

# UC Riverside

## UC Riverside Previously Published Works

### Title

Novel insights into the role of long non-coding RNA in the human malaria parasite, Plasmodium falciparum.

### Permalink

<https://escholarship.org/uc/item/4gx791g7>

### Journal

Nature Communications, 14(1)

### Authors

Batugedara, Gayani  
Lu, Xueqing  
Hristov, Borislav  
et al.

### Publication Date

2023-08-22

### DOI

10.1038/s41467-023-40883-w

Peer reviewed

# Novel insights into the role of long non-coding RNA in the human malaria parasite, *Plasmodium falciparum*

---

Received: 19 August 2022

---

Accepted: 10 August 2023

---

Published online: 22 August 2023

---

 Check for updates

---

Gayani Batugedara<sup>1,5</sup>, Xueqing M. Lu<sup>1,5</sup>, Borislav Hristov<sup>2</sup>, Steven Abel<sup>1</sup>, Zeinab Chahine<sup>1</sup>, Thomas Hollin<sup>1</sup>, Desiree Williams<sup>1</sup>, Tina Wang<sup>1</sup>, Anthony Cort<sup>1</sup>, Todd Lenz<sup>1</sup>, Trevor A. Thompson<sup>1</sup>, Jacques Prudhomme<sup>1</sup>, Abhai K. Tripathi<sup>3</sup>, Guoyue Xu<sup>3</sup>, Juliana Cudini<sup>4</sup>, Sunil Dogga<sup>4</sup>, Mara Lawniczak<sup>4</sup>, William Stafford Noble<sup>2</sup>, Photini Sinnis<sup>3</sup> & Karine G. Le Roch<sup>1</sup> ✉

---

The complex life cycle of *Plasmodium falciparum* requires coordinated gene expression regulation to allow host cell invasion, transmission, and immune evasion. Increasing evidence now suggests a major role for epigenetic mechanisms in gene expression in the parasite. In eukaryotes, many lncRNAs have been identified to be pivotal regulators of genome structure and gene expression. To investigate the regulatory roles of lncRNAs in *P. falciparum* we explore the intergenic lncRNA distribution in nuclear and cytoplasmic subcellular locations. Using nascent RNA expression profiles, we identify a total of 1768 lncRNAs, of which 718 (~41%) are novels in *P. falciparum*. The subcellular localization and stage-specific expression of several putative lncRNAs are validated using RNA-FISH. Additionally, the genome-wide occupancy of several candidate nuclear lncRNAs is explored using ChIRP. The results reveal that lncRNA occupancy sites are focal and sequence-specific with a particular enrichment for several parasite-specific gene families, including those involved in pathogenesis and sexual differentiation. Genomic and phenotypic analysis of one specific lncRNA demonstrate its importance in sexual differentiation and reproduction. Our findings bring a new level of insight into the role of lncRNAs in pathogenicity, gene regulation and sexual differentiation, opening new avenues for targeted therapeutic strategies against the deadly malaria parasite.

Malaria, a mosquito-borne infectious disease, is caused by protozoan parasites of the genus *Plasmodium*. Among the human-infecting species, *Plasmodium falciparum* is the most prevalent and deadly, with an estimated 627 000 deaths in 2020<sup>1</sup>. The parasite has a complex life

cycle involving multiple biological stages in both human and mosquito hosts. As sporozoites are transmitted from a *Plasmodium*-infected mosquito to the human bloodstream, they migrate to the liver to invade hepatocytes and initiate parasite amplification. After this

---

<sup>1</sup>Department of Molecular Cell and Systems Biology, University of California Riverside, Riverside, CA 92521, USA. <sup>2</sup>Department of Genome Sciences, University of Washington, Seattle, WA 98195-5065, USA. <sup>3</sup>Department of Molecular Microbiology and Immunology and the Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. <sup>4</sup>Wellcome Sanger Institute, Hinxton CB10 1SA, UK. <sup>5</sup>These authors contributed equally: Gayani Batugedara, Xueqing M. Lu. ✉e-mail: [karinel@ucr.edu](mailto:karinel@ucr.edu)

pre-erythrocytic cycle, which can last 7 to 10 days, tens of thousands of infectious merozoites are released into the bloodstream to invade red blood cells. Within the erythrocyte, the parasite matures from the ring to the trophozoite and to the multinucleated schizont stages. After 48 h, the newly formed merozoites burst out of the erythrocyte to reinfect new red blood cells. This rupture is usually associated with clinical symptoms. During this intraerythrocytic developmental cycle, a subset of the parasites can differentiate into male and female gametocytes. Once ingested by a female *Anopheles* mosquito during a blood meal, these gametocytes undergo sexual replication inside the mosquito's gut to form a zygote that can differentiate into a mobile ookinete and an oocyst. The oocyst will grow and produce thousands of new sporozoites that will migrate to the mosquito's salivary glands ready to infect a new human host during a subsequent blood meal. This multi-stage developmental life cycle leads to distinct morphological and physiological changes in response to altered environmental conditions and is tightly regulated by coordinated changes in gene expression.

Gene expression profiling<sup>2,3</sup> including bulk RNA-seq experiments<sup>2,4-7</sup>, nascent RNA expression profiles<sup>8</sup>, as well as single cell sequencing<sup>9</sup> has revealed that a majority of the genes in the parasite are transcribed in a cascade of gene expression throughout the parasite life cycle but the exact molecular mechanisms regulating these events are largely unknown.

Compared to other eukaryotes with a similar genome size, *P. falciparum* has an extremely AT-rich genome and a relatively low number of sequence-specific transcription factors (TFs), approximately two-thirds of the TFs expected based on the size of the genome. Only 27 apicomplexan apetal2 (ApiAP2) DNA-binding proteins have been identified as specific TFs in the parasite genome. These ApiAP2 are unique to *Apicomplexa*<sup>10</sup> and have been demonstrated to have a major role as activators or repressors of transcription<sup>11</sup>. Our understanding of the regulation of these TFs, and how various TFs could act together to organize transcriptional networks, is still limited but the patterns of gene expression observed are likely the result of a combination of transcriptional<sup>9,12-14</sup> and post-transcriptional regulatory events<sup>15-18</sup>. Additionally, epigenetic studies<sup>19-25</sup> and chromosome conformation capture methods (Hi-C)<sup>26-28</sup> have suggested that the chromatin state and the three-dimensional (3D) genome structure of *P. falciparum* are strongly connected with transcriptional activity of gene families<sup>28</sup>. Machine learning algorithms have also suggested that the ApiAP2 TFs may indeed work in conjunction with epigenetic factors<sup>29</sup>. However, how all the regulators of transcription are recruited to their DNA binding motifs and their chromatin regions remains to be elucidated. Understanding the exact mechanisms regulating the parasite replication life cycle is essential if we want to identify novel therapeutic targets.

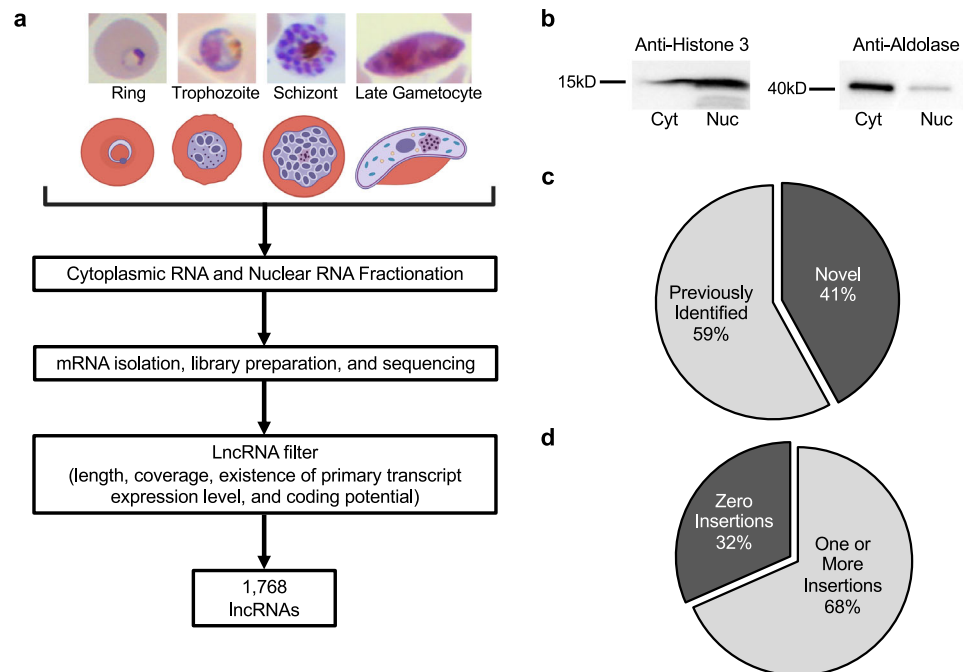
With advances in biotechnology and next generation sequencing technologies, huge strides have been made in genomics studies revealing that the transcriptome of an organism is much larger than expected. In eukaryotes spanning from yeast to human, many non-coding RNAs (ncRNAs) have been detected and linked to diseases ranging from cancers to neurological disorders and are now actively studied for their potential as novel therapeutic and diagnostic agents<sup>30</sup>. Over the past few years, ncRNAs been recognized as key regulators of chromatin states and gene expression<sup>31-33</sup>. One class of ncRNAs, the long noncoding RNAs (lncRNAs), are defined as non-protein coding RNA molecules which are  $\geq 200$  nucleotides in length. Many lncRNAs share features with mature mRNAs including 5' caps, polyadenylated tails as well as introns<sup>34</sup>. lncRNAs are expressed and functionally associated in a cell-type specific manner. Based on their genomic localization, lncRNAs are categorized as sense, antisense, bidirectional, intronic and intergenic<sup>35</sup>. because lncRNAs can bind DNAs, RNAs and proteins, their functions are diverse<sup>34</sup>. lncRNAs enriched in the nuclear fraction often associate with regulation of

transcription<sup>36-39</sup>. By tethering genomic DNA, lncRNAs can control long-range interaction. They can also regulate promoter accessibility by recruiting, guiding or enhancing either TFs or chromatin remodeling enzymes including histone acetyltransferases and methyltransferases. lncRNA have also been shown to interact with spliceosomal factors to affect the frequency and efficiency of mRNA splicing. In the cytosol, lncRNAs can regulate gene expression by mediating mRNA export, RNA stability and translation.

In mammalian systems, the X inactive specific transcript (Xist) is a well-studied example of a lncRNA mediating X-chromosome inactivation during zygotic development<sup>40</sup>. Deposition of Xist on the X-chromosome recruits histone-modifying enzymes that place repressive histone marks, such as H3K9 and H3K27 methylation, leading to gene silencing and the formation of heterochromatin. Two other lncRNAs, the HOX transcript antisense (HOTAIR) and the antisense lncRNA in the INK4 locus (ANRIL), have also been shown to interact with multiprotein Polycomb Protein Complexes (PRC1 and PRC2) to catalyze histone marks and silence gene expression<sup>30</sup>. Similarly, long telomeric repeat-containing lncRNAs (TERRA) have been recently identified as a major component of telomeric heterochromatin<sup>41,42</sup>. With thousands of lncRNAs transcribed in mammalian cells we are only starting to grasp their role in regulating major biological processes.

Although the role of lncRNA in malaria parasite has only been studied more recently, they are emerging as new players in the development of parasite life cycle stages. To date, several studies have already explored the presence of ncRNAs in *P. falciparum*<sup>43-54</sup>. Technological advances including strand-specific and long read sequencing platforms have identified >2500 lncRNA candidates, including 1300 circular lncRNAs<sup>51-53,55,56</sup>. These initial studies confirmed that parasite lncRNAs are developmentally regulated but only a few of these annotated ncRNAs have been functionally characterized. Some have been linked to regulation of virulence genes<sup>57-62</sup>. It has also been established that GC-rich ncRNAs serve as epigenetic regulatory elements that play a role in activating *var* gene transcription as well as several other clonally variant gene families<sup>63</sup>. In addition, a family of twenty-two lncRNAs transcribed from the telomere-associated repetitive elements (TAREs) has been identified in the parasite<sup>45,47,59</sup>. These TARE-lncRNAs show functional similarities to the eukaryotic family of non-coding RNAs involved in telomere and heterochromatin maintenance<sup>64</sup> and could have a role in regulating virulence factors. More recently, the functional characterization of two lncRNAs, *gdu1*-as-lncRNA and *md1*-lncRNA that were detected during gametocytogenesis, has revealed that sexual differentiation and sex determination in *P. falciparum* is at least partially regulated by lncRNAs<sup>65,66</sup>. While it is becoming evident that lncRNAs serve as an integral part of the mechanisms regulating gene expression in *Plasmodium*, the localization and function of most of the identified lncRNAs remain a mystery.

