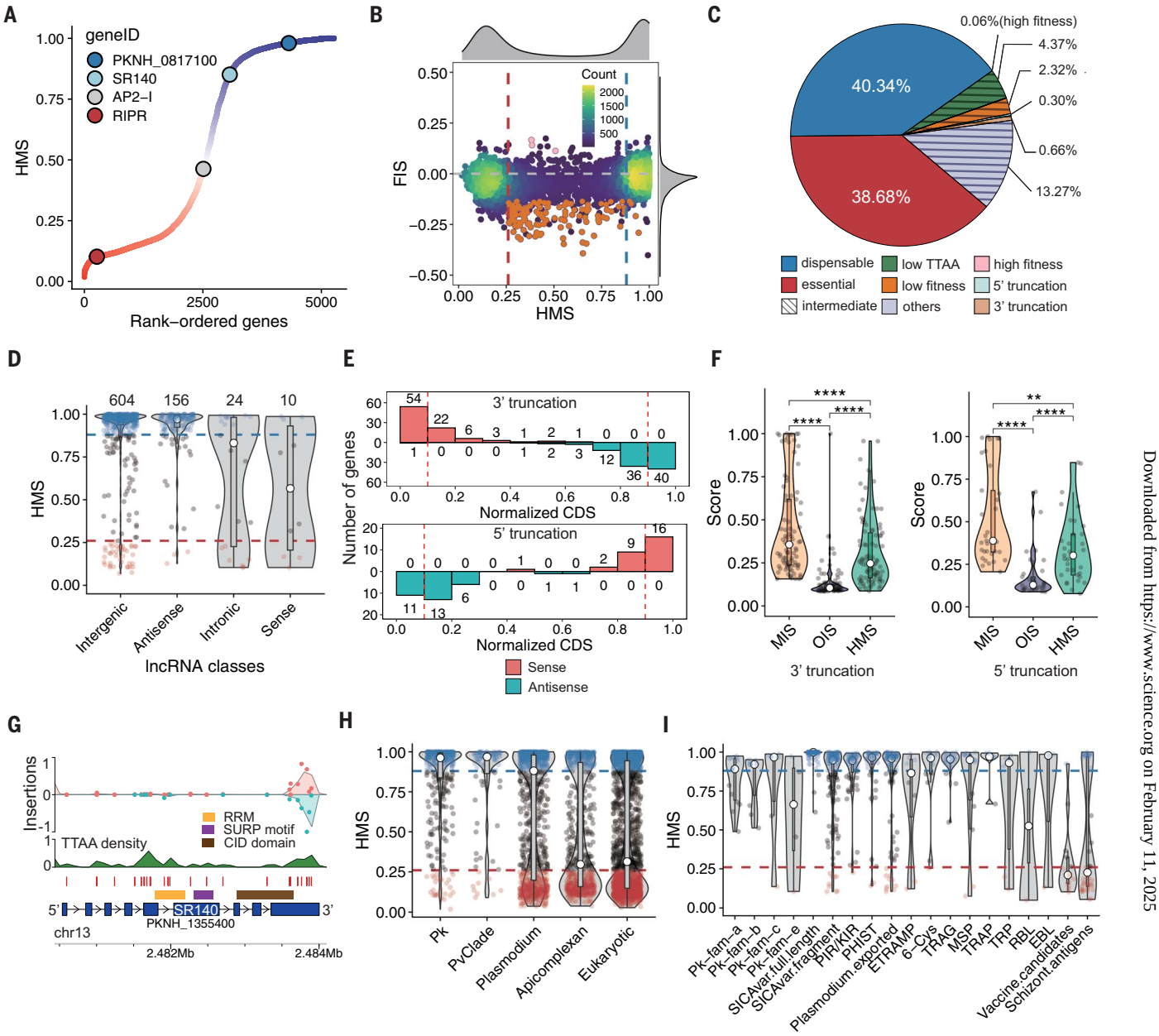



**Fig. 2. Defining the essentiality of the genome for proliferation in RBCs.** (**A**) Distribution of all genes by rank order of HMS. Genes that are more likely to be essential have a HMS closer to 0. Genes more likely to be dispensable have a HMS closer to 1. (**B**) Distribution of genes by HMS and the FIS. Genes with a HMS < 0.26 (red dashed line) are considered essential, and genes with a HMS > 0.88 (blue dashed line) are considered dispensable. Genes with a significantly reduced fitness are highlighted in orange, and genes with increased fitness are highlighted in pink. (**C**) Percentage of genes that are essential, dispensable, or intermediate (HMS > 0.26 and < 0.88). Intermediate HMS genes with <5 TTAA sites (short), significantly reduced FIS (low fitness), significantly increased FIS (high fitness), and genes that are truncatable from the 5′ or 3′ end are shown. Genes with zero TTAA sites are not included. (**D**) HMS distribution of lncRNAs identified in this study. Sense and antisense refer to lncRNAs that overlap with the CDS in the sense and antisense orientation. lncRNAs with a HMS < 0.26 (red dashed line) are considered essential, and lncRNAs with a HMS > 0.88 (blue dashed line) are considered dispensable. (**E**) Number of genes with varying lengths of truncation from the 5′ and 3′ end of the gene (normalized CDS to 0 and 1), respectively. Only genes with a truncation in the normalized CDS between 0.1 and 0.9 (>10% of the gene; red dashed lines) are considered significant. (**F**) Distribution of the essentiality scores of truncatable genes by HMS and OIS have fewer false classification of these genes as dispensable compared with MIS. (**G**) SR140 is an essential but truncatable gene. The lack of insertions in the 5′ end of the gene, which includes an RNA recognition motif (RRM), a SURP motif, and a CID domain, suggests that these domains are essential for function, whereas the 3′ end of the gene contains insertions, suggesting that it is dispensable. (**H**) The essentiality of *P. knowlesi* genes separated by the presence or absence of orthologs in other organisms. Orthologs with conservation across greater evolutionary distance are enriched for essential genes. (**I**) The essentiality of *P. knowlesi* genes involved in host-parasite interactions. For (D), (F), (H), and (I), white circles indicate medians. The bars show the interquartile ranges (IQRs), and the whiskers extend to values within 1.5 times the IQR. ****$P$ < 0.0001; **$P$ < 0.01.

using a 25% cutoff on MSE (89 3′ truncations and 33 5′ truncations). Of the 122 high-confidence truncatable genes, all but two were classified as intermediate or essential by HMS, verifying that the HMS model prevents false dispensable calls from insertion in only part of the gene to which the MIS is prone (Fig. 2F and fig. S10C). The OIS model classified most of these correctly as essential (Fig. 2F). These include known examples of essential but 3′ truncatable genes from *P. falciparum*, such as PTEX150 (*25*), and previously unknown examples, such as U2 small nuclear ribonucleoprotein (snRNP)–associated SURP motif–containing protein (SR140, PKNH_1355400), which has a truncatable 3′ end directly after all annotated domains (Fig. 2G and fig. S10, D to H). A total of 33 genes contained insertions in the 5′ of the CDS, for which there was a strong bias in the orientation of transposon insertion, suggesting that there is promoter activity within the transposon that can drive expression of a 3′ fragment from an alternative start site in a small subset of genes (Fig. 2E and fig. S10, I and J). A more relaxed MSE cutoff of 75% can be used to identify truncatable genes with more complex insertional patterns, such as AP2-I (PKNH_0806500), but may include false calls as seen by the inclusion of dispensable genes (fig. S10, A, B, and H). To identify intermediate genes with fitness defects, we introduced a fitness index score (FIS) that tracks changes in relative abundance of each mutant over three time points (days 9, 14, and 19 posttransfection) (fig. S11, A to C). Overall, 126 genes were identified as having a significantly reduced fitness (excluding genes with HMS < 0.26), of which 91% had an intermediate HMS (Fig. 2, B and C, and fig. S9E). This is fewer than was observed in *P. berghei*, which may be due to mutants with large fitness defects being lost by day 9 in our screen (compared with measurements from days 4 to 8 in *P. berghei*), before which accurate quantification of mutants is difficult owing to high background from plasmid DNA, and is supported by an enrichment of *P. berghei* slow orthologs being classified as essential in our screen (fig. S11D) (*13*). Overall, ~36% genes with an intermediate HMS can be explained by genes that are truncatable, have few (<5) TTAA sites, or are of intermediate fitness (Fig. 2C and fig. S11E). The remaining genes may have an intermediate HMS for biological or technical reasons, including subtle growth effects, being able to tolerate insertion in some sites, having reduced chromosomal accessibility, or lacking sufficient data to make an accurate call owing to the random nature of transposon insertion.

### *P. knowlesi* genes involved in host-parasite interactions are mostly dispensable in vitro

In agreement with observations in other organisms, a greater proportion of *P. knowlesi* genes that have orthologs across greater evolutionary distance are essential. *P. knowlesi* genes, with orthologs in apicomplexans or eukaryotes, contain 46.4% and 46.5% essential genes, respectively, whereas genes found only in *P. knowlesi*, the *P. vivax* clade, and *Plasmodium* spp. contain 6.3%, 9%, and 29.8% essential genes, respectively (Fig. 2H, fig. S12A, and data S5). Genes with syntenic orthologs between *Plasmodium* spp. are further enriched for essential genes compared with those with nonsyntenic orthologs, whereas genes with paralogs are more likely to be dispensable, presumably because of their redundant or diversified functions in the parasite (fig. S12, B and C, and data S5). Of the 400 essential *Plasmodium*-specific genes, 52.5% are annotated as conserved proteins of unknown function. Essential genes conserved across Apicomplexa and higher eukaryotes were enriched for those involved in ribosome biogenesis, RNA metabolism, and transcription (fig. S12D). The essentiality of *P. knowlesi* genes was mapped onto the putative subcellular localization based on the hyperLOPIT data of their orthologs in the apicomplexan *T. gondii* (*26*). Organelles involved in host-parasite interactions, including apical, microneme, rhoptry, and dense granule genes, contained fewer essential genes, whereas protein trafficking, mitochondrial, apicoplast, nuclear, proteasomal, and ribosomal genes contained a high proportion of essential genes, consistent with previous studies, which highlighted these as promising drug targets (fig. S13, A and B).

Genes specific to *P. knowlesi* and the *P. vivax* clade predominantly belong to large gene families encoding proteins involved in host-parasite interactions (Fig. 2I and data S5 and S6). The majority of proteins exported into the host RBC, including all full-length SICAvar genes, are dispensable in vitro, which is expected because they are primarily involved in immune evasion (Fig. 2I, fig. S12D, and data S6). Notably, 12 of 300 exported genes are essential, including one PIR/KIR (PKNH_1454600) and two PHIST genes (PKNH_0201300 and PKNH_0117900), and nine genes annotated as "*Plasmodium* exported protein, unknown function," which suggests essential functions under in vitro culture conditions (Fig. 2I and data S6). The lncRNAs overlapping with CDS are biased toward antisense transcripts and are associated with nonessential genes belonging to multigene families (data S4). Among genes that encode for zoite proteins (merozoite, sporozoite, and ookinete forms) involved in host cell invasion (TRAP, TRP, MSP, EBL, RBL, TRAG, and 6-Cys families), few are essential; this is presumably due to functional redundancy (as evident in *P. falciparum*) (*27*), roles in invasion of other life cycle stages (e.g., CSP and celTOS, although celTOS was proposed to be essential in a *P. falciparum* piggyBac screen) (*11*, *28*), or membership in multigene families, such as TRAG proteins (*29*–*31*) (Fig. 2I, fig. S13C, and data S6). Most validated and putative vaccine antigens from *P. falciparum* and *P. vivax* were essential in *P. knowlesi*, including MSP1, AMA1, GAMA, SRA, all RON proteins, the CLAMP complex, and PCRCR complex proteins (which in non-Laveranian parasites lack an obvious Rh5 ortholog and have been genetically validated as essential in *P. knowlesi*) (Fig. 2I, fig. S13C, and data S6) (*12*). There are several *P. knowlesi*–specific differences in essentiality, including RON1 and MSP5, which are uniquely essential in *P. knowlesi*, and MA and MSP4, which are uniquely dispensable in *P. knowlesi* (fig. S13C and data S6). Within the EBL and RBL families that have divergent orthologs in *P. knowlesi*, only NBPXb is essential; however, we were unable to verify the essentiality of DBPα or DBPγ because they are duplicated and deleted in the YH1 strain, respectively (fig. S13C) (*32*). We curated an extended list of 38 schizont proteins with potential as vaccine antigens and found 17 to be essential, including the recently identified prohibitin 1 (*33*) (Fig. 2I, fig. S13D, and data S6). This information can help prioritize vaccine targets for *P. vivax* clade parasites, which, because of the unculturability of *P. vivax* in vitro, have little to no validation from previous studies.

### Comparative analysis predicts drug susceptibility in *Plasmodium* spp.

We compared our essentiality data for *P. knowlesi* with previous essentialome studies of apicomplexan parasites, including *P. falciparum*, *P. berghei*, and *T. gondii*, and found a high level of overall agreement, with *P. berghei* having the highest concurrence with the *P. knowlesi* data and *T. gondii* the least, likely due to a combination of technical differences between the screens (gene disruption method, culture conditions, and scoring methods) and evolutionary distance (Fig. 3A and fig. S14A) (*11*, *13*, *14*). Conservative criteria were used to identify high-confidence conserved essential or dispensable genes across *Plasmodium* spp. (Materials and methods). Of the 691 genes with 1:1 orthologs and high-confidence essentiality calls across all three species, 225 (32.6%) were dispensable and 329 (47.6%) were essential in all species, whereas 137 (19.8%) had discrepant essentiality in at least one species (Fig. 3B and data S7). Pairwise comparisons show that the *P. falciparum* piggyBac screen has the highest level of discrepant genes with *P. knowlesi* and *P. berghei* (16.9% and 14.4% discrepant calls, respectively) compared with *P. knowlesi* and *P. berghei* (7% discrepant calls), which may be due to divergent biology or the lower saturation of the *P. falciparum* screen (fig. S14, B and C). Conserved essential genes between *Plasmodium* spp. were enriched for central metabolic pathways, such as glycolytic processes (Fig. 3C). We also investigated a larger set of 1238 orthologs with lower-confidence essentiality calls

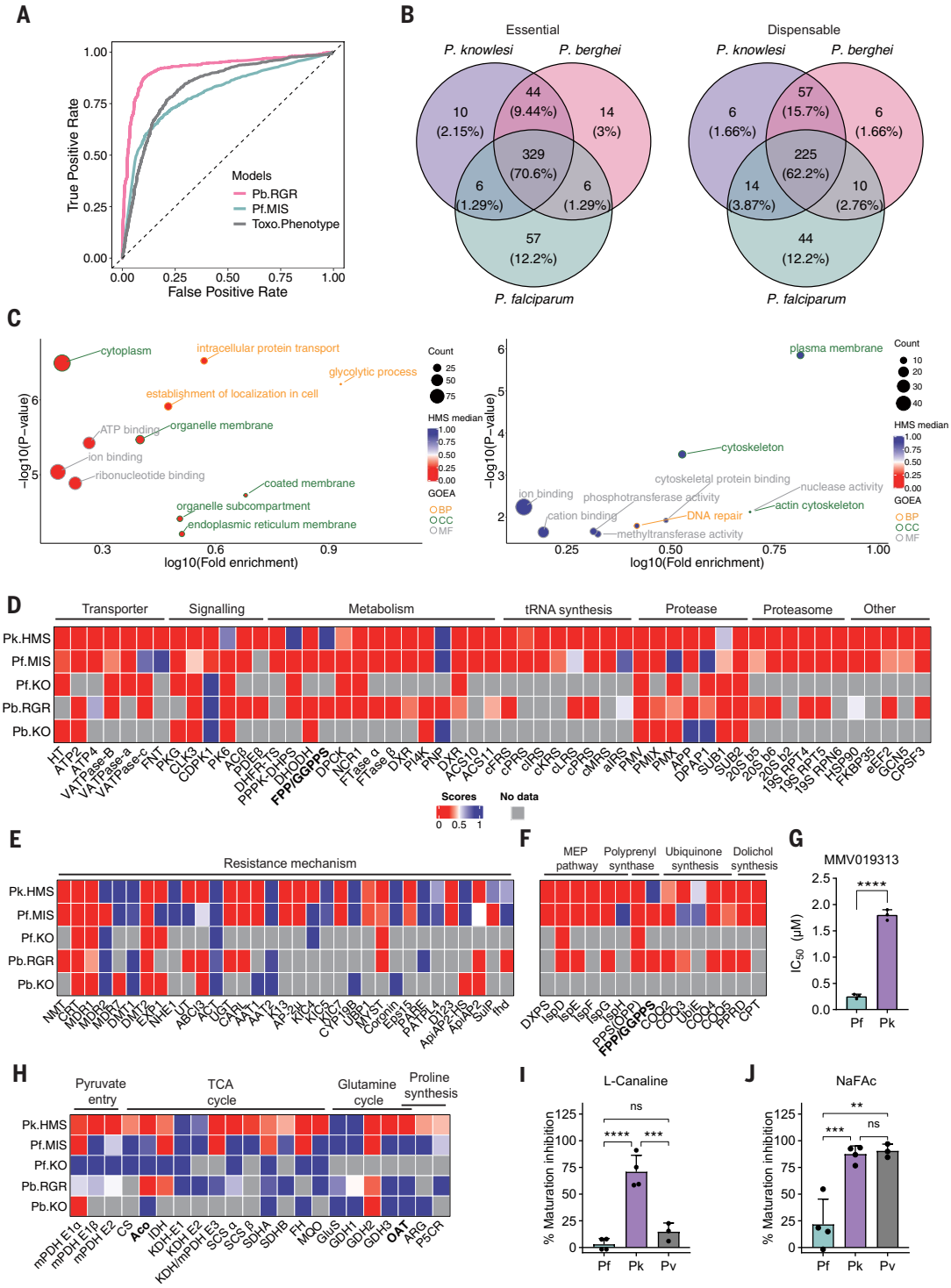



**Fig. 3. A comparative analysis of gene essentiality across apicomplexan parasites.** (**A**) Receiver operating characteristic (ROC) curve quantifying the predictive power of the *P. knowlesi* essentiality analysis for classifying genes from other high-throughput screens. HMS highly correlates with previous apicomplexan essentiality screens. (**B**) Venn diagrams of conserved and species-specific essential (left) and dispensable (right) genes with 1:1 orthologs and high-confidence essentiality calls between *Plasmodium* spp. (**C**) GOEA of conserved essential (left) and dispensable (right) *Plasmodium* genes at the intersection identified in (B). The inner color of each point is the median HMS. The outer color and text color show the GOEA category: biological process (BP), brown; cellular component (CC), green; and molecular function (MF), gray. (**D** and **E**) Heatmaps of essentiality scores of validated *Plasmodium* drug targets (D) and resistance mechanisms (E). Genes studied further are bolded in the heatmaps. The scale for all heatmaps is shown on (D). (**F**) Heatmap of essentiality and fitness scores of genes up- and downstream of FPP/GGPPS function in isoprenoid biosynthesis and utilization. Color scheme matches that of (D). (**G**) $IC_{50}$ of the FPP/GGPPS inhibitor, MMV019313, for *P. falciparum* (Pf) and *P. knowlesi* (Pk) in a 1.5-cycle growth inhibition assay. Three independent replicates were performed in technical duplicate. Means and SDs are shown. Statistic shown is an unpaired *t* test (*****P* < 0.0001). (**H**) Heatmap showing essentiality and fitness of TCA cycle genes and their orthologs in different *Plasmodium* spp. Color scheme matches that of (D). (**I** and **J**) Percentage inhibition of schizont maturation for *P. falciparum* (Pf), *P. knowlesi* (Pk), and *P. vivax* (Pv) by the aconitase inhibitor sodium fluoroacetate at 2 mM (NaFAc) (I) and the OAT inhibitor L-canaline at 8.2 μM (J) is shown. Three independent replicates were performed in technical duplicate. Means and SDs are shown. Statistic shown is ANOVA with multiple comparisons (adjusted ***P* < 0.005; ****P* < 0.001; ns, not significant). For all panels, Pf.MIS data are from the *P. falciparum* piggyBac screen (*11*). Pb.RGR data are from the large-scale screen in *P. berghei* (*13*). Toxo.Phenotype data are from the *T. gondii* CRISPR-Cas9 screen (*14*). Pf.KO and Pb.KO refer to targeted gene disruptions in *P. falciparum* (Pf) and *P. berghei* (Pb) from the literature and from phenoplasm.org (*90*).

in each species. As expected, this group contained a greater number of discrepent genes, and the enriched pathways were similar between the two gene sets (fig. S15 and data S7).

We investigated the cross-species essentiality of 54 previously validated drug targets and 32 known resistance proteins curated from literature (Fig. 3, D and E, and data S8). The majority of validated drug targets were essential across all species, whereas fewer resistance mechanism genes were essential (87% and 53% have HMS < 0.26, respectively) (data S8). Only six genes were

uniquely dispensable in *P. knowlesi* from the high-confidence set of 691 1:1 *Plasmodium* spp. orthologs and had no significant fitness defect in *P. knowlesi* (data S7). This included hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase (PPPK-DHPS, PKNH_1429900) and bifunctional farnesyl/geranylgeranyl diphosphate synthase (FPP/GGPPS, PKNH_0926500). Both genes contained insertions throughout the gene, and we confirmed the gene model using RNA-seq data and that they are not amplified in the genome by qPCR and using existing whole-genome sequencing data (figs. S7 and S16). It is possible that the high folate in the culture medium bypasses the essential function of PPPK-DHPS, although the conditions are the same as in the *P. falciparum* piggyBac study that found PPPK-DHPS to be essential, or alternatively that growth in rhesus macaque RBCs alters their metabolic requirements. The dispensability of FPP/GGPPS, which is listed as a high-priority drug target (*34*), is not explained by altered essentiality of the up- or downstream pathways (Fig. 3F and data S8). Drug susceptibility testing with the *Pf*FPP/GGPPS inhibitor, MMV019313, exhibited a 9.5-fold higher median inhibitory concentration ($IC_{50}$) against *P. knowlesi* compared with *P. falciparum*, which has been suggested to have a secondary target at similar higher concentrations in *P. falciparum*, supporting the observed dispensability of *Pk*FPP/GGPPS (Fig. 3G) (*35*).

Notably, the 10 genes uniquely essential in *P. knowlesi* included two, ornithine aminotransferase (OAT) and glutamate dehydrogenase 3 (GDH3), that bidirectionally convert glutamate to α-ketoglutarate (Fig. 3H and data S7 and S8). These metabolic intermediates, along with glycolysis-derived pyruvate, participate in anaplerosis to replenish the tricarboxylic acid (TCA) cycle of *Plasmodium* spp. OAT essentiality could result from its role in the proline synthesis pathway, which includes other genes with intermediate HMS (Fig. 3H) (*36*). To directly test the dependency of *P. knowlesi* and *P. vivax* on OAT function, the OAT inhibitor L-canaline was used. It was active against *P. knowlesi* but not *P. falciparum* or *P. vivax* (Fig. 3I and fig. S14E). We then investigated the essentiality of other TCA cycle genes in *P. knowlesi*, given that OAT and GDH3 feed metabolic intermediates into the TCA cycle in other *Plasmodium* spp. Notably, all TCA cycle enzymes except for KDH are essential or have significantly reduced FIS (Fig. 3H, fig. S14F, and data S8). The enzymes mPDH and KDH, which govern pyruvate entry into the TCA cycle, are redundant in *P. falciparum*; however, mPDH is essential in *P. knowlesi* (Fig. 3H) (*37*). To assess the dependence of *P. knowlesi*, *P. vivax*, and *P. falciparum* on the TCA cycle, we inhibited a key step catalyzed by aconitase using the inhibitor sodium fluoroacetate. This aconitase inhibitor was found to be inhibitory against both *P. knowlesi* and *P. vivax* but not *P. falciparum* (Fig. 3J and fig. S14D). Taken together, our data suggest that there is an increased reliance on the TCA cycle in *P. vivax* clade asexual-stage parasites but with different metabolite utilization requirements. Empirically, we hypothesize that any TCA cycle inhibitors being pursued as transmission-blocking drugs for *P. falciparum* may be active against asexual *P. vivax* clade parasites (*38*).