Here, to investigate the localization and subsequently the potential role of lncRNAs in *P. falciparum*, we explore the intergenic lncRNA distribution separately in nuclear and cytoplasmic subcellular locations. Using deep sequencing and nascent RNA expression profiles<sup>8</sup>, we identify a total of 1768 lncRNAs, of which 41% are novels in *P. falciparum*. We further validate the subcellular localization and stage-specific expression of several putative lncRNAs using RNA fluorescence in situ hybridization (RNA-FISH) and single-cell RNA sequencing (scRNA-seq). Additionally, the genome-wide occupancy of 7 candidate nuclear lncRNAs is explored using Chromatin Isolation by RNA Purification followed by deep sequencing (ChIRP-seq). Our ChIRP-seq experiments on our candidate lncRNAs reveal that lncRNA occupancy sites within the parasite genome are sequence-specific with a particular enrichment for several parasite-specific gene families, including those involved in pathogenesis, remodeling of the red blood cell, and regulation of sexual differentiation. We also demonstrate that the presence of some of these lncRNAs correlates with changes in gene



**Fig. 1 | Nuclear and cytoplasmic lncRNA identification.** **a** A general overview of the lncRNA identification pipeline. Created with BioRender.com. **b** Validation of cell fractionation efficiency using anti-histone H3 and anti-aldolase as nuclear (Nuc) and cytoplasmic (Cyt) markers. Blot is representative of two independent biological

replicates. **c** Comparison of lncRNA candidates with lncRNAs identified from previous publications. **d** Essentiality of lncRNAs using piggyBac insertion (Zhang et al., 2018). lncRNAs that cannot be disrupted are more likely to be essential.

expression demonstrating that these lncRNAs can possibly work in cooperation with TFs and epigenetic factors. We further validate the role of a lncRNA identified as enriched in gametocytes. Using the CRISPR-Cas9 editing tool, we functionally characterize lncRNA-ch14 and validate its role during sexual differentiation and development, particularly affecting female gametocytes. Transmission studies demonstrate that even partial deletion of this lncRNA significantly affects parasite development throughout all mosquito stages. Collectively, our results provide evidence that in addition to being developmentally regulated, lncRNAs are distributed in distinct cellular compartments in *P. falciparum*. Depending on their nuclear or cytoplasmic localization, they may play important roles in gene regulation at the transcriptional or translational levels respectively, ultimately regulating the malaria parasite life cycle progression.

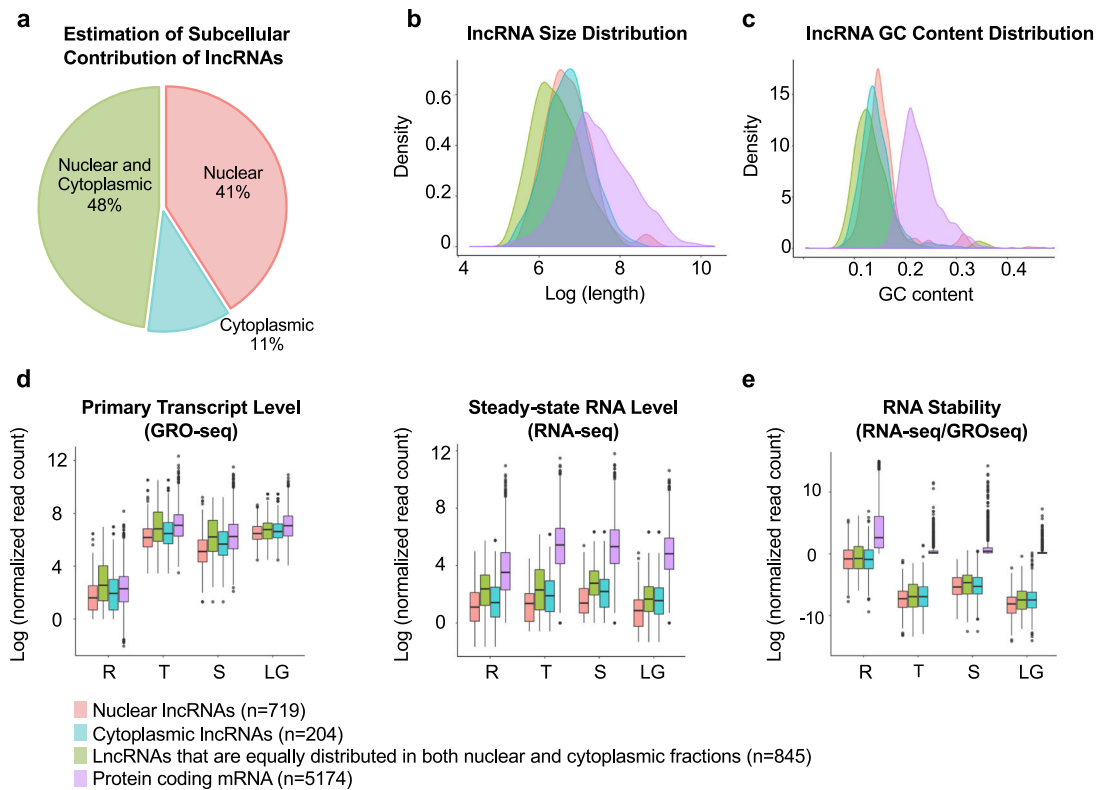
## Results

### Identification of lncRNAs

To comprehensively identify lncRNA populations in *P. falciparum*, we extracted total RNA from both nuclear and cytoplasmic fractions using synchronized parasite cultures at early ring, early trophozoite, late schizont, and gametocyte stages (Fig. 1a). The samples collected here allow for gene expression profiling during the critical processes of parasite egress, invasion, and sexual differentiation. In brief, extracted parasites were subjected to a modified cell fractionation procedure described in the PARIS kit (ThermoFisher) (see “Methods”). Successful isolation of both subcellular fractions was validated using western blot with an anti-histone H3 antibody as a nuclear marker and an anti-aldolase antibody as a cytoplasmic marker (Fig. 1b). After separation of nuclear material from the cytoplasmic material, total RNA and subsequent polyadenylated mRNA was isolated from both fractions. Strand-specific libraries were then prepared and sequenced (see methods for details, Supplementary data 1a). For verification, Spearman correlations in gene expression levels were calculated among nuclear samples and cytoplasmic samples<sup>67</sup> (Fig. S1). Once validated, a computational pipeline was

implemented for the identification of lncRNAs. Briefly, all nuclear and cytoplasmic RNA libraries were merged, resulting in one nuclear and one cytoplasmic merged file, then assembled into nuclear and cytoplasmic transcriptomes independently using cufflinks. Subsequently, transcripts were filtered based on length, expression level, presence of nascent transcript from previously published GRO-seq dataset<sup>8</sup>, and sequence coding potential (Fig. 1a). To specifically identify lncRNA candidates within the intergenic regions and avoid any potential artefacts introduced by PCR amplification of the AT rich genome, we removed any predicted transcripts that have at least 30% overlap with annotated genes. Our goal was to select transcripts that are  $\geq 200$  bp in length, consistently expressed in both published nascent RNA and steady-state RNA expression profiles, and that are likely to be non-protein-coding genes. As a result, we identified a total of 1768 intergenic lncRNAs in *P. falciparum* irrespective of the developmental stage (Supplementary data 1b). Nine hundred fifty-one lncRNAs have no overlap with any UTR regions. Overall, 1050 lncRNAs (~59%) overlapped with previously identified intergenic lncRNAs<sup>49,51–53,55,56</sup> and 718 lncRNAs were identified as novel in *P. falciparum* (Fig. 1c).

To evaluate the essentiality of the lncRNAs identified in this study, we used piggyBac insertion sites from Zhang and colleagues<sup>68</sup>. In this work, the authors used a high-throughput transposon insertional mutagenesis method to distinguish essential and dispensable genes in the *P. falciparum* genome during the asexual stages of the parasite life cycle. We focused our analysis on the integration of the transposon that occurred within each of the identified lncRNAs. The piggyBac insertion site coordinates were overlapped with the genomic ranges for all detected lncRNAs. This was performed after accounting for differences in the parasite strains used in the two studies. Overall, we were unable to uncover piggyBac insertion for 558 lncRNAs (31.6%), suggesting that these lncRNAs are either difficult to disrupt or are potentially essential for the parasite asexual development (Fig. 1d). Because we observed an insertion in 292 lncRNAs (16.5%) that were specifically detected in gametocyte stage, it will be important to



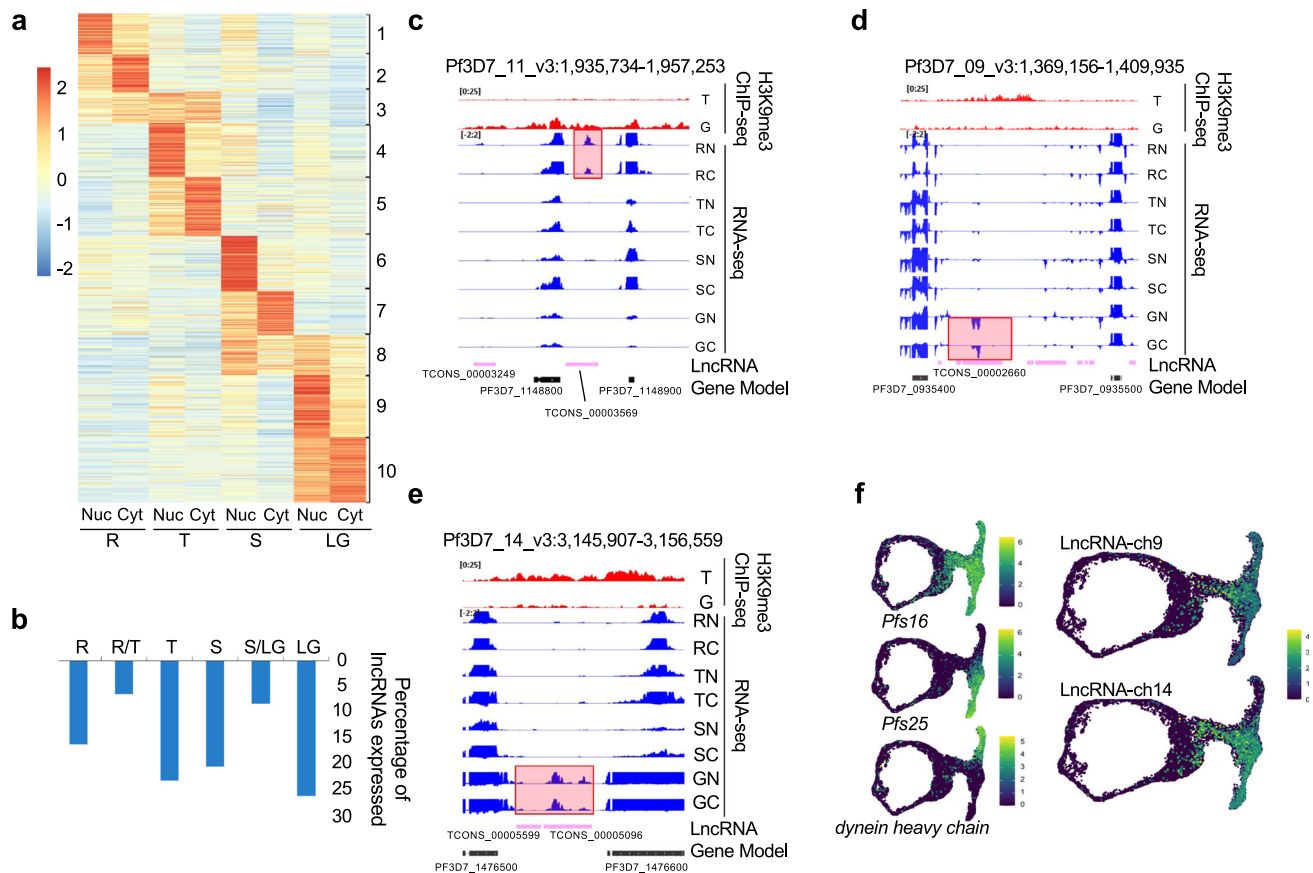
**Fig. 2 | Candidate lncRNA categorization.** **a** A total of 1768 lncRNA candidates were identified, covering 719 nuclear enriched lncRNAs (red), 204 cytoplasmic enriched lncRNAs (blue), and 845 lncRNAs found in both fractions (green). Cellular distribution of predicted lncRNAs is based on  $\log_2$  fold change  $>$  or  $<$  0.5 of summed nuclear vs cytoplasmic expression level. Density plots of size (**b**) and GC content (**c**) of lncRNA candidates and annotated protein encoding mRNAs (purple). **d** Expression levels of primary transcripts (left), steady-state RNA (middle) of

lncRNA candidates and annotated protein encoding mRNAs for Ring (R), Trophozoite (T), Schizont (S) and Late Gametocyte (LG) stages. **e** Relative stability of lncRNA candidates and annotated protein encoding mRNAs. The stability is based on the ratio of RNA-seq/GRO-seq transcript level. Each box represents the 25/75 percentiles, the line across the box represents the median, and the whiskers represent maximum and minimum values. Outliers are indicated with dots.

validate at the phenotypic level whether those lncRNAs are essential during sexual differentiation rather than the asexual cycle. Additionally, significantly fewer insertions per possible insertion site (TTAA sequence) were found for telomeric as compared to subtelomeric lncRNAs, and for subtelomeric as compared to other lncRNAs (Fig. S2a). This suggests that piggyBac insertions in telomeric and subtelomeric lncRNAs are also either difficult to disrupt due to their co-localization with heterochromatin or are more likely to be essential than others. It was also found that the 5' flanking regions of lncRNAs (Fig. S2b) had more insertions per possible site as compared to the rest of the lncRNAs, suggesting that, as a general trend, these regions of the lncRNAs are the most disposable while the gene body and 3' flanking regions may be more important for their function. Further analysis of the 558 lncRNA with zero piggyBac insertions illustrated that 158 of them (28.3%) overlapped by at least 1 bp with a gene, including UTRs, found to be essential in the Zhang study. For the remaining 400 lncRNAs (71.7%), the lack of insertions cannot be tied to a nearby gene rather than the lncRNA itself. In addition, novel lncRNAs were found to have fewer insertions (and thus be more likely essential) than previously known lncRNAs (38.6% of novel lncRNAs had zero insertions, whereas 26.8% of previously known lncRNAs had zero insertions, and novel lncRNAs had fewer insertions per TTAA site (0.184 vs. 0.204)). Thus, many lncRNAs identified in this study may be essential for parasite survival. Although additional experiments will be needed to validate these results, it is also possible that some of the genes found to be essential in the Zhang study including some of the *rifin*, *stevor*, and pseudogenes, could be due to extensive overlap with the identified lncRNAs.