### Perturbation analysis reveals determinants of antimalarial susceptibility

We next investigated the utility of our high-saturation transposon mutagenesis in *P. knowlesi* to identify genes that modulate sensitivity to known antimalarials through unbiased genome-wide perturbation screens (Fig. 4A; fig. S17, A to D; and data S9). GNF179 is a close analog of ganaplacide (KAF156), which has shown positive results in a phase 2 clinical trial (*39*). After selection of the transposon library with GNF179, two statistical models, EdgeR and a site-level model, were used to identify genes with significantly changed numbers of insertions (*40*). Genes were considered hits only when either model classified them as statistically significant. We observed a significant enrichment (~1000-fold) for transposon mutants in one gene, acetyl-CoA transporter (ACT, PKNH_0621300), indicating that its disruption results in resistance (Fig. 4B, fig. S17E, and data S9). Disruption of ACT is known to induce GNF179 resistance in *P. falciparum* (*41*). A limitation of transposon perturbation screens is that the lack of insertions in essential genes prevents their study, including drug targets, with the possible exception of those in UTRs that alter expression. This is observed with the other known mediators of GNF179 resistance in *P. falciparum*, cyclic amine resistance locus (CARL, PKNH_0818900) and UDP-galactose transporter (UGT, PKNH_0910900), which were both essential in our screen, and therefore their role in GNF179 resistance could not be determined (*41*–*43*). Two genes (PKNH_0722900 and PKNH_0929900) were significantly depleted, both of which are unique to *Plasmodium* spp. and have no predicted domains. However, because these genes are not essential, they are unlikely to be the target but instead are modulators of GNF179 sensitivity that may provide pathways to better understanding the mechanism of action of GNF179 (Fig. 4B, fig. S17E, and data S9).

The mechanism of action and resistance of artemisinin (ART), the mainstay of antimalarial treatment for *P. falciparum* globally and increasingly for *P. vivax*, are highly complex (*3*, *44*, *45*). Perturbation of *P. knowlesi* with dihydroartemisinin (DHA) revealed 10 genes with significantly altered levels of insertions, six enriched and four depleted. These included orthologs of genes known to influence ART susceptibility as well as new genes (Fig. 4C, fig. S17F, and data S9). We compared our hits with previous *P. falciparum* DHA transposon perturbation studies; however, these genes were not included in the smaller number of mutants used in those screens (*46*, *47*). Among 21 genetically validated genes that modulate DHA sensitivity in *P. falciparum*, only six were non-essential (data S9). Although some of these genes displayed changes in line with their role in *P. falciparum* ART sensitivity, only *P. knowlesi* knowpain-3 (PKNH_0912900, the closest ortholog of *P. falciparum* falcipain-3) was significantly enriched in our screen, in support of a shared role of hemoglobin digestion in DHA activity between *P. knowlesi* and *P. falciparum* (data S9) (*48*–*50*). Multiple genes with no previous link to ART sensitivity were also identified, including three genes with enriched insertions—a phospholipid or glycerol acyltransferase (PGAT, PKNH_0712100), a protein of unknown function (PKNH_0830900), and protein transport protein SEC16 (SEC16, PKNH_0917600)—along with four genes with depleted insertions—vacuolar fusion protein CCZ1 (CCZ1, PKNH__0904800), a protein of unknown function (PKNH_0840900), F-box protein (FBX07, PKNH_1008000), and patatin-like phospholipase 1 (PATPL1, PKNH_0412000) (Fig. 4C, fig. S17F, and data S9).

Two of the most highly enriched genes, type II NADH:ubiquinone oxidoreductase (NDH2, PKNH_0713000) and riboflavin kinase (RFK, PKNH_1112200), have functions linked to mitochondrial metabolism. NDH2 is one of five mitochondrial dehydrogenases that form the electron transport chain (ETC) to produce reduced ubiquinone and is not essential in *P. falciparum* or *P. berghei*, but it has an intermediate HMS score in *P. knowlesi* (0.64) (*51*, *52*). Notably, NDH2 in *P. falciparum* is up-regulated in Kelch 13 (K13) DHA-resistant mutants, and down-regulation of the ETC has been associated with increased resistance to artemisinins in yeast and *T. gondii* (*53*–*55*). RFK is necessary for the conversion of riboflavin (vitamin $B_2$) to FMN and FAD, which are required for multiple pathways, including oxidative-reduction processes and energy metabolism in the mitochondria. We observed inhibitory activity of the RFK inhibitor roseoflavin against *P. knowlesi* at a similar concentration to that of *P. falciparum* (fig. S17G) (*56*). This inhibitor had an antagonistic effect on DHA activity, confirming the increased fitness of RFK mutants observed after DHA selection and highlighting a potential previously unidentified resistance mechanism (Fig. 4D).

### Conclusions

We have made a comprehensive classification of gene essentiality in a *Plasmodium* spp. malaria parasite during blood-stage growth. The genes that we have identified present promising targets for synergistic inhibition with DHA and other antimalarials, a strategy that is being actively pursued (*57*–*66*). The high efficiency of

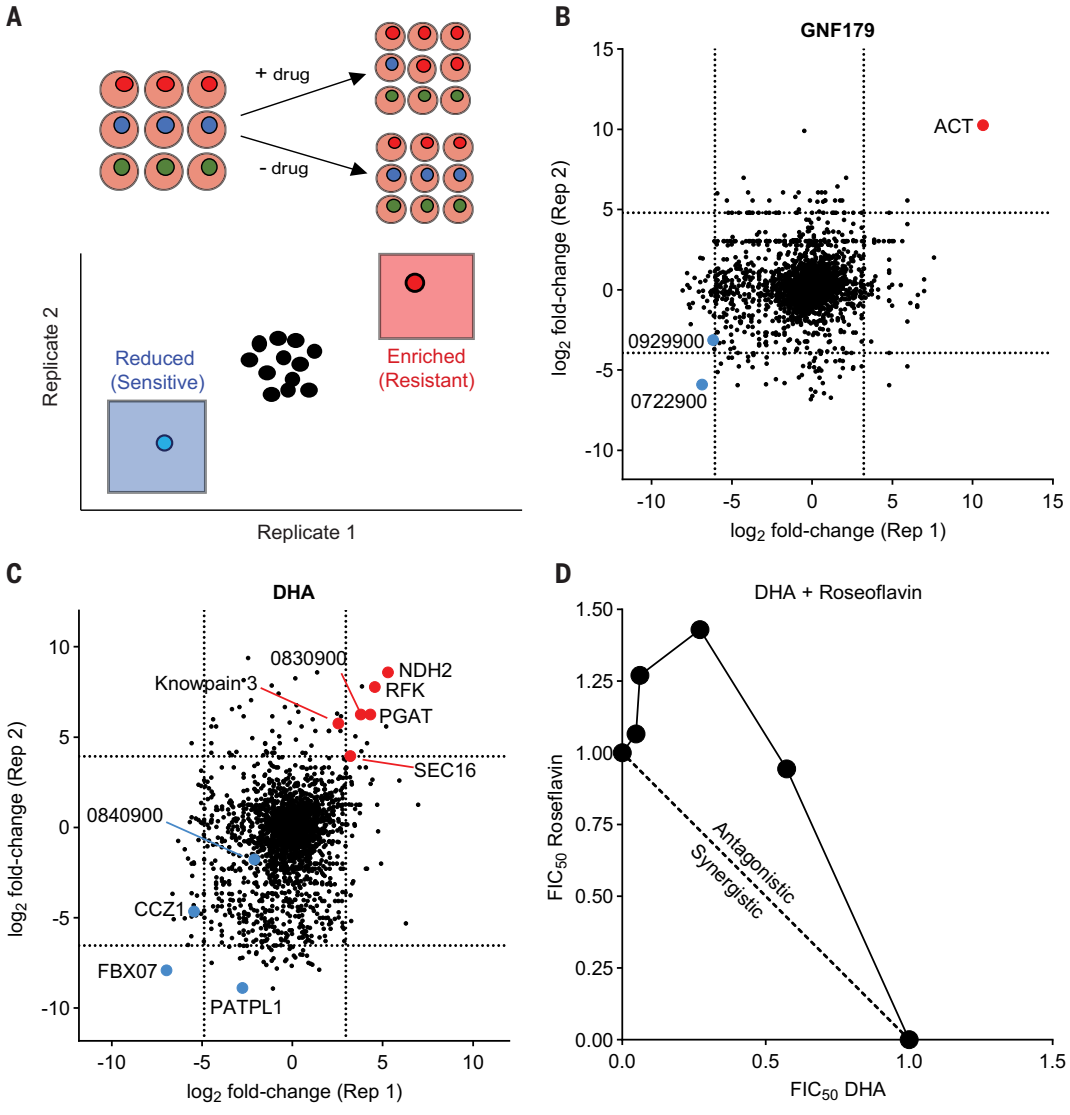



**Fig. 4. Perturbation analysis reveals previously unknown mechanisms of drug action and resistance.** (**A**) Schematic of *P. knowlesi* perturbation screens. Two transposon mutant pools were generated by combining the original transposon pools (replicate 1 and replicate 2 that each originate from independent transfections). Each replicate was exposed to either GNF or DHA (fig. S17) and compared with an untreated control. After drug selection, the relative abundance of each mutant was compared with the untreated control to determine fold-change and statistical significance (Materials and methods). Enrichment of transposon insertions in specific genes implicate them in drug resistance, whereas reduced insertions indicate drug sensitization. (**B** and **C**) Change in relative abundance of transposon mutants of all genes in the library by the EdgeR model after GNF179 (B) or DHA (C) selection. Genes with transposons that are enriched (resistant) are marked in red, and genes that are depleted (sensitized) are marked in blue (see also fig. S17 for the site-level model). (**D**) The fractional inhibitory concentration (FIC) of the riboflavin kinase inhibitor (RFK), roseoflavin, shows antagonism with DHA in *P. knowlesi*. The dashed line represents the theoretical values expected when the drug interaction is additive. Values above the line show that the interaction is antagonistic, and those below the line show that the interaction is synergistic.

transfection of the *P. knowlesi* in vitro culture system will make genome-wide screens more accessible to the malaria community, including using modified transposon strategies, such as outward-facing promoters that alter expression of essential genes in perturbation screens. The models developed in this work can also be applied to *P. falciparum* and other organisms if high-density insertional saturation can be achieved. Comparative studies and genetic perturbation screens in parasites of the *P. vivax* clade will aid prioritization of drug and vaccine targets and reveal alternative mechanisms of resistance that will inform therapeutic development and surveillance.

## Materials and methods

### *Parasite culture*

The *P. knowlesi* YH1 strain used was generated in (*6*). Parasites were maintained in rhesus (*Macaca mulatta*) RBCs provided from Yerkes National Primate Research Center, Emory University. Parasites were cultured in RPMI-1640 media supplemented with 25 mM HEPES, 11.50 mg/l hypoxanthine, 2.42 mM sodium bicarbonate, and 4.31 mg/ml AlbuMAX II (Invitrogen), at 37°C in a 1% oxygen, 5% carbon dioxide and 94% nitrogen environment.

### *Plasmid construction*

Plasmids were synthesized by Twist Bioscience using their custom ampicillin resistance plasmid backbone. The Pk3xNLS-HypPB plasmid, to transiently express hyperactive piggyBac transposase, was generated using (in order 5′ to 3′), 1574 bp of the 5′UTR from CYP19A (PKNH_0818800), 3x nuclear localization signals (NLS), a linker, hyPBase (*9*), 3xHA tag and 600 bp of the 3′UTR of HSP70 (PKNH_1312700). The PkPfRh3HR-BeRFP-hDHFR transposon plasmid, which contains the transposon and selectable marker, was generated using (in order 5′ to 3′), 5′Homolgous recombination (HR) region PfRh3 (not used in this study), Sbf1 and NotI sites flanking the transposon on both sides (to reduce intact plasmid DNA in library prep), 5′ piggyBac inverted terminal repeat, 1056 bp of the 5′UTR CPN10 (PKNH_1434800), mBeRFP fluorescent protein (not used in this study), a T2A skip peptide, hDHFR selectable marker, 500 bp of the 3′UTR of BIP (PKNH_0715900), 3′ piggyBac inverted termal repeat and a 3′Homolgous recombination (HR) region PfRh3. UTRs include a small number of mutations to aid synthesis. Sequences of each component and full plasmids can be found in data S10.

### *Transfection and pooling*

Cotransfection of the transposase and transposon plasmids into *P. knowlesi* YH1 was performed using a modified protocol from (*5*) using the Amaxa 4D nucleofector (Lonza) and the P3 Primary Cell 4D-Nucleofector X Kit L

(Lonza, V4XP-3024). Briefly, mature schizonts were purified on a 60% percoll gradient (900 × g for 15 min), washed with complete RPMI and allowed to recover in RPMI at 37°C in a 1% oxygen, 5% carbon dioxide and 94% nitrogen environment for ~1 hour. Purified schizonts were pelleted and all RPMI was removed. 7.5 μL of each plasmid (7.5 μg of each plasmid) were combined with 100 μL of complete P3 primary cell solution and 20 μL of parasites (~80 to 90% purity, ~2 × $10^8$ cells). This mixture was transferred to the supplied cuvette and electroporated using the FP158 program. Parasites were transferred to a 1.4 mL tube containing 1 ml complete RPMI plus 100 μL rhesus RBCs, and incubated at 37°C on a thermomixer, shaking at 600 rpm for ~20 to 30 min. Parasites were washed with complete RPMI and pooled with 2 to 5 transfections per 10 ml of culture (10 total pools). 100 nM pyrimethamine was added to select for the transposon after 1 day, and media replaced daily.

### *P. knowlesi* TPN sample and DNA preparation

*P. knowlesi* TPN mutant parasites were collected for gDNA prep, 400 to 600 μl of infected RBCs. The *P. knowlesi* TPN mutant blood pellets were lysed with 0.03% saponin followed by genomic DNA extraction using a QIAamp DNA Blood Mini Kit (QIAGEN). The gDNA was quantified using Qubit High Sensitivity assay according to the manufacturers protocol (Life Technologies Ltd, UK).

### *QIseq-based P. knowlesi* TPN DNA library preparation

A modified QIseq approach was used (*10*). For each TPN mutant library, up to 500 ng of gDNA was fragmented combined together with end Repair, dA-tailing and adaptor ligation using the NEBNext Ultra II FS DNA Library Prep kit. The adaptor cleavage for further multiplex library preparation was done using USER enzyme (NEB) at 37°C for 15 min. The adapter-ligated DNA libraries were cleaned using 0.8x volume Ampure XP (Beckman Coulter (UK) Ltd, UK). The piggyBac-genome junction was amplified through two nested PCR reactions (PCR1 and PCR2) using NEBNext Ultra II Q5 Master Mix. Both the nested PCR reactions amplified the 5′ end of the piggyBac transposable element sequence inserted within the genome. For PCR1, primers BE_582- 5′-GACGGATTCGCGCTATTTAGAAAGAGAG-3′ (priming within the 5′ transposon repeat element) and BE_583- 5′-ACACTCTTTCCCTACACGAC (priming with the NEB adaptor sequence) were used to amplify using conditions of annealing at 67°C and extension at 68°C with 14 cycles of PCR. The PCR1 product was purified using Qiagen PCR cleanup kit and further subjected to restriction digestion with enzymes NotI and SbfI to reduce episomal plasmid DNA (non-integrated transposon). The samples were again purified with Qiagen PCR cleanup kit and used for PCR2. The PCR2 conditions were as follows: BE_584- 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAATTTTACGCAGACTATCTTTCTAGGGTTAA-3′ (Rv primer; underlined seq specific to 5′ transposable element along with Illumina adapter) and BE_585-5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCTACACGACGCTCTTCCGATCT-3′ (Fw primer, underlined sequence specific to the NEB adapter along with Illumina adapter; 12 cycles of PCR; annealing and extension at 68°C). The final PCR product was purified and size selected using Ampure XP beads. The size selection was two-step with 0.6x volume to remove large DNA fragments, followed by 0.8x volume to remove small DNA fragments (<200 bp and residual PCR primers). The final PCR reaction was performed to add Illumina sequences and indexes to each sample using commercially available index primers (Nextera XT Index kit, Illumina). The final TPN amplicon product was purified with 0.9x volume Ampure XP beads. The library size and concentration were determined using Bioanalyzer high sensitivity DNA analysis (Agilent Technologies) and Qubit High Sensitivity assay according to the manufacturers protocol (Life Technologies Ltd, UK).

### *P. knowlesi* transposon library–based sequencing

To ensure quality control, all individual *P. knowlesi* TPN-based DNA libraries were pooled and sequenced using a MiSeq-Nano kit (Illumina) to estimate the correct concentration and size of the libraries. For deep coverage, the pooled libraries were then sequenced using the NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles, Illumina). Up to 35% of PhiX was spiked into the transposon library pooled samples to improve cluster generation.

### *P. knowlesi* lncRNA sequencing

*P. knowlesi* YH1 was synchronized using percoll gradient centrifugation method. The parasites were harvested and saved in TRIzol (ThermoFisher) at different stages of intraerythrocytic cycle: rings, trophozoites, and schizonts in the parasite concentration ratio of 2:2:1, with parasitemia between 4 and 5% grown at 2% hematocrit in rhesus RBCs. The different parasite stage samples were pooled together for long noncoding RNA sequencing at BGI Genomics sequencing platform (San Jose, California). Briefly, for the lncRNA library preparation, ribosomal RNA was removed using RNase H kit, followed by RNA fragmentation. The cDNAs were synthesized using a standard protocol at BGI Genomics (*67*). For second-strand synthesis, dUTP was used instead of dTTP. Double-stranded cDNA fragments are subjected to end-repair, and then a single "A" nucleotide is added to the 3′ ends of the blunt fragments. The reaction system and program for adaptor ligation were subsequently configured and set up to ligate adaptors with the cDNAs. This was followed by the final PCR for amplification of the cDNAs. The cDNAs were then denatured to generate single-stranded PCR products, which were cyclized. At the BGI Genomics facility, sequencing was performed where the single-stranded circular DNA molecules were replicated via rolling circle amplification, thereby generating DNA nanoballs (DNBs) with multiple copies of DNA. High-quality DNBs were then loaded into patterned nanoarrays using a high-intensity DNA nanochip technique and sequenced through combinatorial probe-anchor synthesis (cPAS).

### *Schizont maturation assays to test drug susceptibility*

A cryopreserved Brazilian *P. vivax* isolate was thawed as reported previously (*68*, *69*). The samples were then enriched on 1.080 g/mL KCl high Percoll gradients. Briefly, 2 mL of parasitized RBCs resuspended using Iscove's modified Dulbecco's medium (IMDM, Gibco) were layered on 3 mL 1.080 g/mL KCl high Percoll gradient and spun for 15 min at 1200× g with slow acceleration and no break. The interface post Percoll-gradient spin was removed, washed in incomplete IMDM, and used for drug assays. As a part of the drug assay setup, the enriched *P. vivax* parasites, predominantly early rings, were subjected to in vitro culture conditions with 0.2% hematocrit in IMDM containing 10% AB+ heat-inactivated sera and 50 μg/mL gentamicin at 37 °C in 5% CO2, 1% O2, and 94% N2. The effect of L-canaline (Cayman Chemicals), Sodium fluoroacetate (Molport,supplier Advamacs) and MMV019313 on parasite development was assessed by measuring the proportion of parasites maturing from ring to schizont stage based on their DNA content. Specifically, ring stage *P. vivax*, *P. knowlesi* and *P. falciparum* parasites at 0.5 to 1% parasitemia and 0.2% hematocrit were exposed to the drugs in 384-well flat clear bottom black plates (Corning) and incubated at 37°C. The final concentrations of the drugs in the plate ranged from 1.3 nM to 100 μM for MMV01913, L-canaline and 0.01 mM to 2 mM for sodium fluoroacetate. All three parasite species, *P. vivax*, *P. knowlesi*, and *P. falciparum*, were monitored for maturation from rings toward late stage schizonts in the presence of given drug versus no drug control. The drug assays were harvested at 28 to 30 hours for *P. knowlesi* and 44 to 48 hours for *P. vivax* and P. *falciparum* by staining with 1X SYBR green I dye for 15 min (working stock: 1:1000 dilution in 1X PBS). SYBR green intensities corresponding to parasite DNA content were measured using Cytek flow cytometer and analyzed using the FlowJo software (version 10.10.0). DHA-treated parasites served as a kill control to identify parasite gating for rings.

### *1.5 cycle assays to test drug susceptibility*

Drug sensitivities of *P. knowlesi* and *P. falciparum* asexual stage parasites were assessed using a

SYBR green I-based fluorescence assay (70). Stock solutions of dihydroartemisinin (DHA), MMV01913, and Roseoflavin (Cayman Chemicals) were prepared in DMSO at 10 mM, whereas L-canaline and sodium fluoroacetate were prepared in water with 0.1% Triton X-100 at 10 mM and 50 mM, respectively. The drugs were printed directly onto 96-well black clear bottom plates (Corning) using HP D300e digital dispenser. Synchronized ring-stage *Plasmodium* spp. parasites used for measurement with or without drugs were prepared in 96-well black clear bottom plates at a culture volume of 100 µl, 1%-hematocrit, and starting parasitemia of 1%, in RPMI-1640 as described previously (71). Subsequently, *P. knowlesi* and *P. falciparum* cultures were gassed and incubated at 37°C for 48 hours and 72 hours, respectively. To quantify the DNA content, the parasite cultures were lysed by adding 20 µl of a 6x lysis buffer (0.16% saponin, 20 mM Tris-HCl, 5 mM EDTA, 1.6% Triton X-100, pH 7.4) containing 1:1000 SYBR green I (Invitrogen, S7567). Fluorescence intensities were measured using a microplate reader (SpectraMax iD5, Molecular Devices) at the excitation wavelength 494 nm and the emission wavelength 530 nm. The raw fluorescence values were normalized to untreated controls after subtraction of background fluorescence.

### Fixed ratio method (isobologram) analysis for drug synergy

The fixed ratio method was used to assess the impact of combinatorial effect of DHA and Roseoflavin on *P. knowlesi*, as described previously (72). The two drugs, DHA and Roseoflavin, were combined to generate six dose-response curves based on combinatorial ratios, ensuring the $IC_{50}$ of each drug fell at a midpoint within a two-fold serial dilution. Out of the six combinatorial ratios two included the two drugs alone (100Roseoflavin:0DHA and 0Roseoflavin:0DHA) and the remaining four each had a component relating to each of the two drugs in the combination (80Roseoflavin:20DHA; 60Roseoflavin:40DHA; 40Roseoflavin:60DHA). SYBR-green I based drug assay was performed as described above. The SYBR fluorescence readout data obtained was used to generate dose-response curves using GraphPad. This included four dose-response curves relating to each combination of DHA and Roseoflavin and two curves for each drug alone. The $IC_{50}$ values calculated from each of these dose response curves were used to construct an isobologram using GraphPad (GraphPad Prism, 10.1.1).

### Quantification and statistical analysis for drug susceptibility

All statistical analysis was performed in Graphpad PRISM 10. $IC_{50}$ values were calculated using the nonlinear regression function (variable slope – four parameters, least squares regression). Statistical significance of $IC_{50}$ changes were determined using analysis of variance (ANOVA) or *t* test. Details of each analysis can be found in the corresponding figure legend.

### Drug selections of transposon pools

Initial transposon pools were combined into replicate 1 and replicate 2 (each originating from unique transfections) 14 to 25 days after initial transfections. Each perturbation selection was started with 1 ml packed RBCs at 2% parasitemia (~$2 \times 10^8$ mixed stage parasites). Drug was applied continuously at the concentrations shown in fig. S17. To ensure sufficient selection on parasites, drug concentration was increased through the experiment until a significant difference in the growth rate was observed between the treated and untreated control. If parasitemia dropped below 0.5%, then parasites were taken off drug to recover (fig. S17). Parasites were diluted as needed to maintain parasitemia between 1 and 10%, and culture media was replaced daily.

### Determination of gene copy number

The qPCR reactions were conducted using the StepOnePlus Real-Time PCR System (Applied Biosystems). Each reaction contained 10 µL of Fast SYBR Green Master Mix (Applied Biosystems, cat. no. 4385610), 0.15 µL of each forward and reverse primer (100 µM stock, final concentration 0.5 µM each), 2 µL of DNA, and nuclease-free water to a final volume of 20 µL. The following cycling conditions were used: an initial denaturation at 95°C for 20 s, followed by 40 cycles of denaturation at 95°C for 3 s and annealing/extension at 60°C for 30 s. Standard curves were generated using 10-fold serial dilutions of gel-extracted amplicons to ensure accurate quantification. Relative copy numbers were calculated using the StepOne Software (version 2.3). Primer sequences are provided in data S10.