### Length, GC content, and RNA stability of cytoplasmic and nuclear lncRNAs

lncRNAs exhibit diverse subcellular distribution patterns, ranging from nuclear foci to cytoplasmic localization. Their localization patterns are linked to their distinct regulatory effects at their site of action<sup>69,70</sup>. Therefore, to better understand the potential function of the lncRNA in *Plasmodium*, we categorized the subcellular localization of our candidate lncRNAs into nuclear lncRNAs, cytoplasmic lncRNAs, or indistinguishable lncRNAs that are equally distributed in both fractions. Among the total identified 1768 lncRNAs, 719 lncRNAs (41%) were enriched in the nuclear fraction, 204 lncRNAs (11%) were enriched in the cytoplasmic fraction, and 845 lncRNAs (48%) showed similar distribution between both subcellular fractions (Fig. 2a). Further, we explored the physical properties of these lncRNAs. We observed that lncRNAs are in general shorter in length and less GC-rich as compared to protein-encoding mRNAs (Fig. 2b, c). Using total steady-state RNA expression profiles and nascent RNA expression profiles (GRO-seq) (Fig. 2d), we then estimated the expression levels and stability of the lncRNAs. RNA stability was calculated as the ratio between steady-state RNA expression levels over nascent RNA expression levels. We discovered that, although the overall life cycle gene expression pattern of the lncRNAs is similar to the expression pattern of coding mRNAs, lncRNAs are less abundant and less stable than coding mRNAs; nuclear lncRNAs are particularly lowly expressed and unstable as compared to the other two groups of lncRNAs (Fig. 2e). These observations are consistent with previous lncRNA annotation studies in human breast cancer cells<sup>71</sup> and noncoding RNA stability studies in mammalian genomes<sup>72</sup>. Our results suggest that the low expression level and the



**Fig. 3 | Gene expression pattern of lncRNAs.** **a** lncRNAs are grouped into 10 clusters based on their life cycle expression patterns in the Nuclear (Nuc) or Cytoplasmic (Cyt) fraction for the Ring (R), Trophozoite (T), Schizont (S) and Late Gametocyte (LT) stages. **b** Percentage of lncRNAs that are highly expressed at ring, trophozoite, schizont, and late gametocyte stages. Genome browser views of H3K9me3 ChIP-seq and RNA-seq datasets in the region of a representative asexual stage-specific lncRNA on chromosome 11 (**c**), two gametocyte-specific lncRNAs located at the intergenic regions of chromosome 9, lncRNA-ch9 (**d**) and 14, lncRNA-ch14 (**e**). RNA-seq data of Nuclear (N) and Cytoplasmic (C) fraction at Ring (R),

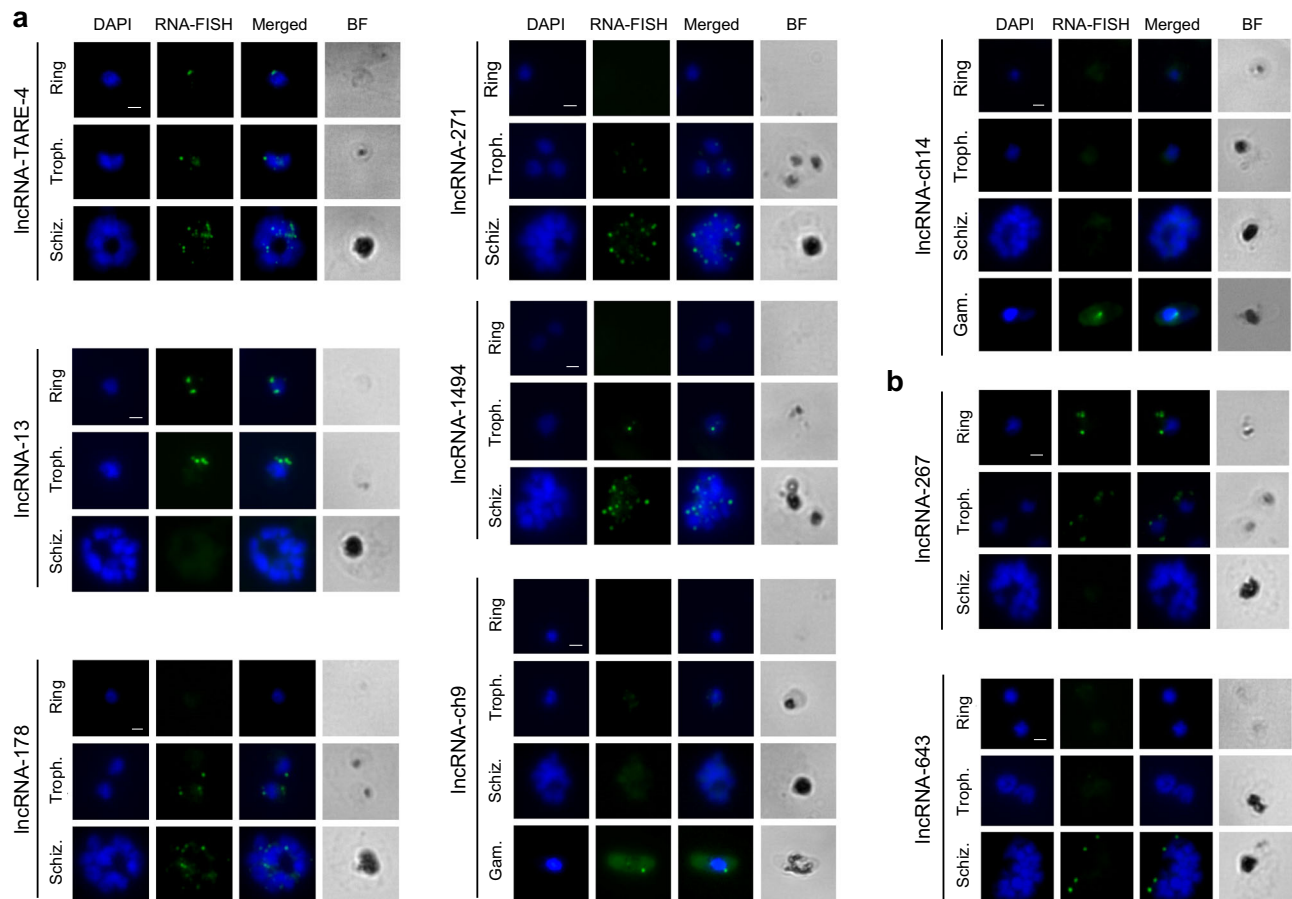
Trophozoite (T), Schizont (S) and Gametocyte (G) stages are shown (e.g., RN=nuclear fraction at ring stage). **f** scRNA-seq analysis. 2-dimensional UMAP projection of *P. falciparum* parasites, both asexual (ring) and sexual (T-shape). Each dot represents a single cell. Left panel: Cells colored according to log-normalized gene expression values for gametocytes (*Pfs16* (PF3D7\_0406200), top), females (*Pfs25* (PF3D7\_1031000), middle), and males (*dynein heavy chain* (PF3D7\_0905300), bottom). Right panel: Log-normalized expression of lncRNA-ch9 (top) and lncRNA-ch14 (bottom) across the *P. falciparum* life cycle.

low stability of these lncRNAs may be the reason why they failed to be detected in previous identification attempts. By taking advantage of primary transcripts detected in our GRO-seq dataset, we significantly improved the sensitivity of lncRNA detection, especially for those localized in the nuclear fraction and expressed at a lower level.

#### Stage-specific expression of cytosolic and nuclear lncRNAs

As lncRNAs often exhibit specific expression patterns in a tissue dependent manner, we investigated the stage specificity of identified candidate lncRNAs across the asexual and sexual life cycle stages. Using k-means clustering, we were able to group lncRNAs into 10 distinct clusters (Fig. 3a). Generally, nearly all lncRNAs showed a strong coordinated cascade throughout the parasite's life cycle. Similar to what was observed with mRNA, a large fraction of the lncRNAs was highly expressed at mature stages compared to the ring stages (Fig. 3b). Cluster 1 contains lncRNAs that are more abundantly expressed in the nuclear fraction of ring stage parasites and are lowly expressed in the nuclear fraction of schizont stage parasites. lncRNAs representative of this cluster are the lncRNA-TAREs. We observed that most lncRNA-TAREs identified in this study (19 out of 21) are clustered into this group with an average expression of 1.18  $\log_2$  fold change of nuclear to cytoplasmic ratio (Fig. 3a). The remaining two identified lncRNA-TAREs were found in cluster 6, where transcription peaks at the schizont stage. This finding validates our approach and suggests that lncRNAs in

this cluster may contribute to the maintenance and regulation of chromatin structure and telomere ends. Approximately 28% of the identified lncRNAs are more abundantly found in either the nuclear or cytoplasmic fraction at the schizont stage (cluster 6, 7 and 8), after DNA replication and the peak of transcriptional activity observed at the trophozoite stage. We observed a few lncRNAs that are solely expressed during the asexual cycle with distinct changes in heterochromatin marks (Fig. 3c). Based on clustering analysis, we also found that 460 of our detected lncRNAs are exclusively expressed at a high level at the gametocyte stage (cluster 9 and 10). Interestingly, two unique lncRNAs in this cluster, lncRNA-ch9 (PF3D7\_09\_v3:1,384,241-1,386,630) and lncRNA-ch14 (PF3D7\_14\_v3:3,148,960 - 3,150,115), were identified in a previous study to be located within heterochromatin regions marked by repressive histone marks H3K9me3 at the trophozoite stage<sup>28</sup> (Fig. 3d and 3e). At the gametocyte stage however, the H3K9me3 was lost. Additionally, both lncRNAs are transcribed from regions adjacent to gametocyte-specific genes. To validate the expression of these gametocyte-specific lncRNAs, we performed RT-PCR (Fig. S3 and Supplementary data 1c) as well as single cell RNA-seq (scRNA-seq) across key-stages of the parasite life cycle. lncRNA expression was visualized on the UMAP embedding generated from coding gene expression (Fig. 3f). lncRNA-ch9 and lncRNA-ch14 were expressed in sexual-stage parasites, with lncRNA-ch9 and lncRNA-ch14 were expressed in sexual-stage parasites, with specific enrichment in male



**Fig. 4 | RNA-FISH experiments to show localization of several candidate lncRNAs. a** Nuclear lncRNAs (lncRNA-TARE4, lncRNA-13, lncRNA-178, lncRNA-271, lncRNA-1494, lncRNA-ch9 and lncRNA-ch14) colocalize with nuclei stained with DAPI in Ring, Trophozoite (Troph.), Schizont (Schiz.) and Gametocyte (Gam.)

stages. **b** Cytoplasmic lncRNA-267 and lncRNA-643 do not colocalize with the nuclei stained with DAPI. Scale bar indicates 2  $\mu$ m. Hybridization images are representative of approximately 15 stained parasites from two independent experiments. BF brightfield.

and female gametocytes, respectively (Fig. 3f, right panel). Collectively, these results emphasize the stage-specific expression of parasite lncRNAs and the potential function of gametocyte-specific lncRNAs in regulating sexual development.

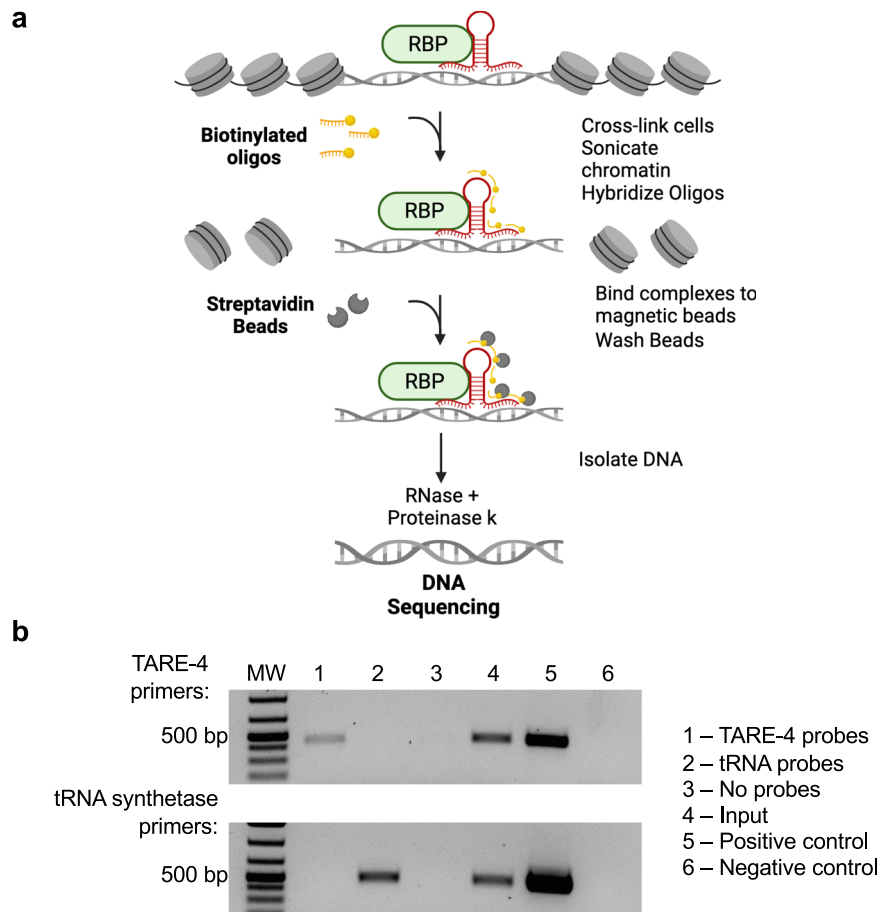
#### Validation of lncRNA localization and stage-specific expression

To validate the cellular localization of several candidate lncRNAs, we utilized RNA fluorescence in situ hybridization (RNA-FISH). We selected two candidates that were detected as enriched in the cytoplasmic fraction (lncRNA-267 (Pf3D7\_08\_v3:1382128-1382689) and lncRNA-643 (Pf3D7\_14\_v3:1606672-1607587)), four candidates detected as enriched in the nuclear fraction of asexual parasites (lncRNA-13 (Pf3D7\_01\_v3:491225-494291), lncRNA-178 (Pf3D7\_06\_v3:53758-54745), lncRNA-271 (Pf3D7\_02\_v3:590844-592940) and lncRNA-1494 (Pf3D7\_06\_v3:1311694-1312858)) and two candidates detected as enriched in the nuclear fraction of sexually mature gametocytes (lncRNA-ch9 and lncRNA-ch14). Finally, we also selected a lncRNA that had been previously identified and known to be transcribed from the telomere region on chromosome 4, termed lncRNA-TARE4 (Pf3D7\_04\_v3:1194786-1199684)<sup>47</sup> as our positive control. Briefly, mixed stage parasites were fixed and hybridized to fluorescently labeled -200-300 nucleotide antisense RNA probes (Supplementary data 1c, d). The hybridization images, representative of 15-20 parasites, demonstrate that the nuclear lncRNAs localize to distinct foci within or close to the DAPI-stained nuclei (Fig. 4a), while cytoplasmic lncRNAs are localized outside the DAPI-stained genomic DNA (Fig. 4b). Additionally, using RNA-FISH, we validated the stage-specific expression of

our candidate lncRNAs. Specifically, expression of lncRNA-267 and lncRNA-13 were enriched at the ring and trophozoite stages; lncRNA-178 was expressed at the trophozoite and schizont stages; lncRNA-643 was expressed at the schizont stage only and lncRNA-TARE4 was expressed at all three asexual stages. lncRNA-ch9 and lncRNA-ch14 were only expressed at the gametocyte stage. These results highlight that, similar to protein-coding transcripts, these candidate lncRNAs are developmentally regulated.