### Tn-seq read alignment and preprocessing

The reference genome and gene annotations for *P. knowlesi* (H strain, release 58) were downloaded from https://plasmodb.org/. Read adaptors were trimmed using Cutadapt v.2.10 (73) from the 5′ end with parameter settings: –pair-adapters –pair-filter=any –quality-base=33 -m 30 -q 20 –discard-untrimmed. Trimmed paired-end reads were filtered using a custom R script to remove reads with no TTAA at the beginning of the read as well as reads with TTAA but without the transposon sequence in read1. Filtered paired-end reads were mapped to the genome using BWA-mem v.0.7.17 (74). PCR duplicates were removed using samtools v.1.10 markdup (75). High quality usable reads were extracted using samtools with parameter settings: samtools view -q 30 -f 66 -b -h. Filtered bam files from read1 were converted into bed files and all potential TTAA sites and their unique identifiers were identified and assembled into a count matrix using bedtools v.2.31.0 (76) and custom R scripts. Only read1s were used to construct the count matrix. Reads in the final bed files were matched to their unique TTAA identifiers. Read counts were determined in two ways: unidirectional and directional, where read in opposite orientation were considered as independent insertions.

### Background correction and gold plus lists

We curated a list of 30 putative background genes in *P. knowlesi* based on their orthologs having a targeted gene disruption in *P. falciparum,* evidence of essentiality in *P. berghei* from either targeted disruption or genomic screening and are essential in the *T. gondii* CRISPR-Cas9 screen. The extended gold-plus lists were generated using the following criteria: shared essential/dispensable calls in the *P. falciparum*, *P. berghei*, and *T. gondii* screens and no conflicting data with targeted gene disruptions, or essential/dispensable in both *P. falciparum* and *P. berghei* screens and targeted gene disruptions in both species. Essentiality data for targeted disruptions are from phenoplasm.org (61). Screening data are from the *P. falciparum* piggyBac screen (6), the *P. berghei* large-scale screen (7), and *T. gondii* CRISPR-Cas9 screen (8). Our criteria for high-confidence calls in each of these screens were: For *P. falciparum*, genes <650 bp were excluded; essential genes have MIS < 0.2 and classified as nonmutable; dispensable genes have a MIS > 0.8 and classified as mutable. For *P. berghei*, essential genes have RGR < 0.2 and classified as essential; dispensable genes have RGR > 0.9 and dispensable classification. For *T. gondii*, essential genes have CRISPR score < −3, and dispensable genes have CRISPR score > 0. For background correction, genes were ranked by essentiality according to the total number of reads mapped to the CDS of the genes. For this analysis, TTAA sites locates close to the end (>99%) of the transcript were not considered. Outliers were flagged as those with read counts over the 75% quantile and filtered out from subsequent analysis, resulting in 22 genes with high confidence (fig. S1B). The mean value of total reads mapped to TTAA sites within the CDS of the 22 genes were used as "background insertions" and subtracted from all TTAA sites from every gene per sample separately.

### Comparative essentiality analysis

Genes lacking orthologs in pairwise or across the three species of *Plasmodium* were filtered. For the remaining genes, the comparisons were performed under both high and low confidence cutoffs. Essential and dispensable groups with high confidence cutoff were flagged using the criteria for high-confidence calls in each of the screen as described in the "Background correction and gold plus lists" section. For low

confidence cutoff, genes labels were taken directly from the *P. berghei* and *P. falciparum* studies without additional filters. No change was applied to *P. knowlesi*.

### Assessment of saturation level

To assess the saturation level of insertions within the genome, we performed a simulation as follows. Genes were randomly split into two groups, insertable and un-insertable assuming (i) 60% of genes are insertable (lower bound) and (ii) 100% of genes are insertable (upper bound). TTAA sites in the CDS regions of genes were classified according to the genes they reside in. Sites were sampled at random, and the proportion of insertions and targeted genes were calculated. Total number of sampled sites ranged from 0 to 1000000. Saturation was estimated at 95% level of insertable genes being targeted.

### MIS calculation

The MIS was calculated as previously described (*11*). We applied the following modifications to account for noise in the data. First, the count data was corrected to remove background noise and count per million (CPM) normalization was performed. All TTAA sites locates at the >=99% the transcript were removed. We calculated the $MS_g$ score as $MS_g = \log\left(\frac{N_g+1}{D_g}\right)$. Here, $N_g$ is the total number of reads mapped to the CDS of gene $g$ and $D_g = \frac{N_{TTAA}\times 1000}{L_g}$ is the TTAA density of the gene $g$, i.e., the number of unique TTAA sites per kb of the CDS of gene $g$. $N_{TTAA}$ represents the total number of TTAA sites within the CDS of gene $g$ and $L_g$ is the CDS length of gene $g$. The $MS_{g_{bg}}$ of the 22 gold list of essential genes were used to estimate the background mean and standard deviation $\mu_{bg}$ and $\sigma_{bg}$. Then, $MS_g$ of all genes were corrected by the z-score normalization with $\mu_{bg}$ and $\sigma_{bg}$ and shifted using by max normalized score of gold list ($GL$) to calculate the final $MS_{g_{new}}$ score

$$MS_{g_{new}} = \frac{MS_g - \mu_{bg}}{\sigma_{bg}} - \max_{bg\in GL}\left(\frac{MS_{g_{bg}} - \mu_{bg}}{\sigma_{bg}}\right)$$

A Gaussian mixture model was then fitted to the $MS_{g_{new}}$ and the parameters were inferred using the expectation maximization (EM) algorithm. Finally, the MIS was calculated as the posterior probability that gene $g$ belongs to the dispensable gene group using the Bayes formula

$$\begin{aligned} MIS_g &= P(\pi_g = 1 | MS_{g_{new}}, \mu_{\pi_g=1}, \sigma_{\pi_g=1}) \\ &= \frac{P(\pi_g = 1) f(MS_{g_{new}} | \mu_{\pi_g=1}, \sigma_{\pi_g=1})}{\sum_{\pi_g=1,0} P(\pi_g) f(MS_{g_{new}} | \mu_{\pi_g}, \sigma_{\pi_g})} \end{aligned}$$

where $\pi_g$ is a Bernoulli random variable representing the essentiality of gene $g$ with $\pi_g = 1$ corresponding to dispensable and $\pi_g = 0$ corresponding to essential. Here $f$ represents the normal density.

### Weight calculation for TTAA site

To account for the distribution of TTAA sites and their relative position within a gene, we devised a weighting scheme that proportionally assigns weights based on distance from the TSS. Specifically, introns were removed from every gene $g$ and exons were concatenated. The CDS length of every gene $g$ was normalized to 0 to 1 and the relative distance $R_i$ of each TTAA site $i$ was calculated by

$$R_i = \frac{|d_i - L_g|}{L_g} - \frac{1}{2}$$

where $d_i$ is the distance of site $i$ to the TSS and $L_g$ is the CDS length. Subsequently, the weight of each TTAA site $i$ was calculated based on the quadratic equation

$$W_i = -\frac{5}{8}R_i^2 + \frac{2}{5}R_i + 1$$

The coefficients $-\frac{5}{8}$ and $\frac{2}{5}$ were calculated to optimize the weights in a way that weight is the highest closer to the center and quadratically drops down closer to the ends, with the 3′ end having the lowest weight. The rationale behind this formulation is that the disruption of TTAA sites at the 3′ end should be less effective in disrupting the protein product. $W_i$ was then passed through a sigmoid function to cause a quicker drop at the ends and reduce the weight of TTAA locating at extreme 3′ end of the CDS with settings: α = 1000, β = 0.48.

$$y = \frac{1}{1 + e^{-\alpha(W_i - \beta)}}$$

These weights were used in all subsequent models.

### Modified MIS

A modified MIS model was devised to reduce the bias introduce by gene length as follows

$$\widetilde{MS}_g = \log\left(\frac{\sum N_i \times W_i + 1}{N_{TTAA}}\right)$$

Here, $N_i$ denotes the total number of reads mapped at the TTAA site $i$ within the CDS of gene $g$, $W_i$ is the weight of the TTAA site $i$, and $N_{TTAA}$ is the total number of unique TTAA. All the theoretical TTAA sites were weighted with the weight function introduced in the last paragraph. We calculated the $\widetilde{MS}_g$ of gene based on a modified equation without CDS length $L_g$ because it introduces a bias in dispensability calls for genes with long CDS length. We background corrected and normalize the $\widetilde{MS}_g$ and calculate the posterior probability of dispensability of the genes as in the MIS calculations.

### OIS

To leverage the statistical power of independent transfection pools, we devised an OIS as follows. First, samples from the same transfection pool were merged, resulting in 10 sets. The occupancy score of gene $g$ is defined by

$$OS_g = \log\left(\frac{\sum O_i \times W_i + 1}{N_{TTAA}}\right)$$

where $O_i$ is the occupancy count for site $i$ within the CDS of gene $g$ and $W_i$ is the weight for site $i$ based on the previously defined weight function. Occupancy counts for the OIS model were collapsed into binary values after background noise removal and CPM normalization (cutoff = 0.5) for each transfection pool (occupancy per site per transfection pool). $OS_g$ of all genes were then normalized using the same approach as in the $MS_g$ and MIS model.

$$OS_{g_{new}} = \frac{OS_g - \mu_{bg}}{\sigma_{bg}} - \max\left(\frac{OS_{g_{bg}} - \mu_{bg}}{\sigma_{bg}}\right)$$

The $OS_{g_{new}}$ was modeled using a mixed Gaussian distribution and parameters were fitted using the EM algorithm. Same procedure as in MIS was applied to calculate the posterior probability of gene essentiality with OIS.

### Bayesian network model

We developed a Bayesian methodology to infer gene essentiality as well as the chromatin state of the TTAA sites within genes.

#### Model overview

The Bayesian approach probabilistically estimates the essentiality of each gene as well as mutability of the sites and their innate chromatin state. As input, the model takes in the observed insertion counts at each site within the CDS of the genes, runs a probabilistic query to find most probable configuration of the gene essentiality and site mutability, and outputs the estimates of gene essentiality, chromatin state, and site mutability. To achieve this, we first construct a directed acyclic graph (DAG) for each gene (fig. S4A). The graph consists of the following layers. The head node is a binary random variable representing the gene. The value of this node will probabilistically determine the essentiality of the gene (0: essential and 1: nonessential). Directly below the node, we introduce a layer of nodes to model mutability of the (ordered) sites within the CDS of the gene (one node per site). The value of these nodes model represents the site state (0: immutable; 1: mutable). Each of these nodes has a child node directly underneath representing the chromatin state of the site (0: inaccessible/1: accessible). In this layer, nodes are connected to the immediate downstream node, forming a Markov chain. The rationale for these connections is to account for spatial chromatin state, which may affect consecutive sites (sites near (in)accessible sites are more likely to be (in)accessible). Finally, each of these nodes has a direct child node representing the

observable counts. Input count data (evidence) is provided to the model and through a sampling procedure, values and corresponding probabilities of unobserved nodes are estimated to identify the most probable configuration.

### *Mathematical details*

For each gene we constructed a directed acyclic graph

$$AG = (N, E)$$

where $N$ represents the nodes and $E$ represents the connections (edges) between the nodes. The network consists of several layers. The terminal nodes in the network $X_i$, represent the observed counts for each site in the original count matrix. Directly above the count nodes, latent variable $C_i$ are introduced to model the latent chromatin state. These random variables are Bernoulli with two states (0: inaccessible/1: accessible). Moreover, to account for chromatin state in the neighboring regions, we introduced a Markov Chain where state of $C_{i-1}$ influences the state of $C_i$. Nodes $S_i$ are also Bernoulli random variables representing the site state (0: immutable/1: mutable) within the CDS of gene $G$. Finally, $G_i$ represents the gene state (0: essential/1: dispensability). Figure S4A shows the schematic of the Bayesian network.