#### Genomic maps of RNA-chromatin interactions

The locations of the binding sites of most lncRNAs remain unknown. Accordingly, the role of lncRNAs in chromatin and gene regulation in eukaryotes, including the malaria parasite, has been mostly deduced from the indirect effects of lncRNA perturbation. To explore the role of lncRNAs in gene expression, we sought to identify occupancy sites of our selected candidate lncRNAs within the parasite genome. For an unbiased high-throughput discovery of RNA-bound DNA in *P. falciparum*, we adapted a method termed Chromatin Isolation by RNA Purification (ChIRP) (Fig. 5a)<sup>73,74</sup>. ChIRP-seq is based on affinity capture of target lncRNA:chromatin complex by tiling antisense-biotinylated-oligos to allow the identification of lncRNA-DNA binding sites at single base-pair resolution with high sensitivity and low background<sup>75,76</sup>. Such experiments can identify whether lncRNAs are working in *cis* on neighboring genes or in *trans* to regulate distant genes. ChIRP-seq is applicable to all detected lncRNAs and requires no knowledge of the RNA's structure. This method has recently been used in *Plasmodium* to investigate the role of ncRNA RUF6 on heterochromatin formation<sup>76</sup>. In



**Fig. 5 | Chromatin Isolation by RNA Purification (ChIRP).** **a** Schematic representation of the ChIRP methodology. Created with BioRender.com. RBP: RNA-binding protein. **b** RT-PCR following the ChIRP protocol validates the specificity of the biotinylated antisense oligonucleotides. Following the ChIRP-pull-down, the RNA fraction was analyzed, and RT-PCR results confirm that lncRNA-TARE-4 probes retrieve the lncRNA-TARE-4 RNA (535 bp PCR product) and the control serine tRNA ligase

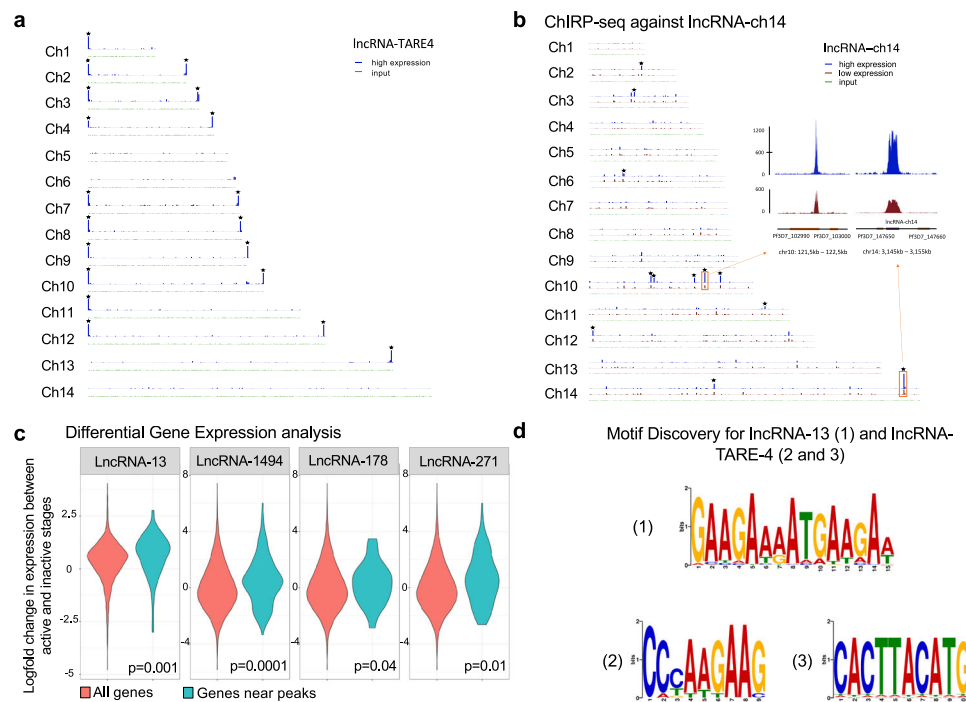
probes retrieve the serine tRNA ligase RNA (505 bp PCR product), respectively. No RNA was retrieved in the no probe control. RNA from a ChIRP-input sample as well as WT 3D7 parasites (wells 4 and 5, respectively) was used to confirm the lncRNA-TARE-4 and serine tRNA ligase primers. The negative controls (well 6) represent no template controls. RT-PCRs are representative of two independent replicates.

our experiments, synchronized parasites were extracted and cross-linked. Parasite nuclei were then extracted, and chromatin was solubilized and sonicated. Biotinylated antisense oligonucleotides tiling our candidate lncRNAs (Supplementary data 1e, f) were hybridized to target RNAs and isolated using magnetic beads. These candidates correspond to the seven nuclear lncRNAs validated using RNA-FISH: lncRNA-TARE4, lncRNA-13, lncRNA-178, lncRNA-1494 and lncRNA-271 detected in the nuclear fraction of the asexual stages, as well as lncRNA-ch9 and lncRNA-ch14 detected in the nuclear fraction of the sexual stages. To validate the specificity of the biotinylated oligonucleotides to target our RNA of interest, we performed RT-PCR following our RNA pull-down. RT-PCR results confirmed that lncRNAs and control serine tRNA ligase probes retrieve the selected lncRNA and the serine tRNA ligase RNA (PF3D7\_0717700), respectively (Fig. 5b). No RNA was retrieved in the negative controls that were incubated with no probes or templates. These results confirm that the biotinylated probes target the RNA of interest with specificity. Purified DNA fragments were then sequenced using next-generation sequencing technology. An input control was used to normalize the signal from ChIRP enrichment.

For all seven lncRNAs, ChIRP-seq experiments were performed in duplicate at the stages when the lncRNA was either highly or lowly expressed. We also generated several input controls for each analyzed stage to demonstrate the consistency of our coverage in various heterochromatin and euchromatin regions (Fig. S4). For all lncRNAs

investigated here, ChIRP-seq performed at time points where the lncRNAs were least expressed retrieved low to no significant signals (Fig. 6a, b and Fig. S4). The lncRNA-TARE4, transcribed from the telomere region on chromosome 4, and expressed throughout the parasite life cycle stages investigated in this study was found to strongly interact with most telomeres in a very specific manner (Fig. 6a). Interestingly, one focus per cell for lncRNA-TARE4 was detected by RNA-FISH experiment at the ring stage. One focus per cell was also observed by IFA using an antibody against histone H3K9me3, known to also localize in the telomeres, as well as the subtelomeric and internal *var* gene cluster of the parasite genome<sup>28,65</sup> (Fig. S5). This data correlates nicely with the Hi-C data published previously<sup>27,28,77</sup> that demonstrate that most telomere ends, including *var* genes, interact with each other in a large heterochromatin cluster before and after DNA replication. lncRNA-13, transcribed on chromosome 1 and highly expressed at the trophozoite stage (Fig. S4), was found to be around surface antigen genes, including PF3D7\_0113100 (SURFIN4.1) and PF3D7\_1149200 (RESA, ring-infected erythrocyte surface antigen). Both the SURFIN and RESA families of proteins have been implicated in erythrocyte invasion-related processes and are transcribed in mature-stage parasites<sup>78,79</sup>. Given that a trophozoite stage lncRNA was identified adjacent to surface antigen genes which are transcribed at the schizont stage, lncRNA-13 is possibly playing a role in recruiting chromatin modifying enzymes to edit the epigenetic state of the chromatin and allow the recruitment of transcription factor(s) needed





**Fig. 6 | ChIRP-seq reveals candidate lncRNA binding sites. a** Genome-wide binding sites of lncRNA-TARE-4 and **(b)** lncRNA-ch14. Mapped, normalized reads from the active stage (top, blue track) and inactive stage (bottom, red track) are shown for each chromosome (Ch). Significant peaks are highlighted with an asterisk. **c** Differential gene expression analysis. The log<sub>2</sub>-fold change of gene expression was calculated for the genes closest to the lncRNA peaks between the inactive and active stage (right violin). This distribution was compared to the log<sub>2</sub> fold change in expression for all other genes in the *Plasmodium* genome (left

violin) using a two-sided *t* test, and the *p* values are reported at the bottom of each panel. **d** Motif identification. 100 bp sequences centered at the peaks' summits were extracted, and we used STREME specifying 2nd order Markov model and default for the rest of parameters to search for possible motifs. We identified one motif for lncRNA-13 ( $p = 1.8e^{-3}$ ) occurring in 131 of the 138 (95%) lncRNA-13 sequences and two motifs for lncRNA-TARE-4, ( $p = 3.7e^{-6}$  and  $p = 5.1e^{-5}$ ), occurring in 72% and 61%, respectively, of the TARE binding areas.

to activate the transcription of these genes. To further investigate the possible link between lncRNAs and their role in epigenetics and regulation of gene expression, we developed a software pipeline in Python to identify all specific binding sites in the genome using Bowtie for mapping and PePr for peak calling<sup>80</sup>. We then aligned lncRNA ChIRP-seq signals across all 5' and 3' UTRs as well as the gene bodies. Similar to what was detected in higher eukaryotes, we discovered that the lncRNA occupancy is enriched either in the gene bodies for lncRNA-13 and lncRNA-178 or near the end of the 5' UTR of each gene (lncRNA-1494 and lncRNA-271) (Fig. S6). While the data will need to be further validated at the molecular level, this pattern provides support for our candidate lncRNAs to promote either transcriptional elongation or transcriptional initiation, respectively. We next retrieved the genes closest to the identified lncRNA binding sites and calculated the log<sub>2</sub> fold change of their expression from inactive to active stage. We compared the resultant information to the change in the expression profiles for all other genes in the *P. falciparum* genome (Fig. 6c). For each of the lncRNAs investigated, we detected a significant increase in the expression of the genes near the ChIRP signals. These results indicate that overall, the presence of lncRNA correlates with a significantly increased gene expression. To further demonstrate that lncRNAs interact specifically with DNA, we looked for motif enrichment (Fig. 6d). Motif analysis of ChIRP-seq data revealed one motif for lncRNA-13 ( $p$ val =  $1.8e^{-3}$ ) occurring in 131 of the 138 retrieved lncRNA-13 sequences and two motifs for lncRNA-TARE-4 ( $p$ val =  $3.7e^{-7}$  and  $p$ val =  $5.1e^{-5}$ ) occurring in 72% and 61% of the TARE-4 binding sites. These data demonstrate that our ChIRP-seq experiments were highly sensitive and specific, and that we were able to retrieve biological insights into their function.

We then focused our attention on ChIRP-seq data generated using probes against lncRNA-ch9 and lncRNA-ch14, two lncRNAs enriched in

gametocytes. ChIRP-seq signals (Fig. 6b, Fig. S4, and Supplementary data 1k, l) showed significant enrichment in the genomic regions where the lncRNAs are transcribed. The lncRNA-ch9 lie between genes that have been implicated in gametocyte differentiation. These include PF3D7\_0935500<sup>81</sup>, a *Plasmodium* exported protein of unknown function, PF3D7\_0935600, a gametocytogenesis-implicated protein, and PF3D7\_0935400, Gametocyte development protein 1. These three genes are known to be significantly up regulated in gametocytes and have been demonstrated to be essential to sexual commitment<sup>65</sup>. For lncRNA-ch14, the genes are PF3D7\_1476500, a probable protein of unknown function, PF3D7\_1476600<sup>82</sup>, a *Plasmodium* exported protein of unknown function and PF3D7\_1476700, a lysophospholipase, three genes on chromosome 14 that are only detected either in gametocyte or ookinete stages. When overlaid with previous ChIP-seq data generated against histone H3K9me3 during the asexual and sexual stages of the parasite life cycle, we noticed that the presence of these lncRNAs correlate with a loss of H3K9me3 marks at the gametocyte stage. For lncRNA-ch14, 11 additional peaks were detected as statistically significant in the gametocyte stage (Fig. 6b, and Supplementary data 1). Most of these peaks were identified in the promoters of genes that were described as conserved *Plasmodium* protein of unknown function but were also known to be expressed in gametocyte including PF3D7\_1145400, a dynamin-like protein overexpressed in female gametocytes<sup>83</sup>. While these data will need to be further validated, our results suggest that these lncRNAs may recruit histone demethylase and/or histone acetyl transferase to change the epigenetic state of the chromatin and activate the expression of these genes during sexual differentiation. Collectively, these experiments propose that lncRNAs in the parasite could be essential to recruit chromatin remodeling and modifying enzymes as well as sequence-specific transcription factors to regulate gene expression.

## Role of lncRNA-ch14 in sexual commitment and development

We next sought to validate the role of one lncRNA in the parasite development. We therefore selected lncRNA-ch14, which was detected as upregulated in female gametocytes. We began by disrupting its full length via the CRISPR-Cas9 editing tool. After four unsuccessful attempts, we concluded that either the system was not able to target this genomic region or this large chromosomal deletion was lethal to the parasite. However, we were able to successfully disrupt the lncRNA-ch14 gene through insertion of a resistance marker spanning the position (Ch14:3,148,960–3,150,115) of the gene (Fig. S7a). Parasite lncRNA14 disruptive lines (two clones, named  $\Delta$ lncRNA-ch14 B1 and F2) were recovered and validated via PCR and RT-PCR (Fig. S7b). We also confirmed the absence of obvious off-target effects via whole genome sequencing (WGS). We then examined parasite growth using our two selected clones along with wild-type NF54 parasites during the erythrocytic cycle. Growth was monitored in triplicates using Giemsa-stained blood culture smears for two full cycles. No significant difference was observed in the asexual stages of the  $\Delta$ lncRNA-ch14 clones compared to the WT (Fig. S8a). This indicates that partial disruption of lncRNA-ch14 does not have a significant role in the asexual blood stage replication.

We subsequently aimed to analyze the effect of  $\Delta$ lncRNA-ch14 in gametogenesis. Relative gametocyte numbers were determined by microscopic examination of Giemsa-stained blood smears prepared from day 16 gametocyte cultures. To consider the impact prolonged culturing times may have on gametogenesis, the assays were conducted between our two  $\Delta$ lncRNA-ch14 clones as well as two NF54 strains; a NF54 WT lab strain as well as the NF54 parental line used for the initial transfection. The NF54 parental line was maintained in culture in parallel with our selected  $\Delta$ lncRNA-ch14 clones. We detected a significant decrease in mature stage V gametocytes in clones B1 and F2 compared to the WT NF54 line ( $n=4$ ,  $p<0.05$ ). This decrease was however not detected as significant between the parental NF54 line and  $\Delta$ lncRNA-ch14 clones (Fig. 7a). To better understand discrepancies observed between the NF54 lines, we purified gDNA for WGS. While we confirmed successful disruption of ncRNA-ch14 in our two selected clones, we also identified a nonsense mutation in the gametocyte developmental protein 1 (*gdu1*) gene (PF3D7\_0935400), a gene essential in sexual differentiation, in the parental line. The mutation detected indicated a premature stop codon leading to a C-terminal truncation of 39 amino acids (GDV39) similar to what was previously observed by Tibúrcio and colleagues<sup>84,85</sup>. Of the 110 reads covering the mutation site, 43 were shown to retain the reference base while 67 displayed the *gdu1* mutation described. This result suggests the development of a spontaneous mutation after transfection and that the reduced number of gametocytes observed in the parental line were most likely a result of a significant portion (60%) of parasites with the *gdu1* mutation, not producing mature gametocytes. This mutation was however absent in both of our  $\Delta$ lncRNA-ch14 clones as well as our NF54 WT (Supplementary data 2), explaining the discrepancies observed in our gametocyte induction assays with the NF54 parental line. Development of spontaneous mutations in culture attests the need of WGS to validate the phenotypes observed in the parasites for both the WT and genetically modified strains. As the parental line was still capable of generating healthy mature gametocytes, albeit at a lower frequency, and gametocytemia is normalized prior to mosquito feeds, the use of the NF54 parental line as our control to analyze the impacts of lncRNA-ch14 disruption on transmission to the mosquito was not considered to be a major issue because GDV1 is only essential for early sexual commitment. Our reasoning was that the gametocytes produced from the parental line would still closely resemble the  $\Delta$ lncRNA-ch14 clones at the genomic level.

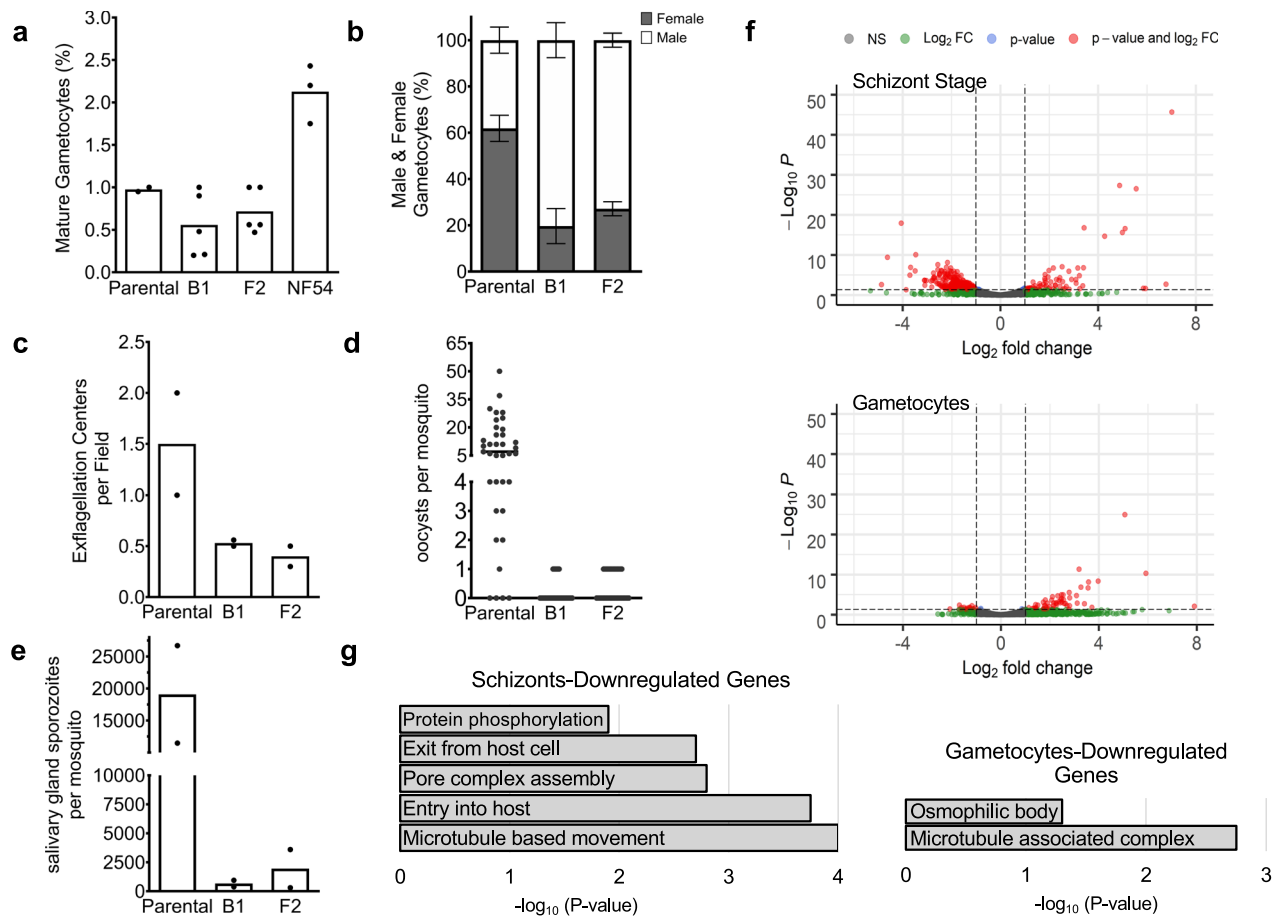
We then analyzed the formation and ratio of mature male and female gametocytes between the parental line and  $\Delta$ lncRNA-ch14 clones. To discriminate between male and female gametocytes, blood

smears prepared from day 16 gametocyte cultures were stained with Giemsa and  $\geq 100$  mature stage V gametocytes were counted to determine sex ratio in each line. Male gametocytes can be distinguished from their female counterparts as they are less elongated with rounder ends and their cytoplasm is distinctly pink while the cytoplasm of female gametocytes, with their large stores of RNA, are dark blue. As shown in Fig. 7b, the male to female ratio was significantly affected in the  $\Delta$ lncRNA-ch14 clones compared to control parasites. In control lines the ratio of female to male gametocytes was approximately 2:1, with the expected larger number of females than males. In contrast, in the lncRNA-ch14 clones it was approximately 2:7, with significantly more males than females ( $n=2$ ,  $p<0.05$ ) (Fig. 7b). These data suggest that lncRNA-ch14 has a role in the ratio of male and female gametocytes produced under our culture conditions. Exflagellation assays revealed a dramatic drop in microgametocyte exflagellation with an average of a 65% decrease in exflagellation centers observed in the  $\Delta$ lncRNA-ch14 clones compared to the parental lines, indicating a defect in male gametogenesis and microgamete formation in  $\Delta$ lncRNA-ch14 parasites (Fig. 7c). All together these data demonstrate a role of lncRNA-ch14 in gametogenesis.

We next investigated the transmissibility of  $\Delta$ lncRNA-ch14 gametocytes to mosquitoes. Infectious blood meals were prepared with parental and  $\Delta$ lncRNA-ch14 stage V gametocytes. Mosquitoes were fed with 0.2% mature stage V gametocyte infected blood using membrane feeders as described earlier<sup>86</sup>. Mosquito midguts were dissected and analyzed on day 11 post blood feeding. As expected, our mutated parental line was able to produce a high number of oocysts and sporozoites. However, we detected a significant decrease in the number of oocysts per midgut in the  $\Delta$ lncRNA-ch14 clones compared to the parental control. Both prevalence and intensity of infection were impacted in the  $\Delta$ lncRNA-ch14 clones. While 90% of control infected mosquitoes had oocysts, only 17% and 37% of the  $\Delta$ lncRNA-ch14 clones, B1 and F2, respectively were positive for oocysts. Additionally, the mosquitoes infected with the  $\Delta$ lncRNA-ch14 clones had only 1 oocyst while the control had a median of 7 oocysts ( $n=2$ ,  $p<0.05$ ) (Fig. 7d). As expected, these lower oocyst numbers resulted in significantly lower salivary gland sporozoite numbers. On day 17 salivary glands were dissected from control and  $\Delta$ lncRNA-ch14 infected mosquitoes and the average of number of salivary glands sporozoites from 2 biological replicates was 19,000 for the control line and 667 and 1993 for the  $\Delta$ lncRNA-ch14 clones B2 and F1, respectively, a decrease of about 96% ( $n=2$ ,  $p<0.05$ ) (Fig. 7e). Altogether, though we could only partially disrupt our candidate lncRNA, we clearly demonstrate that lncRNA-ch14 has an important role in gametocyte development and in the infectivity of these gametocytes for mosquitoes.

## Transcriptome perturbation in asexual and sexual stages

Based on our phenotypic assays, we predicted that perturbation of lncRNA-ch14 would affect parasite gene expression. We therefore performed RNA-seq analysis for two biological replicates each of our WT NF54 and  $\Delta$ lncRNA-ch14 clones. For each parasite line, RNA was extracted at both schizont stage, before sexual commitment and at the late gametocyte stages after sexual commitment. For each respective stage, we observed up-regulation of 78 and 57 genes in the  $\Delta$ lncRNA-ch14 lines compared to controls as well as downregulation of 383 and 18 genes (Fig. 7f and Supplementary data 3). Gene Ontology (GO) enrichment analysis revealed that up-regulated genes were involved in translation and cytoadherence ( $p=10e^{-14}$ ) that are known to be expressed in a stage specific manner between the human and vector hosts (Fig. S8b). Downregulated genes were mostly involved in microtubule movement, cell signaling and oxidation-reduction processes that are known to be critical during sexual differentiation (Fig. 7g). Five specific genes known to be upregulated in female gametocytes were detected to be significantly downregulated in our  $\Delta$ lncRNA-ch14 clones compared to our control line. These genes



**Fig. 7 | LncRNA-ch14 disruption design and characterization.** **a** Percentage of mature gametocytes. Gametocyte cultures were sampled by Giemsa-stained blood smears to assess the percent of Stage V gametocytes. Assays were performed at least in duplicate and significance of the results was calculated using the one-way ANOVA with Dunnett's multiple comparison test (\*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

**b** Percentage of gametocytes identified as male and female. Two independent biological experiments were performed and data are presented as mean values  $\pm$ SD ( $n > 500$  mature gametocytes). Significance of the results was calculated using the two-way ANOVA with Holm-Šidák correction (\*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

**c** Exflagellation assays were performed and the number of exflagellation centers per field ( $n > 10$ ) were counted. Shown are results from two biological replicates. Bars indicate the mean. **(d&e)** Mosquito passage: Gametocyte cultures were fed to *Anopheles stephensi* mosquitoes. On day 11 post-infection, midguts were removed, and oocysts were counted **(d)** and on day 17 post-infection salivary glands were

harvested and sporozoites were counted **(e)**. Data are pooled from 2 biological replicates. Oocyst counts were performed with 15–25 mosquito midguts per experiment. Significance of the results was calculated using the Kruskal-Wallis test with Dunn's post-test (\*\*\*\* $p < 0.0001$ ). For salivary gland sporozoite counts, salivary glands from 20 mosquitoes were harvested, homogenized and sporozoites counted using a hemocytometer. Shown is the average number of salivary gland sporozoites for each of two experiments. Significance was calculated using the Chi-Squared tests (\*\*\*\* $p < 0.0001$ ). **f** Volcano plots for gene expression profile between the WT and  $\Delta$ lncRNA-ch14 lines by  $-\text{Log}_{10} P$  (y-axis) and  $\text{log}_2$  Fold change (x-axis) in asexual stage parasites (Top) and in mature gametocytes (Bottom). NS: Non-significant; FC: Fold Change. **g** Bar graph representations of selected Gene Ontology (GO) enrichment of downregulated genes between WT and  $\Delta$ lncRNA-ch14 lines are presented by  $\text{Log}_{10} P$  (y-axis) in asexual mature (Left) and mature gametocyte (Right) stages. Exact  $p$  values and raw data are indicated in the Source data file.

include PF3D7\_1250100, the osmiophilic body protein G377; PF3D7\_1407000, LCCL domain-containing protein; PF3D7\_0719200, the NIMA related kinase 4; PF3D7\_0525900, the NIMA related kinase 2, and PF3D7\_1031000, the ookinete surface protein P25. Four male-specific genes were also detected as downregulated in the lncRNA-ch14 clones compared to the control. Those genes included PF3D7\_1113900, the mitogen-activated protein kinase 2, PF3D7\_1014200, the male gamete fusion factor HAP2; PF3D7\_1216700, the perforin-like protein 2 and PF3D7\_1465800, the putative dynein beta chain coding gene. All together this data confirms that lncRNA-ch14 controls, at least partially, the regulation of the transcripts known to be critical in gametocyte development including several key kinases involved in cell signaling relevant to gametogenesis.