For terminal observed counts notes, we assumed that each $X_i$ is independent and used a mixture of two negative binomial distribution representing essential and dispensable sites) to model $X_i$.

$$X_i \sim \pi NB(p_0, u_0) + (\pi - 1)NB(p_1, u_1)$$

The parameters ($\pi$, $p_0$, $u_0$, $p_1$, $u_1$) of the mixed negative binomial distribution model were inferred by EM algorithm via maximizing $P(G_i, S_i, C_i \mid X_i)$.

To fit the model, we developed a Gibbs simpler that was used to identify the most likely state of each node in the network. The joint probability distribution of each node was factorized given its Markov blanket (MB). The factorization and posterior state of each state were calculated as follows

$$P(C_i = 0|MB) = P(C_i = 0|S_i = s_i, C_{i-1} = c_{i-1}, X_i = x_i)$$
$$= \frac{P(X_i = x_i|S_i = s_i, C_{i-1} = c_{i-1}, C_i = 0) \times P(S_i = s_i, C_{i-1} = c_{i-1}, C_i = 0)}{P(S_i = s_i, C_{i-1} = c_{i-1}, X_i = x_i)}$$
$$= \frac{P(X_i = x_i|C_i = 0) \times P(C_i = 0|C_{i-1} = c_{i-1}, S_i = s_i) \times P(C_{i-1} = c_{i-1}) \times P(S_i = s_i)}{P(X_i = x_i) \times P(C_{i-1} = c_{i-1}) \times P(S_i = s_i)}$$

$P(C_i = 0|C_{i-1} = c_{i-1}, S_i = 0)$ can be calculated by the transition probability matrix

$$\begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

$P(C_i = 0|C_{i-1} = c_{i-1}, S_i = 1)$ can be calculated by the transition probability matrix

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$$

$P(C_i = 1|MB)$ was calculated by $P(C_i = 1|MB) = 1 - P(C_i = 0|MB)$. The posterior of $S_i$ nodes are given by

$$P(S_i = 0|MB) = P(S_i = 0|G_i = g, C_{i-1} = c_{i-1}, C_i = c_i)$$
$$= \frac{P(C_i = c_i|C_{i-1} = c_{i-1}, S_i = 0) \times P \times S_i = 0|(G_i = g) \times P(C_{i-1} = c_{i-1}) \times P(G_i = g)}{P(G_i = g, C_{i-1} = c_{i-1}, C_i = c_i)}$$
$$= \frac{P(C_i = c_i|C_{i-1} = c_{i-1}, S_i = 0) \times P(S_i = 0|G_i = g) \times P(C_{i-1} = c_{i-1}) \times P(G_i = g)}{P(G_i = g) \times P(C_{i-1} = c_{i-1}) \times P(C_i = c_i)}$$

$P(S_i = 0|G_i = g)$ can be calculated by the transition probability matrix

$$\begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}$$

Similarly, $P(S_i = 1|MB)$ was calculated by $P(S_i = 1|MB) = 1 - P(S_i = 0|MB)$. The network was fully instantiated at random and node values were updated using a Gibbs Sampling procedure. Finally, the posterior probability of gene $G_i$ modeled using a saturation logistic function with adaptive weight $w_i$

$$P(G_i = 0|MB) = P(G_i = 0|S) = \frac{e^{\sum w_i S_i}}{1 + e^{-\sum w_i S_i}}$$

$$w_i = \frac{5}{1 + e^{5 - N_{TTAA}}} + 1$$

where $N_{TTAA}$ is the total number of theoretical TTAA sites locating at the CDS of gene $G_i$. Note that the more sites a gene has, the more confident the essentiality of the gene will be. The $P(G_i = 1|MB)$ was calculated by: $P(G_i = 1|MB) = P(G_i = 1|S) = 1 - P(G_i = 0|S)$. The probability $P(G_i = 1|MB)$ is defined as the BMS of the dispensability of gene.

### *HMS calculation*

To take advantage of both MMIS and BMS, we devised a hybrid model combining the two using a convex sum with an adaptive weight as follows

$$P(G_i|S) = (1 - \lambda)p_a + \lambda p_b$$

$$\lambda = \frac{1}{1 + e^{5 - N_{TTAA}}}$$

where $p_a$ and $p_b$ represent the posterior probability of gene $G$ being dispensable in MMIS model and Bayesian network model, respectively. The transition point was set to 5, which causes the hybrid models to softly switch to MMIS for genes with lower than 5 TTAA sites, and the BMS for genes with more than 5 TTAA sites. The rationale behind this hybrid model is that MMIS tends to be less conservative for calling dispensability for genes with low number of TTAA sites. Hence this soft switch allows for confident calls for all genes with at least 3 TTAA sites (fig. S5).

### *FIS*

To access the fitness advantages and defects of genes over time, regardless of essentiality score, we devised a FIS to track counts across timepoints. A total of three time points (day 9, day 14, and day 19) were used. Samples before day 9 were excluded from this analysis. Background noise removed and CPM normalized count matrix was used to calculate the relative abundance of each mutant $m$ for each transfection pool at each time point separately as

$$MFS_g = \log\left[\frac{\left(\sum_m r_{t,g}^{m,k} + 1\right) \times \frac{1}{\sum_k I(t \in TPN_k)} \times \frac{1}{\sum_m I(m \in g) + 1}}{N_{TTAA}}\right]$$

Here, $r_{t,g}^{m,k}$ represents the normalized reads number of mutant $m$ targeting gene $g$ on its CDS in transfection pool $k$, $\sum_k I(t \in TPN_k)$ is the total number of transfections that contains the timepoint $t$, $\sum_m I(m \in g)$is the total number of mutants targeting gene $g$ in all samples. $I(t \in Tr_k)$ and $I(m \in g)$ are two indicator functions which are introduced to normalize differences in sample numbers and time points.

$$I(t \in TPN_k) = \begin{cases} 1; & \text{if } t \text{ in } k \\ 0; & \text{else} \end{cases}, I(m \in g) = \begin{cases} 1; & \text{if } m \neq 0 \\ 0; & \text{else} \end{cases}$$

A linear regression model was fitted for each gene's $MFS_g$ using the three time points. $MFS_g$ is defined as the slope of fitted regression line. This slope captures the fitness trend of gene $g$ over time. Significance of slope was estimated as the linear regression slope coefficient $P$ values corrected for multiple hypothesis testing using the false discovery rate (FDR) method.

### *Identification of truncatable genes*

To identify truncatable genes, we devised a simple change point (cp) detection algorithm that fits a step function to a binary vector $S$ tracking the mutability of the TTAA sites associated with the gene. This vector is calculated by the Bayesian Network, which consists of nodes representing site insertability. The cp values range from 0 to 1, covering all points on the normalized length of the gene. The goodness of the fit was estimated by least squares approximations. The best fit was used to rank gene truncatablility. This procedure was applied 5′ to 3′ and 3′ to 5′ to identify both 3′ and 5′ truncatable genes. We used cutoff $0.1 < R_i < 0.9$ and $0 < cp < 1$ and MSE<25 percentile/75 percentile (fig. S10) to identify truncatable genes with high confidence, where MSE represents the Mean squared error between fitted step function and the vector $S$. cp is the relative location of the changing point within the gene based on ranked TTAA sites and $R_i$ represents the relative distance of changing point to TSS.

### *HMM for identifying depleted regions*

We developed a HMM to identify depleted and enriched TTAA insertion regions across

all chromosomes. The hidden state of the HMM is a binary random variable $X$ representing the status of the site as 0 for depleted and 1 for nondepleted. The observable node $O$ downstream of $X$ represents the actual observed counts of the background corrected counts matrix. This random variable was modeled as a geometric random variable with parameters specific to the hidden state. The geometric distribution parameters for each state were estimated by modelling the observed background counts as a mixture of two geometric distributions with state specific parameters estimated using observed read counts. These probabilities were used as emission probabilities. The transition probabilities between states were set by trial and error as

$$\begin{pmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{pmatrix}$$

The Vitterbi algorithm was used to estimate the posterior probability of hidden state. We modified the source Python code from (*77*) to generate a custom R script. Final depleted regions were marked using a cutoff of ≥ 5 consecutive TTAA sites labeled as "depleted" by the model. Identified depleted regions were visualized on the circos plot.

### Hot-spot and cold-spot detection

A similar HMM was trained to identify hotspots (i.e., TTAA sites with very high insertion levels) and coldspots (depleted regions). The HMMs consisted of two states and the emission probabilities were estimated using the quantiles of the count distribution (counts on the upper 0.999 quantiles for hot-spots and reads < 1 count for coldspots). Transition probability matrices were set by trial and error as

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$$

An independent analysis was used to validate hotsopts and coldspots as follows: 0 counts after background noise removal in exons of dispensable genes and intergenic regions were directly extracted for motif enrichment. Regions 10 nt upstream and downstream of the coldspots and hotspots were extracted for motif enrichment analysis.

### Motif analysis

Extracted fasta sequences were used to perform motif search. Motif search was performed using BAMMmotif (*78*).

### Orthology analysis

All orthologs genes between Apicomplexan species were identified by OrthoMCL (https://orthomcl.org/) (*79*). Search terms used for defining conservation groups in OrthoMCL (release 6.2) were by phyetic pattern: *P. knowlesi* (Pkno=1T), *P. vivax* clade (pcym+pkno+pvip=3T), *Plasmodium* ((pber+pfal+pcha+pgal+prei+pyoe>=1T) AND (pkno=1T)), apicomplexan ((PIRO+COCC>=1T) AND (pkno=1T), and eukaryotic (CILI+FUNG+META+VIRI>=3T AND (pkno=1T)). Reciprocal blast on protein sequences by R package orthologr (v.0.4.0) (*80*) with settings: eval = 0.001. OrthoFinder2 (*81*) was used to further narrow down the list and increase the confidence in ortholog calls.

### Paralogs, synteny, and protein localization data

Paralogs and syntenic genes were downloaded from PlasmoDB (*21*). Protein localization was identified using orthologous genes in *T. gondii* and hyperLopit data (*26*).

### Isoform essentiality

Isoforms were identified using the same pipeline as in in the lncRNAs. Same scores as HMS, OIS and MIS were used to call the essentiality for all identified isoforms. The transcripts flagged as "=" in transcript assembly result were filtered out.

### Perturbation analysis at gene and site level

We devised two approaches to statistically assess the difference in insertions between treated and control groups. In the first approach, we aggregate the site level insertions to gene level by summing the background corrected counts into a single count. This was done for each replicate (2 biological and 2 technical) and the fold change enrichments and $P$ values were calculated using the background corrected and CPM normalized counts with the edgeR (v.3.40.2) package (*40*).

In the second approach, background corrected, CPM normalized count matrix was used to calculate the fold change for every site. Sites with 0 count were removed from all samples. To further increase statistical confidence, genes with less than 6 sites (bidirectional) were excluded from the analysis. The FC was calculated for each site and the inverse of coefficient of variation (CV) was calculated by

$$\text{CV\_inverse} = \frac{\text{Mean}(\log_2 \text{FC})}{\text{Sd}(\log_2 \text{FC})}$$

In both approaches, significance was determined using log2FC at 95 and 5; 97 and 3; 99 and 1 percentile, for insertional increasing and decreasing genes, respectively. Final "hits" were determined using the following criteria: gene must contain >1 TTAA site, have a HMS >0.26 and one of the following: in the 99% percentile in both replicates in either the site-level model or the EdgeR model (for EdgeR must also have a $P$ value and FDR < 0.05), or be the single highest change in either direction in either replicate by either model.

### lncRNA processing

Trimmed reads were mapped to the genome described above by using HISAT2 v.2.1.0 (*82*) with the following parameters: –rna-strandness RF –dta –score-min L,0,-0.6 –no-mixed –no-discordant. Nonuniquely mapped reads and low-quality reads were filtered out by samtools v.1.19 (*82*) using parameters: -q 10 -f 2 -b -h. The quality of filtered reads was analyzed by RseQC v.2.6.4 (*83*) and MultiQC v.1.6 (*84*). Samtools was used to sort bam files. The transcriptome was assembled using stringtie v.2.1.7 (*85*) with and without GTF annotation. The GTF annotation and assembled GTF files were compared using gffcompare v.0.12.6 (*86*). Novel lncRNA were filtered out by customized bash script using the following criteria: 1. Transcripts igreater than 200bp with number of exons greater than 1; 2. Read coverage of transcript greater than 3; 3. Transcript is not protein-coding. The fasta files of filtered potential lncRNAs were extracted by gffread v.0.12.7 (*86*) and processed by CPC2 (*87*) to assess coding potentiality. Transcripts have coding potentiality flagged by CPC2 with default setting were filtered out. Novel lncRNAs flagged with "i" (intronic), "x" (antisense), "o" (sense), "u" (intergenic) classes were filtered using a customized bash script.