## Discussion

Many lncRNAs are now recognized as essential regulators of chromatin structure and gene expression in eukaryotes. While some of the

identified lncRNAs have been shown to work in *cis* on neighboring genes, others seem to work in *trans* to regulate distantly located genes. Specifically, functions of nuclear lncRNAs have been determined as either directly promoting or repressing gene expression activity<sup>87,88</sup>, guiding or enhancing the functions of regulatory proteins<sup>37,88–91</sup>, or assisting the alteration of chromatin structures by shaping 3D genome organization<sup>38,92,93</sup>.

The extent of lncRNA regulation in the human malaria parasite is only now starting to emerge. A few lncRNAs have already been suggested to regulate *var* gene expression<sup>57,63,94</sup> or drive sexual commitment<sup>65,66</sup> confirming that at least some of the identified *P. falciparum* lncRNA candidates may have a functional role in the parasite life cycle progression.

In *P. falciparum*, emerging evidence has shown that chromatin structure and genome organization are of vital importance for the parasite's gene expression and regulation system<sup>28,95</sup>. Depending on their localization and their specific interactions with DNA, RNA and

proteins, lncRNAs can modulate chromatin and epigenetics in the nucleus or mRNA stability and translation in the cytosol, ultimately affecting gene expression. Therefore, identification and characterization of nuclear or cytoplasmic enriched lncRNAs may support the discovery of lncRNAs that are either chromatin-associated or translational-associated regulators of gene expression in the parasite. The dataset generated in this study presents the first global detection of lncRNAs from different subcellular locations throughout several *P. falciparum* life cycle stages. By utilizing published total and nascent RNA expression profiles (GRO-seq<sup>8</sup>), we were able to significantly improve the sensitivity of lncRNA detection, especially for the identification of nuclear lncRNAs. Using both experimental and computational pipelines, we identified 1768 lncRNAs covering 204 cytoplasmic enriched, 719 nuclear enriched, and 845 lncRNAs that localized to both fractions. Our data suggest that nuclear and cytoplasmic lncRNAs are coordinately expressed but cytoplasmic lncRNAs are less abundant as compared to the number of nuclear lncRNAs in the parasite. In addition, we observed that a small group of cytoplasmic lncRNAs is highly expressed at the trophozoite stage, the stage where a large proportion of genes are transcribed<sup>8</sup>. Though more in-depth studies will be required to confirm the functions of these trophozoite-expressed cytoplasmic lncRNAs, it is possible that some of these lncRNAs are involved in mRNA stability or translational regulation.

In our present work, we also observed that many lncRNAs enriched in the nuclear fraction, including the lncRNA-TAREs, are highly abundant at the ring and schizont stages. This finding suggests that some of these lncRNAs (cluster 1, Fig. 3a) are likely to be involved in heterochromatin maintenance or chromatin structure reorganization events, as previous ChIP-seq and Hi-C experiments have shown that epigenetics and chromatin are critical to gene expression at the initiation level<sup>27</sup>. Additionally, ChIRP experiments mapping genome-wide binding sites of lncRNAs revealed that lncRNA-TARE4 binds to subtelomeric regions on multiple chromosomes as well as regulatory regions around genes involved in pathogenesis and immune evasion. Previous reports showed that subtelomeric regions as well as virulence gene families cluster in perinuclear heterochromatin. Therefore, evidence suggests a role for lncRNA-TARE4 in transcriptional and/or epigenetic regulation of parasite telomeric and subtelomeric regions by interacting with or recruiting histone-modifying complexes to targeted regions to maintain them in a heterochromatin state, much like the case of X chromosome inactivation regulated via lncRNA Xist<sup>96</sup>.

Genomic occupancy of other lncRNAs explored here, suggest that these lncRNAs bind around the gene regions. In all cases investigated, a positive correlation was observed between the lncRNA expression and the expression of genes around the lncRNA occupancy sites (Fig. 6). Given already existing evidence for lncRNA-associated epigenetic modification and transcriptional regulation in other eukaryotes<sup>97–99</sup>, it is likely that the lncRNAs identified in the parasite nucleus are responsible for coordinated recruitment of distinct repressing proteins and/or histone-modifying complexes to target loci. Additionally, we uncovered that the lncRNA binding sites were situated upstream of the start codon of target genes (Fig. S6). This pattern of lncRNA occupancy provides additional support for the idea that the lncRNAs explored here might have a role in recruiting protein complexes to promoter regions of target genes to regulate transcription, either by activating the formation of the pre-initiation complex or recruiting histone modifiers. However, while additional experiments are needed to confirm the roles of these nuclear lncRNAs in the parasite, using ChIRP-seq, we demonstrate that genome-wide collections of RNA binding sites can be used to discover the DNA sequence motifs enriched by lncRNAs. These findings signify the existence of lncRNA target sites in the genome, an entirely new class of regulatory elements that could be essential for transcriptional regulation in the malaria parasite.

Genetic disruption of lncRNA-ch14, a transcript detected specifically in gametocytes, demonstrates that this lncRNA plays an important role in sexual differentiation and is required for onward transmission to the mosquito (Fig. 7a and b). This finding is supported by our transcriptomic analysis where we identified significant down-regulation of genes involved in sexual differentiation including NEK and MAP kinases<sup>100,101</sup> ookinete/oocyst development<sup>102</sup> and microtubule function (i.e., dyneins and kinesins) most likely important in reshaping the parasite into sexual stages (Fig. 7f and 7g, Fig. S8b and Supplementary data 3). Importantly, the skewed sex ratio of the  $\Delta$ lncRNA-ch14 parasites does not completely account for the dramatic decrease in the ability of these parasites to be transmitted to mosquitoes. Indeed, our data suggest that the gametocytes of the  $\Delta$ lncRNA-ch14 parasites are less infectious, a result that is also supported by the decrease in exflagellation of the male gametocytes (Fig. 7c). It is currently difficult to assess infectiousness of female gametocytes, but we would hypothesize that these are also impacted by the disruption of lncRNA-ch14. While further experiments will be needed to further validate the effect of the full deletion or down-regulation of the lncRNA-ch14 transcript in the mosquito stage, the results presented here confirm that some of the lncRNAs identified in this study play a role in the parasite's sexual development and onward transmission to the mosquito.

Compared to the progress made in understanding lncRNA biology in higher eukaryotes, the field of lncRNAs in *Plasmodium* is still evolving. Analysis of promoter and gene body regions with available histone modification datasets (H3K9me3, H3K36me3, and H3K9ac) are still needed for further annotation of these candidate lncRNAs. It is clear that lncRNAs represent a new paradigm in chromatin remodeling and genome regulation. Therefore, this newly generated dataset will not only assist future lncRNA studies in the malaria parasite but will also help in identifying parasite-specific gene expression regulators that can ultimately be used as new anti-malarial drug targets.

## Methods

### Parasite culture

*P. falciparum* 3D7 strain at ~8% parasitemia was cultured in human erythrocytes at 5% hematocrit in 25 mL of culture as previously described in<sup>103</sup>. Two synchronization steps were performed with 5% D-sorbitol treatments at ring stage within eight hours. Parasites were collected at early ring, early trophozoite, and late schizont stages. Parasite developmental stages were assessed using Giemsa-stained blood smears.

### Nuclear and cytosolic RNA isolation

Highly synchronized parasites were first extracted using 0.15% saponin solution followed by centrifugation at 1500  $\times$ g for 10 min at 4 °C. Parasite pellets were then washed twice with ice cold PBS and re-collected at 1500  $\times$ g. Parasite pellets were resuspended in 500  $\mu$ L ice cold Cell Fractionation Buffer (PARIS kit, ThermoFisher; AMI921) with 10  $\mu$ L of RNase Inhibitor (SUPERaseIn 20U/ $\mu$ L, Invitrogen; AM2694) and incubated on ice for 10 minutes. Samples were centrifuged at 500  $\times$ g for 5 min at 4 °C. After centrifugation, the supernatant containing the cytoplasmic fraction was collected. Nuclei were resuspended in 500  $\mu$ L Cell fractionation buffer and 15  $\mu$ L RNase Inhibitor as described above. To obtain a more purified nuclear fraction, the pellet was syringed with a 26 G inch needle five times. The sample was incubated on ice for 10 min and centrifuged at 500  $\times$ g for 5 min at 4 °C. The nuclear pellet was resuspended in 500  $\mu$ L of ice-cold Cell Disruption Buffer (PARIS kit, ThermoFisher; AMI921). For both cytoplasmic and nuclear fractions, RNA was isolated by adding 5 volumes of Trizol LS Reagent (Life Technologies, Carlsbad, CA, USA) followed by a 5 min incubation at 37 °C. RNA was then isolated according to manufacturer's instructions. DNA-free DNA removal kit (ThermoFisher; AMI906) was used to remove potential genomic DNA contamination

according to manufacturer's instruction, and the absence of genomic DNA was confirmed by performing a 40-cycle PCR on the PfAlba3 gene using 200 to 500 ng input RNA.

### mRNA isolation and library preparation

Messenger RNA was purified from total cytoplasmic and nuclear RNA samples using NEBNext Poly(A) nRNA Magnetic Isolation module (NEB; E7490S) with manufacturer's instructions. Once mRNA was isolated, strand-specific RNA-seq libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB; E7420S) with library amplification specifically modified to accommodate the high AT content of *P. falciparum* genome: libraries were amplified for a total of 12 PCR cycles (45 s at 98 °C followed by 15 cycles of 15 s at 98 °C, 30 s at 55 °C, 30 s at [62 °C], 5 min 62 °C). Libraries were then sequenced on Illumina NExtSeq500 generating 75 bp paired-end sequence reads.

### Sequence mapping

After sequencing, the quality of raw reads was analyzed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The first 15 bases and the last base were trimmed. Contaminating adaptor reads, reads that were unpaired, bases below 28 that contained Ns, and reads shorter than 18 bases were also filtered using Sickle (<https://github.com/najoshi/sickle>)<sup>104</sup>. All trimmed reads were then mapped to the *P. falciparum* genome (v34) using HISAT2<sup>105</sup> with the following parameters: `-t`, `-downstream-transcriptome assembly`, `-max-intron-len 3000`, `-no-discordant`, `-summary-file`, `-known-splicesite-infile`, `-rna-strandness RF`, and `-novel-splicesite-outfile`. After mapping, we removed all reads that were not uniquely mapped, not properly paired (samtools v 0.1.19-44428 cd<sup>106</sup>) and are likely to be PCR duplicates (Picard tools v1.78, [broadinstitute.github.io/picard/](http://broadinstitute.github.io/picard/)). The final number of working reads for each library is listed in Supplementary data 1. For genome browser tracks, read coverage per nucleotide was first determined using BEDTools<sup>107</sup> and normalized per million mapped reads.

### Transcriptome assembly and lncRNA identification

To identify lncRNAs in the nuclear and cytoplasmic fractions, we first merged all nuclear libraries and cytoplasmic libraries for each replicate, resulting in one pair of nuclear and cytoplasmic dataset per replicates. Next, we assembled the transcriptome (cufflinks v2.1.1)<sup>108</sup> for each of the datasets using the following parameters: `-p 8 -b PlasmoDB-34_Pfalciparum3D7_Genome.fasta -M PlasmoDB-34_Pfalciparum3D7.gff -library-type first strand -l 5000`. After transcriptome assembly, we filtered out transcripts that are less than 200 basepairs and were predicted to be protein-coding (CPAT, <http://lilab.research.bcm.edu>). We then merge transcripts in both replicates using cuffmerge and removed any transcripts that are on the same strand and have more than 30% overlap with annotated regions (BEDTools intersect). Lastly, we further selected transcripts that have both primary and steady-state transcriptional evidence. For primary transcription, we used GRO-seq dataset (GSE85478) and removed any transcript that has a read coverage below 15% of the median expression of protein-encoding genes, as well as transcript that has an FPKM count less than 10 at any given stage. The same filtering criteria were also applied to the steady-state RNA-seq expression profiles.

To estimate the cellular location for each predicted lncRNA, we first calculated the summed read count of all nuclear libraries and the summed read count of all cytoplasmic libraries. Then, we measured the log<sub>2</sub>-fold change (log<sub>2</sub> FC) of the summed nuclear signal to the summed cytoplasmic signal. Any transcripts with a log<sub>2</sub> FC value above 0.5 were classified as nuclear enriched lncRNAs, and any transcripts with a log<sub>2</sub> FC value below -0.5 were classified as cytoplasmic enriched lncRNAs. In addition, lncRNAs with a log<sub>2</sub> FC value between the above thresholds were classified as lncRNAs expressed equally in both fractions.

### Western blot

Mixed-stage parasites were collected as described above. Parasite pellets were gently resuspended in 500 µL of ice-cold Cell Fractionation Buffer (PARIS kit, ThermoFisher; AM1921) and 50 µL of 10X EDTA-free Protease inhibitor (cOmplete Tablets, Mini EDTA-free, EASY pack, Roche; 05 892 791 001). Solution was incubated on ice for 10 min and the sample was centrifuged for 5 min at 4 °C and 500 × g. The supernatant containing cytoplasmic fraction was collected carefully and the nuclear pellet was resuspended in 500 µL Cell Fractionation Buffer followed by needle-lysis 5x using 26 G inch needle. Nuclei were collected again at 4 °C and 500 × g. The supernatant was discarded and the nuclei pellet in 500 µL of Cell Disruption Buffer (PARIS kit, ThermoFisher; AM1921) and incubated on ice for 10 min. The nuclear fraction was then sonicated 7x with 10 s on/30 s off using a probe sonicator. Extracted nuclear protein lysates were incubated for 10 min at room temperature and centrifuged for 2 min at 10,000 × g to remove cell debris. Seven micrograms of parasite cytoplasmic and nuclear protein lysates were diluted in a 2X Laemmli buffer at a 1:1 ratio followed by heating at 95 °C for 10 min. Protein lysates are then loaded on an Any-KD SDS-PAGE gel (Bio-Rad, 569033) and run for 1 h at 125 V. Proteins were transferred to a PVDF membrane for 1 h at 18 V, then stained using commercial antibodies generated against histone H3 (1:3000 dilution, Abcam; ab1791) and PfAldolase (1:1000 dilution, Abcam; ab207494), and secondary antibody, Goat Anti-Rabbit IgG HRP Conjugate (1:25,000 dilution, Bio-Rad; 1706515). Membranes were visualized using the Bio-Rad ChemiDoc MP Gel Imager and images were treated using Image Lab software. Uncropped blots are presented in Source data file.

### PiggyBac insertion analysis

To analyze lncRNA essentiality, we used piggyBac insertion sites from<sup>68</sup>, who performed saturation mutagenesis to uncover essential genes in *P. falciparum*. Since that study used an NF54 reference genome, we converted the coordinates to be applicable to the 3D7 reference genome (v38, PlasmoDB), using liftOver (Kent tools v427, UCSC Genome Bioinformatics group, <https://github.com/ucscGenomeBrowser/kent>). A chain file for the two genomes, needed for liftOver, was manually constructed as described here: [http://genomewiki.ucsc.edu/index.php/LiftOver\\_Howto](http://genomewiki.ucsc.edu/index.php/LiftOver_Howto).

Custom Python scripts were used to overlap insertion site coordinates with lncRNA ranges to count the number of insertions that occurred in each lncRNA, as well as to locate TTAA sites (sites where piggyBac insertions could potentially occur) in the genome and count the number of TTAA sites in each lncRNA. These scripts were also used to determine the normalized location of each TTAA site and insertion site, in one of 50 windows either across the lncRNA range or also including the 5' and 3' flanking regions, which were each given 50% of the length of the lncRNA. The ratio of number of insertion sites to number of TTAA sites within a lncRNA was used as a loose measure of essentiality.

### Estimation of transcript stability

Read coverage values were calculated from total steady-state RNA datasets (SRP026367, SRS417027, SRS417268, SRS417269) using BEDTools v2.25.0. The read counts were then normalized as described in the original publication, and ratios between RNA-seq and GRO-seq coverage values were calculated for each lncRNA and gene. This ratio reflects the relative abundance of the mature RNA transcript over its corresponding primary transcript and is a simple but convenient measurement for transcript stability.

### Reverse transcription PCR

Total RNA was isolated from 10 mL of mixed-stage asexual *P. falciparum* culture and 25 mL of late gametocyte stage culture. Total RNA quality was checked on an agarose gel and genomic DNA

contamination was removed using a DNA-free DNA removal kit (ThermoFisher; AM1906) according to manufacturer's instructions. The absence of genomic DNA was validated using a primer set targeting an intergenic region within *PfAlba3* (PF3D7\_1006200). Approximately 1 µg of DNase I treated RNA from each sample was used in a 35-cycle PCR reaction to confirm the absence of genomic DNA contamination. DNase-treated total RNA was then mixed with 0.1 µg of random hexamers, 0.6 µg of oligo-dT (20), and 2 µL 10 mM dNTP mix (Life Technologies) in total volume of 10 µL, incubated for 10 minutes at 70 °C and then chilled on ice for 5 minutes. This mixture was added to a solution containing 4 µL 10X RT buffer, 8 µL 20 mM MgCl<sub>2</sub>, 4 µL 0.1 M DTT, 2 µL 20U/µl Superscript III and 1 µL 200 U/µL SuperScript III Reverse Transcriptase (Invitrogen, 18080044). First-strand cDNA was synthesized by incubating the sample for 10 min at 25 °C, 50 min at 50 °C, and finally 5 minutes at 85 °C. First strand cDNA is then mixed with 70 µL of nuclease free water, 30 µL 5x second-strand buffer (Invitrogen, 10812014), 3 µL 10 mM dNTP mix (Life Technologies), 4 µL 10 U/µL *E. coli* DNA Polymerase (NEB, M0209), 1 µL 10 U/µL *E. coli* DNA ligase (NEB, M0205) and 1 µL 2 U/µL *E. coli* RNase H (Invitrogen, 18021014). Samples were incubated for 2 h at 16 °C and double stranded cDNA was purified using AMPure XP beads (Beckman Coulter, A63881). For testing transcription activity of predicted genes, 450 ng of double stranded cDNA was mixed with 10 pmole of both forward and reverse primers. DNA was incubated for 5 minutes at 95 °C, then 30 s at 98 °C, 30 s at 55 °C, 30 s at 62 °C for 25 cycles. All primers used for PCR validation are listed in Supplementary data 1.

### Single-cell sequencing and data processing

*P. falciparum* strain NF54 was cultured in O+ blood in complete RPMI 1640 culture medium at 37 °C in a gas mixture of 5% O<sub>2</sub>/5% CO<sub>2</sub>/90% N<sub>2</sub>, as described previously<sup>9,109</sup>. Sexual commitment was induced at 1% parasitemia and 3% hematocrit and culture media were supplemented with 10% human serum. After 4, 6 and 10 days post sexual commitment, samples were taken from the culture for single cell sequencing. Cells from each day were loaded into separate inlets in a 10X chromium controller using the manufacturer's instructions for a 10,000-target cell capture. Libraries for the days 4 and 6 samples were obtained using Chromium 10X version 2 chemistry, whereas libraries for the day 10 sample were obtained using version 3 chemistry. Cells were sequenced on a single lane of a HighSeq4000 using 150-bp paired-end reads. Raw reads were mapped to a custom gtf containing lncRNA coordinates appended to the *P. falciparum* 3D7 V3 reference genome ([www.sanger.ac.uk/resources/downloads/protozoa/](http://www.sanger.ac.uk/resources/downloads/protozoa/)). Read mapping, deconvolution of cell barcodes and UMIs and the generation of single cell expression matrices were performed using the CellRanger pipeline v 3.0.0. lncRNA regions were labeled as 'protein coding' to be prioritized in STAR mapping in CellRanger. CellRanger was also run separately for each sample using the 3D7 reference genome that did not contain the appended non-coding regions for comparison. Resultant count matrices were loaded into the R package Seurat (v3.2.2) for pre-processing.

### Quality control and lncRNA expression

Single-cell transcriptomes (SCTs) were log-normalized, and expression scaled using Seurat (v3.2.2). Each cell was assigned a stage by mapping to the Malaria Cell Atlas<sup>109</sup> using scmap-cell (v1.8.0). Cells were assigned the stage of their closest neighbor in the Malaria Cell Atlas if they reached a cosine similarity of > 0.2. Cells identified as an early/late ring or late schizont containing <50 UMI/cell and <50 genes/cell were removed due to poor quality. Cells mapped to late stages, or cells not assigned to a stage in the Malaria Cell Atlas were removed if they contained <100 UMIs/cell or <80 genes/cell. Data from days 4, 6 and 10 were integrated together using Seurat's IntegrateData function using 2000 integration anchors and 10 significant principal components. A variance stabilizing transformation was

performed on the integrated matrix to identify the 750 most highly variable coding genes, and these were used to perform a principal component (PC) analysis. Significant PCs were then used to calculate three-dimensional UMAP embeddings using only coding genes. lncRNA expression was visualized on the UMAP embedding generated from coding gene expression using the package ggplot2 to assign stage-specific expression for the lncRNA.

### RNA in situ hybridization (RNA-FISH)

RNA FISH was performed with slight modifications as described by Sierra-Miranda, 2012<sup>59</sup> on mixed-stage asexual and gametocyte stage parasites. Antisense RNA probes for seven nuclear lncRNAs; -TARE4, -178, -13, -1494, -271, -4076 -ch9, -ch14 and two cytoplasmic lncRNAs; -267, -643, were labeled by in vitro transcription in the presence of fluorescein. RNA FISH was also performed using sense RNA probes as controls. Briefly, fixed and permeabilized parasites were incubated with RNA probes overnight at 37 °C. Parasites were washed with 2x SSC three times for 15 min each at 45 °C followed by one wash with 1x PBS for 5 min at room temperature. The slides were mounted in a Vectashield mounting medium with DAPI and visualized using the Olympus BX40 epifluorescence microscope. Images were treated with ImageJ. Pictures are representative of 15-20 positive parasites examined.

### Immunofluorescence assays

Parasites were fixed with 4% paraformaldehyde and 0.0075% glutaraldehyde for 15 min at 4 °C, and then sedimented on Poly-D-lysine coated coverslips for 1 hr. at room temperature. After PBS washes, parasites were permeabilized and saturated with 0.2% Triton X-100, 5% BSA, 0.1% Tween 20 in PBS for 30 min at room temperature. Anti-H3K9me3 mAb (Abcam, ab184677) was diluted at 1:500 in 5% BSA, 0.1% Tween 20 and PBS, and applied for 1 hr. at room temperature. After PBS washes, Goat anti-Mouse Alexa Fluor 488 (Invitrogen, A11001) was diluted at 1:2000 and applied for 1 h at room temperature. Coverslips were mounted in Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200). Images were acquired using a Keyence BZ-X810 fluorescence microscope and treated with ImageJ.

### Chromatin isolation by RNA purification (ChIRP)

ChIRP-seq experiments were performed in duplicate for all nuclear lncRNAs, at the time point of highest lncRNA expression and lowest expression. Synchronized parasite cultures were collected and incubated in 0.15% saponin for 10 min on ice to lyse red blood cells. Parasites were centrifuged at 3234 x g for 10 min at 4 °C and subsequently washed three times with PBS by resuspending in cold PBS and centrifuging for 10 min at 3234 x g at 4 °C. Parasites were cross-linked for 15 min at RT with 1% glutaraldehyde. Cross-linking was quenched by adding glycine to a final concentration of 0.125 M and incubating for 5 min at 37 °C. Parasites were centrifuged at 2500 x g for 5 min at 4 °C, washed three times with cold PBS and stored at -80 °C.

To extract nuclei, parasite was first incubated in nuclear extraction buffer (10 mM HEPES, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), EDTA-free protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche)) on ice. After 30 min, Igepal CA-630 (Sigma-Aldrich, I8896) was added to a final concentration of 0.25% and needle sheared seven times by passing through a 26 G ½ needle. Parasite nuclei were centrifuged at 2500 x g for 20 min at 4 °C and resuspended in shearing buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl pH 7.5, EDTA-free protease inhibitor cocktail and phosphatase inhibitor cocktail). Chromatin was fragmented using the Covaris UltraSonicator (S220) to obtain 100-500 bp DNA fragments with the following settings: 5% duty cycle, 140 intensity incident power, 200 cycles per burst. Sonicated samples were centrifuged for 10 min at 17000 x g at 4 °C to remove insoluble material.

Fragmented chromatin was precleared using Dynabeads MyOne Streptavidin T1 (Thermo Fisher, 65601) by incubating for 30 min at 37 °C to reduce non-specific background. Per ChIRP sample using 1 mL of lysate, 10 µL each was removed for the RNA input and DNA input, respectively. Each sample was diluted in 2x volume of hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris-Cl pH 7.5, 1 mM EDTA, 15% formamide, 0.0005x volume of AEBSEF, 0.01x volume of SUPERase-In (Ambion, AM2694) and 0.01x volume of protease inhibitor cocktail). ChIRP probes used for each lncRNA (see Supplementary data 1) were pooled, heated at 85 °C for 3 min and cooled on ice. ChIRP probes were added to each sample (2 µL of 100 µM pooled probes per sample) and incubated at 37 °C with end-to-end rotation for 4 h. Prior to completion of hybridization, Dynabeads MyOne Streptavidin T1 beads were washed three times on a magnet stand using lysis buffer (50 mM Tris-Cl pH 7, 10 mM EDTA, 1% SDS). After the hybridization, 100 µL of washed T1 beads were added to each tube and incubated for 30 min at 37 °C. Beads were washed with wash buffer (2x SSC, 0.5% SDS, 0.005x volume of AEBSEF) and split evenly for isolation of DNA and RNA fractions.