### Gene set enrichment

Gene ontology enrichment analysis (GOEA) was performed using available GO on PlasmoDB. Significant GO terms (Benjamini < 0.5 and rank <= 8 and Result count >=3) were identified and the $\log_{10}$ (fold enrichment) and $-\log_{10}$ ($P$ value) were used to visualize the significant GO terms.

### Structural variants

To identify structural vairants (indels, SNPs) of the human-adapted Pk YH1 strain used in this study, the genome DNA (SRA accession number: SRX1553878) was processed as follows: read quality was analyzed by FastQC, pair-end reads were trimmed by Cutadapt v.2.10, sorted by samtools v.1.10 and mapped to genome using BWA-mem v.0.7.17. Structural variants were called by delly v.1.2.6 (*88*). Deletion variants were filtered by bcftools v.1.10.2 (*89*) using the parameter 'SVTYPE="DEL"'. BCF files were converted into vcf files and bed files with setting: bcftools query -f '%CHROM\t%POS\t%INFO/END\t%INFO\t%FILTER\t.\t.\n'. Deletion events with flag "PASS" and "PRECISE" were retained and processed by a custom R scripts. Calls were manually verified in IGV. Duplicated regions were filtered using the parameter 'SVTYPE="DUP" with flag "PASS" or "PRECISE" and visually verified in IGV as well.

## REFERENCES AND NOTES


1. World Health Organization (WHO), "World Malaria Report 2023" (2023); https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023.
2. P. E. Duffy, J. P. Gorres, Malaria vaccines since 2000: Progress, priorities, products. *NPJ Vaccines* **5**, 48 (2020). doi: 10.1038/s41541-020-0196-3; pmid: 32566259
3. K. E. Ward, D. A. Fidock, J. L. Bridgford, *Plasmodium falciparum* resistance to artemisinin-based combination therapies. *Curr.*

*Opin. Microbiol.* **69**, 102193 (2022). doi: 10.1016/j.mib.2022.102193; pmid: 36007459
4. S. Tachibana *et al*., *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* **44**, 1051–1055 (2012). doi: 10.1038/ng.2375; pmid: 22863735
5. R. W. Moon *et al*., Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 531–536 (2013). doi: 10.1073/pnas.1216457110; pmid: 23267069
6. C. Lim *et al*., Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat. Commun.* **4**, 1638 (2013). doi: 10.1038/ncomms2612; pmid: 23535659
7. A. C. Y. Chua *et al*., Robust continuous in vitro culture of the *Plasmodium cynomolgi* erythrocytic stages. *Nat. Commun.* **10**, 3635 (2019). doi: 10.1038/s41467-019-11332-4; pmid: 31406175
8. A. A. Escalante, A. S. Cepeda, M. A. Pacheco, Why *Plasmodium vivax* and *Plasmodium falciparum* are so different? A tale of two clades and their species diversities. *Malar. J.* **21**, 139 (2022). doi: 10.1186/s12936-022-04130-9; pmid: 35505356
9. K. Yusa, L. Zhou, M. A. Li, A. Bradley, N. L. Craig, A hyperactive *piggyBac* transposase for mammalian applications. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1531–1536 (2011). doi: 10.1073/pnas.1008322108; pmid: 21205896
10. I. F. Bronner *et al*., Quantitative insertion-site sequencing (QIseq) for high throughput phenotyping of transposon mutants. *Genome Res.* **26**, 980–989 (2016). doi: 10.1101/gr.200279.115; pmid: 27197223
11. M. Zhang *et al*., Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360**, eaap7847 (2018). doi: 10.1126/science.aap7847; pmid: 29724925
12. E. Knuepfer *et al*., Divergent roles for the RH5 complex components, CyRPA and RIPR in human-infective malaria parasites. *PLOS Pathog.* **15**, e1007809 (2019). doi: 10.1371/journal.ppat.1007809; pmid: 31185066
13. E. Bushell *et al*., Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260–272.e8 (2017). doi: 10.1016/j.cell.2017.06.030; pmid: 28708996
14. S. M. Sidik *et al*., A Genome-wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes. *Cell* **166**, 1423–1435.e12 (2016). doi: 10.1016/j.cell.2016.08.019; pmid: 27594426
15. S.-K. Lee *et al*., Merozoite surface protein 1 paralog is involved in the human erythrocyte invasion of a zoonotic malaria, *Plasmodium knowlesi*. *Front. Cell. Infect. Microbiol.* **13**, 1314533 (2023). doi: 10.3389/fcimb.2023.1314533; pmid: 38111629
16. D. N. Ndegwa *et al*., Using *Plasmodium knowlesi* as a model for screening *Plasmodium vivax* blood-stage malaria vaccine targets reveals new candidates. *PLOS Pathog.* **17**, e1008864 (2021). doi: 10.1371/journal.ppat.1008864; pmid: 34197567
17. F. Mohring *et al*., Rapid and iterative genome editing in the malaria parasite *Plasmodium knowlesi* provides new tools for *P. vivax* research. *eLife* **8**, e45829 (2019). doi: 10.7554/eLife.45829; pmid: 31205002
18. R. W. Moon *et al*., Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7231–7236 (2016). doi: 10.1073/pnas.1522469113; pmid: 27303038
19. R. J. Hill *et al*., Regulation and essentiality of the StAR-related lipid transfer (START) domain-containing phospholipid transfer protein PFA0210c in malaria parasites. *J. Biol. Chem.* **291**, 24280–24292 (2016). doi: 10.1074/jbc.M116.740506; pmid: 27694132
20. C. H. Kocken *et al*., *Plasmodium knowlesi* provides a rapid in vitro and in vivo transfection system that enables double-crossover gene knockout studies. *Infect. Immun.* **70**, 655–660 (2002). doi: 10.1128/IAI.70.2.655-660.2002; pmid: 11796595
21. J. Alvarez-Jarreta *et al*., VEuPathDB: The eukaryotic pathogen, vector and host bioinformatics resource center in 2023. *Nucleic Acids Res.* **52**, D808–D816 (2024). doi: 10.1093/nar/gkad1003; pmid: 37953350
22. T. A. Thompson, Z. Chahine, K. G. Le Roch, The role of long noncoding RNAs in malaria parasites. *Trends Parasitol.* **39**, 517–531 (2023). doi: 10.1016/j.pt.2023.03.016; pmid: 37121862
23. A. Gogol-Döring *et al*., Genome-wide profiling reveals remarkable parallels between insertion site selection properties of the MLV retrovirus and the *piggyBac* transposon in primary human CD4$^+$ T cells. *Mol. Ther.* **24**, 592–606 (2016). doi: 10.1038/mt.2016.11; pmid: 26755332
24. M. A. Li *et al*., The *piggyBac* transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. *Mol. Cell. Biol.* **33**, 1317–1330 (2013). doi: 10.1128/MCB.00670-12; pmid: 23358416
25. B. Elsworth *et al*., Proteomic analysis reveals novel proteins associated with the *Plasmodium* protein exporter PTEX and a loss of complex stability upon truncation of the core PTEX component, PTEX150. *Cell. Microbiol.* **18**, 1551–1569 (2016). doi: 10.1111/cmi.12596; pmid: 27019089
26. K. Barylyuk *et al*., A Comprehensive Subcellular Atlas of the *Toxoplasma* Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. *Cell Host Microbe* **28**, 752–766.e9 (2020). doi: 10.1016/j.chom.2020.09.011; pmid: 33053376
27. A. F. Cowman, C. J. Tonkin, W.-H. Tham, M. T. Duraisingh, The molecular basis of erythrocyte invasion by malaria parasites. *Cell Host Microbe* **22**, 232–245 (2017). doi: 10.1016/j.chom.2017.07.003; pmid: 28799908
28. M. S. Paoletta, S. E. Wilkowsky, Thrombospondin related anonymous protein superfamily in vector-borne apicomplexans: The parasite's toolkit for cell invasion. *Front. Cell. Infect. Microbiol.* **12**, 831592 (2022). doi: 10.3389/fcimb.2022.831592; pmid: 35463644
29. P. Kundu *et al*., The structure of a *Plasmodium vivax* Tryptophan Rich Antigen domain suggests a lipid binding function for a pan-*Plasmodium* multi-gene family. *Nat. Commun.* **14**, 5703 (2023). doi: 10.1038/s41467-023-40885-8; pmid: 37709739
30. J.-A. Gabelich *et al*., A member of the tryptophan-rich protein family is required for efficient sequestration of *Plasmodium berghei* schizonts. *PLOS Pathog.* **18**, e1010846 (2022). doi: 10.1371/journal.ppat.1010846; pmid: 36126089
31. L. Fan *et al*., An erythrocyte membrane-associated antigen, PvTRAg-26 of *Plasmodium vivax*: A study of its antigenicity and immunogenicity. *Front. Public Health* **8**, 148 (2020). doi: 10.3389/fpubh.2020.00148; pmid: 32411650
32. S. Dankwa *et al*., Ancient human sialic acid variant restricts an emerging zoonotic malaria parasite. *Nat. Commun.* **7**, 11187 (2016). doi: 10.1038/ncomms11187; pmid: 27041489
33. M. Marothia *et al*., Targeting *Pf*Prohibitin 2-Hu-Hsp70A1A complex as a unique approach towards malaria vaccine development. *iScience* **27**, 109918 (2024). doi: 10.1016/j.isci.2024.109918; pmid: 38812541
34. B. Forte *et al*., Prioritization of molecular targets for antimalarial drug discovery. *ACS Infect. Dis.* **7**, 2764–2776 (2021). doi: 10.1021/acsinfecdis.1c00322; pmid: 34523908
35. J. E. Gisselberg, Z. Herrera, L. M. Orchard, M. Llinás, E. Yeh, Specific inhibition of the bifunctional farnesyl/geranylgeranyl diphosphate synthase in malaria parasites via a new small-molecule binding site. *Cell Chem. Biol.* **25**, 185–193.e5 (2018). doi: 10.1016/j.chembiol.2017.11.010; pmid: 29276048
36. E. Jortzik *et al*., Redox regulation of *Plasmodium falciparum* ornithine δ-aminotransferase. *J. Mol. Biol.* **402**, 445–459 (2010). doi: 10.1016/j.jmb.2010.07.039; pmid: 20673832
37. S. C. Nair, J. T. Munro, A. Mann, M. Llinás, S. T. Prigge, The mitochondrion of *Plasmodium falciparum* is required for cellular acetyl-CoA metabolism and protein acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2210929120 (2023). doi: 10.1073/pnas.2210929120; pmid: 37068227
38. B. A. Munro, B. J. McMorran, Antimalarial drug strategies to target *Plasmodium* gametocytes. *Parasitologia* **2**, 101–124 (2022). doi: 10.3390/parasitologia2020011
39. B. Ogutu *et al*., Ganaplacide (KAF156) plus lumefantrine solid dispersion formulation combination for uncomplicated Plasmodium falciparum malaria: An open-label, multicentre, parallel-group, randomised, controlled, phase 2 trial. *Lancet Infect. Dis.* **23**, 1051–1061 (2023). doi: 10.1016/S1473-3099(23)00209-8; pmid: 37327809
40. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010). doi: 10.1093/bioinformatics/btp616; pmid: 19910308
41. M. Lim *et al*., UDP-galactose and acetyl-CoA transporters as *Plasmodium* multidrug resistance genes. *Nat. Microbiol.* **1**, 16166 (2016). doi: 10.1038/nmicrobiol.2016.166
42. G. LaMonte *et al*., Mutations in the *Plasmodium falciparum* cyclic amine resistance locus (PfCARL) confer multidrug resistance. *mBio* **7**, e00696-16 (2016). doi: 10.1128/mBio.00696-16; pmid: 27381290
43. P. A. Magistrado *et al*., *Plasmodium falciparum* cyclic amine resistance locus (PfCARL), a resistance mechanism for two distinct compound classes. *ACS Infect. Dis.* **2**, 816–826 (2016). doi: 10.1021/acsinfecdis.6b00025; pmid: 27933786
44. K. Pandit, N. Surolia, S. Bhattacharjee, K. Karmodiya, The many paths to artemisinin resistance in *Plasmodium falciparum*. *Trends Parasitol.* **39**, 1060–1073 (2023). doi: 10.1016/j.pt.2023.09.011; pmid: 37833166
45. S. C. Xie, S. A. Ralph, L. Tilley, K13, the cytostome, and artemisinin resistance. *Trends Parasitol.* **36**, 533–544 (2020). doi: 10.1016/j.pt.2020.03.006; pmid: 32359872
46. C. V. Pires *et al*., Chemogenomic Profiling of a *Plasmodium falciparum* Transposon Mutant Library Reveals Shared Effects of Dihydroartemisinin and Bortezomib on Lipid Metabolism and Exported Proteins. *Microbiol. Spectr.* **11**, e05014-22 (2023). doi: 10.1128/spectrum.05014-22; pmid: 37067430
47. M. Zhang *et al*., The apicoplast link to fever-survival and artemisinin-resistance in the malaria parasite. *Nat. Commun.* **12**, 4563 (2021). doi: 10.1038/s41467-021-24814-1; pmid: 34315897
48. S. C. Xie *et al*., Haemoglobin degradation underpins the sensitivity of early ring stage *Plasmodium falciparum* to artemisinins. *J. Cell Sci.* **129**, 406–416 (2016). doi: 10.1242/jcs.178830; pmid: 26675237
49. F. Rocamora *et al*., Oxidative stress and protein damage responses mediate artemisinin resistance in malaria parasites. *PLOS Pathog.* **14**, e1006930 (2018). doi: 10.1371/journal.ppat.1006930; pmid: 29538461
50. F. Ariey *et al*., A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50–55 (2014). doi: 10.1038/nature12876; pmid: 24352242
51. H. Ke *et al*., Mitochondrial type II NADH dehydrogenase of *Plasmodium falciparum* (PfNDH2) is dispensable in the asexual blood stages. *PLOS ONE* **14**, e0214023 (2019). doi: 10.1371/journal.pone.0214023; pmid: 30964863
52. K. E. Boysen, K. Matuschewski, Arrested oocyst maturation in *Plasmodium* parasites lacking type II NADH:ubiquinone dehydrogenase. *J. Biol. Chem.* **286**, 32661–32671 (2011). doi: 10.1074/jbc.M111.269399; pmid: 21771793
53. S. Mok *et al*., Artemisinin-resistant K13 mutations rewire *Plasmodium falciparum's* intra-erythrocytic metabolic program to enhance survival. *Nat. Commun.* **12**, 530 (2021). doi: 10.1038/s41467-020-20805-w; pmid: 33483501
54. J. Wang *et al*., Artemisinin directly targets malarial mitochondria through its specific mitochondrial activation. *PLOS ONE* **5**, e9582 (2010). doi: 10.1371/journal.pone.0009582; pmid: 20221395
55. C. R. Harding *et al*., Genetic screens reveal a central role for heme metabolism in artemisinin susceptibility. *Nat. Commun.* **11**, 4813 (2020). doi: 10.1038/s41467-020-18624-0; pmid: 32968076
56. A. L. Hemasa, M. Mack, K. J. Saliba, Roseoflavin, a natural riboflavin analogue, possesses *in vitro* and *in vivo* antiplasmodial activity. *Antimicrob. Agents Chemother.* **66**, e00540-22 (2022). doi: 10.1128/aac.00540-22; pmid: 36094195
57. W. Zhan *et al*., Dual-pharmacophore artezomibs hijack the *Plasmodium* ubiquitin-proteasome system to kill malaria parasites while overcoming drug resistance. *Cell Chem. Biol.* **30**, 457–469.e11 (2023). doi: 10.1016/j.chembiol.2023.04.006; pmid: 37148884
58. T. Reyser *et al*., Epidrugs as promising tools to eliminate *Plasmodium falciparum* artemisinin-resistant and quiescent parasites. *Pharmaceutics* **15**, 2440 (2023). doi: 10.3390/pharmaceutics15102440; pmid: 37896200
59. A. Chaurasiya *et al*., Targeting Artemisinin-Resistant Malaria by Repurposing the Anti-Hepatitis C Virus Drug Alisporivir. *Antimicrob. Agents Chemother.* **66**, e00392-22 (2022). doi: 10.1128/aac.00392-22; pmid: 36374050
60. M. R. Rosenthal, C. L. Ng, A Proteasome Mutation Sensitizes *P. falciparum* Cam3.II K13$^{C580Y}$ Parasites to DHA and OZ439. *ACS Infect. Dis.* **7**, 1923–1931 (2021). doi: 10.1021/acsinfecdis.0c00900; pmid: 33971094
61. N. Suthram *et al*., Elucidation of DNA repair function of PfBlm and potentiation of artemisinin action by a small-molecule inhibitor of RecQ helicase. *MSphere* **5**, e00956-20 (2020). doi: 10.1128/mSphere.00956-20; pmid: 33239368
62. N. V. Simwela *et al*., *Plasmodium berghei* K13 mutations mediate *in vivo* artemisinin resistance that is reversed by proteasome inhibition. *mBio* **11**, e02312-20 (2020). doi: 10.1128/mBio.02312-20; pmid: 33173001
63. B. H. Stokes *et al*., Covalent *Plasmodium falciparum*-selective proteasome inhibitors exhibit a low propensity for generating resistance *in vitro* and synergize with multiple antimalarial

agents. *PLOS Pathog.* **15**, e1007722 (2019). doi: 10.1371/journal.ppat.1007722; pmid: 31170268
64. L. A. Kirkman *et al*., Antimalarial proteasome inhibitor reveals collateral sensitivity from intersubunit interactions and fitness cost of resistance. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6863–E6870 (2018). doi: 10.1073/pnas.1806109115; pmid: 29967165
65. B. T. Mott *et al*., High-throughput matrix screening identifies synergistic and antagonistic antimalarial drug combinations. *Sci. Rep.* **5**, 13891 (2015). doi: 10.1038/srep13891; pmid: 26403635
66. C. Dogovski *et al*., Targeting the cell stress response of *Plasmodium falciparum* to overcome artemisinin resistance. *PLOS Biol.* **13**, e1002132 (2015). doi: 10.1371/journal.pbio.1002132; pmid: 25901609
67. K. M. Broadbent *et al*., A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol.* **12**, R56 (2011). doi: 10.1186/gb-2011-12-6-r56; pmid: 21689454
68. G. W. Rangel *et al*., Enhanced *Ex Vivo Plasmodium vivax* Intraerythrocytic Enrichment and Maturation for Rapid and Sensitive Parasite Growth Assays. *Antimicrob. Agents Chemother.* **62**, e02519-17 (2018). doi: 10.1128/AAC.02519-17; pmid: 29378713
69. T. C. de Oliveira *et al*., Genome-wide diversity and differentiation in New World populations of the human malaria parasite *Plasmodium vivax*. *PLOS Negl. Trop. Dis.* **11**, e0005824 (2017). doi: 10.1371/journal.pntd.0005824; pmid: 28759591
70. J. D. Johnson *et al*., Assessment and continued validation of the malaria SYBR green I-based fluorescence assay for use in malaria drug screening. *Antimicrob. Agents Chemother.* **51**, 1926–1933 (2007). doi: 10.1128/AAC.01607-06; pmid: 17371812
71. A. S. Paul, C. K. Moreira, B. Elsworth, D. R. Allred, M. T. Duraisingh, Extensive Shared Chemosensitivity between Malaria and Babesiosis Blood-Stage Parasites. *Antimicrob. Agents Chemother.* **60**, 5059–5063 (2016). doi: 10.1128/AAC.00928-16; pmid: 27246780
72. Q. L. Fivelman, I. S. Adagu, D. C. Warhurst, Modified fixed-ratio isobologram method for studying in vitro interactions between atovaquone and proguanil or dihydroartemisinin against drug-resistant strains of *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **48**, 4097–4102 (2004). doi: 10.1128/AAC.48.11.4097-4102.2004; pmid: 15504827
73. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011). doi: 10.14806/ej.17.1.200
74. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168
75. H. Li *et al*., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
76. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: 10.1093/bioinformatics/btq033; pmid: 20110278
77. M. A. DeJesus, T. R. Ioerger, A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics* **14**, 303 (2013). doi: 10.1186/1471-2105-14-303; pmid: 24103077
78. A. Kiesel *et al*., The BaMM web server for *de-novo* motif discovery and regulatory sequence analysis. *Nucleic Acids Res.* **46**, W215–W220 (2018). doi: 10.1093/nar/gky431; pmid: 29846656
79. L. Li, C. J. Stoeckert Jr., D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003). doi: 10.1101/gr.1224503; pmid: 12952885
80. H.-G. Drost, A. Gabel, I. Grosse, M. Quint, Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol. Biol. Evol.* **32**, 1221–1231 (2015). doi: 10.1093/molbev/msv012; pmid: 25631928
81. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019). doi: 10.1186/s13059-019-1832-y; pmid: 31727128
82. 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). doi: 10.1038/s41587-019-0201-4; pmid: 31375807
83. L. Wang, S. Wang, W. Li, RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012). doi: 10.1093/bioinformatics/bts356; pmid: 22743226
84. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016). doi: 10.1093/bioinformatics/btw354; pmid: 27312411
85. M. Pertea *et al*., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015). doi: 10.1038/nbt.3122; pmid: 25690850
86. G. Pertea, M. Pertea, GFF utilities: GffRead and GffCompare. *F1000 Res.* **9**, 304 (2020). doi: 10.12688/f1000research.23297.1; pmid: 32489650
87. Y.-J. Kang *et al*., CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017). doi: 10.1093/nar/gkx428; pmid: 28521017
88. T. Rausch *et al*., DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012). doi: 10.1093/bioinformatics/bts378; pmid: 22962449
89. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011). doi: 10.1093/bioinformatics/btr509; pmid: 21903627
90. T. Sanderson, J. C. Rayner, PhenoPlasm: A database of disruption phenotypes for malaria parasite genes. *Wellcome Open Res.* **2**, 45 (2017). doi: 10.12688/wellcomeopenres.11896.2
91. S. Ye, umbibio/PkHessentialome: PkH_essentialome, version 1.1, Zenodo (2024); https://doi.org/10.5281/zenodo.13852567.

## ACKNOWLEDGMENTS

We thank the staff at VEuPathDB for their assistance with data deposition. We also thank M. Fraser and D. Degras for helpful discussions and assistance with the manuscript. **Funding:** This study was supported by National Institutes of Health (NIH) grant 5R01AI168163 (M.T.D.), NIH grant 5R01 AI167570 (M.-J.G., M.T.D., and K.Z.), NIH grants ORIP/OD P51OD011132 and U42 PDP11023 (Emory National Primate Research Center), Swiss National Science Foundation Postdoc Mobility fellowships PBSKP3_140144 and P300P3_151146 (C.G.), and the Food and Drug Administration Intramural Research Program (B.E.). **Author contributions:** Conceptualization: B.E., S.Y., S.D., K.Z., M.T.D.; Investigation: B.E., S.Y., S.D., J.A.T., Q.S., B.T.T., U.K., K.Z., M.T.D.; Methodology: B.E., S.Y., S.D., J.A.T., Q.S., B.T.T., A.S.P., U.K., C.G., M.U.F., M.-J.G., K.Z., M.T.D.; Writing – original draft: B.E., S.Y., S.D., K.Z., M.T.D.; Writing – review & editing: B.E., S.Y., S.D., J.A.T., Q.S., B.T.T., A.S.P., U.K., C.G., M.U.F., M.-J.G., K.Z., M.T.D. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All reagents in this study are available from K.Z. and M.T.D. The code used for analysis is available on Zenodo (*91*). Tn-seq and lncRNA-seq data have been deposited to the NCBI Sequence Read Archive (SRA) under the accession no. PRJNA1116591. All data used in this analysis are declared free to access. 

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adq6241
Figs. S1 to S17
Data S1 to S10
MDAR Reproducibility Checklist

Submitted 3 June 2024; accepted 5 December 2024
10.1126/science.adq6241