For RNA isolation, the RNA input and chromatin-bound beads were resuspended in RNA elution buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.5% SDS, 1 mg/mL Proteinase K), incubated at 50 °C for 45 min, boiled at 95 °C for 15 min and subjected to trizol:chloroform extraction. Genomic DNA contamination was removed using a DNA-free DNA removal kit (ThermoFisher, AM1906) according to manufacturer's instructions. The absence of genomic DNA was validated using a primer set targeting an intergenic region within (PF3D7\_1006200) in a 35-cycle PCR reaction. DNase-treated RNA was then mixed with 0.1 µg of random hexamers, 0.6 µg of oligo-dT (20), and 2 µL 10 mM dNTP mix (Life Technologies) in total volume of 10 µL, incubated for 10 min at 70 °C and then chilled on ice for 5 min. This mixture was added to a solution containing 4 µL 10X RT buffer, 8 µL 20 mM MgCl<sub>2</sub>, 4 µL 0.1 M DTT, 2 µL 20 U/µL SUPERase-In and 1 µL 200 U/µL SuperScript III Reverse Transcriptase. First-strand cDNA was synthesized by incubating the sample for 10 min at 25 °C, 50 min at 50 °C, and finally 5 min at 85 °C followed by a 20 min incubation with 1 µL 2 U/µL *E. coli* RNase H at 37 °C. Prepared cDNA was then subjected to quantitative reverse-transcription PCR for the detection of enriched TARE-4 and serine tRNA ligase transcripts with the following program: 5 min at 95 °C, 30 cycles of 30 s at 98 °C, 30 s at 55 °C, 30 s at 62 °C and a final extension 5 min at 62 °C. All primers used for PCR validation are listed in Supplementary data 1.

Libraries from the ChIRP samples were prepared using the KAPA Library Preparation Kit (KAPA Biosystems). Libraries were amplified for a total of 12 PCR cycles (12 cycles of 15 s at 98 °C, 30 s at 55 °C, 30 s at (62 °C]) using the KAPA HiFi HotStart Ready Mix (KAPA Biosystems). Libraries were sequenced with a NextSeq500 DNA sequencer (Illumina). Raw read quality was first analyzed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped to the *P. falciparum* genome (v38, PlasmoDB) using Bowtie2 (v2.4.2). Duplicate, unmapped, and low quality (MAPQ < 20) reads were filtered out using Samtools (v1.9), and only uniquely mapped reads were retained. All libraries, including the input, were then normalized by dividing by the number of mapped reads in each of them. For each nucleotide, the signal from the input library was then subtracted from each of the ChIRP-seq libraries, and any negative value was replaced with a zero. Genome tracks were generated by the R package ggplot2.

### Peak calling

Peaks were called using PePr v1.1. For a given lncRNA of interest, the tool was run in differential binding analysis mode using the filtered ChIRP-seq libraries for when the lncRNA was active versus non-active with the following parameters specified: -peaktype broad -threshold 1e<sup>-10</sup>. The top 25% of all reported peaks were selected because they exhibited the strongest signal (see Fig. S4) and used in downstream analyses. For differential gene expression analysis, the closest gene to

each peak was selected, and its expression in the two stages (active versus inactive) was obtained from<sup>29,67,110</sup>.

### lncRNA-ch14 gene disruption

Gene knockout (KO) for the long non-coding RNA 14 (lncRNA-ch14) spanning position (Ch14:3,148,960–3,150,115) on chromosome 14 was performed using a two-plasmid design. The plasmid pDC2-Cas9-sgRNA-hdhfr<sup>111</sup>, gifted from Marcus Lee (Wellcome Sanger Institute) contains the SpCas9, a site to express the sgRNA, and a positive selectable marker human dihydrofolate reductase (*dhfr*). The sgRNA was selected from the database generated by<sup>112</sup> and cloned into pDC2-Cas9-sgRNA-hdhfr at the BbsI restriction site. The homology directed repair plasmid (modified pDC2-donor-*bsd* without eGFP) was designed to insert a selectable marker, blasticidin S-deaminase (*bsd*), disrupting the lncRNA-ch14 region. The target specifying homology arm sequences were isolated through PCR amplification and gel purification. The right homology regions (RHR), and the left homology regions (LHR) of each gene were then ligated into the linearized vector via Gibson assembly. The final donor vectors were confirmed by restriction digests and Sanger sequencing. All primers are indicated in Supplementary data 1n.

Plasmids were isolated from 250 mL cultures of *Escherichia coli* (XL10-Gold Ultracompetent Cells, Agilent Cat. 200314) and 60 µg of each plasmid was used to transfect ring stage parasites. 24-hrs before transfection, mature parasite cultures (6–8% parasitemia) were magnetically separated using magnetic columns (MACS LD columns, Miltenyi Biotec) and diluted to 1% parasitemia containing 0.5 mL fresh erythrocytes<sup>113</sup>. The next day, ~3% ring stage parasites were pelleted and washed in 4 mL of cytomix<sup>114</sup>. 200 µL of the infected erythrocytes were resuspended with the two plasmids in cytomix to a total volume of 400 µL in a 0.2 cm cuvette. Electroporation was performed with a single pulse at 0.310 kV and 950 µF using the Biorad GenePulser electroporator. Cells were immediately transferred to a flask containing 12 mL media and 400 µL erythrocytes. The media was exchanged 5 hrs post electroporation with 12 mL of fresh media. The following day, fresh culture media was added and supplemented with 2.5 nM WR99210 and 2.5 µg/mL blasticidin (RPI Corp, B12150-0.1). Media and drug selection were replenished every 48 h. After 14 days, the culture was split into two flasks and 50 µL of erythrocytes were added every two weeks. Once parasites were detected by microscopy, WR99210 was removed (selection for Cas9). Integration of the *bsd* gene was confirmed by gDNA extraction and PCR.

### Isolation of ΔlncRNA-ch14 clone

To generate genetically homogenous parasite lines, the transfected parasites were serially diluted to approximately 0.5%, into 96 well plates. 200 µL final volume of cultured parasites were incubated with *bsd* drug selection for 1 month with weekly erythrocytic and media changes for the first 2 weeks of dilution followed by media changes every 2 days until parasite recovery is observed through Giemsa-stained smears.

### Verification of ΔlncRNA-ch14 line

Genomic DNA (gDNA) was extracted and purified using DNeasy Blood & Tissue kit (Qiagen, 69504) following instructions from the manufacturer. The genotyping PCR analysis was used to genotype the KO lines using primer indicated in Supplementary data 1n. The PCR amplification was done using 2xKAPA master mix for thirty cycles with an annealing temperature of 50 °C and an extension temperature of 62 °C. The PCR amplicons were analyzed on a 1% agarose gel electrophoresis.

For whole genome sequencing, genomic DNAs were fragmented using a Covaris S220 ultrasonicator and libraries were generated using KAPA LTP Library Preparation Kit (Roche, KK8230). To verify that the insertion was present in the genome at the correct location in both

transfected lines, reads were mapped using Bowtie2 (version 2.4.4) to the *P. falciparum* 3D7 reference genome (v48, PlasmoDB), edited to include the insertion sequence in the intended location. IGV (Broad Institute) was used to verify that reads aligned to the insertion sequence.

### **ΔlncRNA-ch14 line genome-wide sequencing and variant analysis**

Libraries were sequenced using a NovaSeq 6000 DNA sequencer (Illumina), producing paired-end 100-bp reads. To verify that the insertion was present in the genome at the correct location in both transfected lines, reads were mapped using Bowtie2 (version 2.4.4) to the *P. falciparum* 3D7 reference genome (v48, PlasmoDB), edited to include the insertion sequence in the intended location. IGV (Broad Institute) was used to verify that reads aligned to the insertion sequence. To call variants (SNPs/indels) in the transfected lines compared to a previously sequenced NF54 control line, genomic DNA reads were first trimmed of adapters and aligned to the *Homo sapiens* genome (assembly GRCh38) to remove human-mapped reads. Remaining reads were aligned to the *P. falciparum* 3D7 genome using bwa (version 0.7.17) and PCR duplicates were removed using PicardTools (Broad Institute). GATK HaplotypeCaller (<https://gatk.broadinstitute.org/hc/en-us>) was used to call variants between the sample and the 3D7 reference genome for both the transfected lines and the NF54 control. Only variants that were present in both transfected lines but not the NF54 control line were kept. We examined only coding-region variants and removed those that were synonymous variants or were in *var*, *rifin*, or *stevor* genes. Quality control of variants was done by hard filtering using GATK guidelines.

### **Assessment of gametocyte development**

Viability of gametocytes was assessed via microscopy in parasite laboratory strains NF54 and two of the ΔlncRNA-ch14 clones, F2 and B1. The morphology of parasite gametocytes was assessed in a Giemsa-stained thin blood smear. Gametocytes were classified either as viable (normal intact morphology of mature gametocytes) or dead (deformed cells with a decrease in width, a thin needle-like appearance or degraded cytoplasmic content).

### **Gametocyte cultures and mosquito feeding**

This was performed as outlined previously [50]. Briefly, asexual stage cultures were grown in RPMI-1640 containing 2 mM L-glutamine, 50 mg/L hypoxanthine, 25 mM HEPES, 0.225% NaHCO<sub>3</sub>, contained 10% *v/v* human serum in 4% human erythrocytes. Five mL of an asexual stage culture at 5% parasitemia was centrifuged at 500 × *g* for 5 min at room temperature. Gametocyte cultures were initiated at 0.5% asynchronous asexual parasitemia from low passage stock and maintained up to day 18 with daily media changes but without any addition of fresh erythrocytes. The culture medium was changed daily for 15–18 days, by carefully aspirating ~ 70–80% of the supernatant medium to avoid removing cells, and 5 mL of fresh complete culture medium was added to each well. Giemsa-stained blood smears were made every alternate day to confirm that the parasites remained viable. Instead of a gas incubator, cultures were maintained at 37 °C in a candle jar made of glass desiccators. On day 15 to 18, gametocyte culture, containing largely mature gametocytes, were used for mosquito feeds. Cultures were transferred to pre-warmed tubes and centrifuged at 500 × *g* for 5 min. The cells were diluted in a pre-warmed 50:50 mixture of uninfected erythrocytes and normal human serum to achieve a mature gametocytemia of 0.2% and the resulting ‘feeding mixture’ was placed into a pre-warmed glass feeder. Uninfected *Anopheles stephensi* mosquitoes, starved overnight of sugar water, were allowed to feed on the culture for 30 min. Unfed mosquitoes were removed, and the mosquito cups were placed in a humidified 26 °C incubator, with 10% sugar-soaked cotton pads placed on top of the mosquito cage.

### **Oocyst and salivary gland sporozoite quantification**

On days 11 and 17 after the infective-blood meal, mosquitoes were dissected and midguts or salivary glands, respectively, were harvested for sporozoite counts. Day 11 midguts were stained with mercurochrome and photographed for oocyst counts by brightfield and phase microscopy using an upright Nikon E600 microscope with a PlanApo 4× objective. On day 17, salivary glands from ~20 mosquitoes were pooled, homogenized, and released sporozoites were counted using a haemocytometer.

### **Gametocyte quantification, sex determination and exflagellation assay**

Between days 15 to 18, blood smears were prepared from gametocyte cultures, fixed with methanol and stained with Giemsa (Sigma, GS500), prepared as a 1:5 dilution in buffer (pH = 7.2) made using Gurr buffered tablets (VWR, 331942 F) and filtered. Slides were stained for 20 min, washed with buffer, and allowed to dry before observation using a Nikon E600 microscope with a PlanApo 100× oil objective. For calculation of gametocytemias and male: female ratios, at least 500 mature gametocytes were scored per slide. To count exflagellation centers, 500 μL of mature gametocyte culture was centrifuged at 500 × *g* for 4 min and the resulting pellet was resuspended in equal volume of prewarmed normal human serum. Temperature was dropped to room temperature to activate gametogenesis and after 15 min incubation 10 μL of culture was transferred to a glass slide and covered with a cover slip. Exflagellation centers were counted at 40× objective in at least ten fields for each gametocyte culture. To avoid bias, microscopic examination was performed in a blinded fashion by a trained reader.

### **ΔlncRNA-ch14 transcriptome analysis**

Libraries were prepared from the extracted total RNA, first by isolating mRNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490), then using the NEBNext Ultra Directional RNA Library Prep Kit (NEB, E7420). Libraries were amplified for a total of 12 PCR cycles (12 cycles of 15 s at 98 °C, 30 s at 55 °C, 30 s at [62 °C]) using the KAPA HiFi HotStart Ready Mix (KAPA Biosystems). Libraries were sequenced using a NovaSeq 6000 DNA sequencer (Illumina), producing paired-end 100-bp reads.

FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), was used to assess raw read quality and characteristics, and based on this information, the first 11 bp of each read and any adapter sequences were removed using Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>). Bases were trimmed from reads using Sickle with a Phred quality threshold of 20 (<https://github.com/najoshi/sickle>). These reads were mapped against the *Homo sapiens* genome (assembly GRCh38) using Bowtie2 (version 2.4.4) and mapped reads were removed. The remaining reads were mapped against the *Plasmodium falciparum* 3D7 genome (v48, PlasmoDB) using HISAT2 (version 2.2.1), using default parameters. Uniquely mapped, properly paired reads with mapping quality 40 or higher were retained using SAMtools (<http://samtools.sourceforge.net/>), and PCR duplicates were removed using PicardTools (Broad Institute). Genome browser tracks were generated and viewed using the Integrative Genomic Viewer (IGV) (Broad Institute).

Raw read counts were determined for each gene in the *P. falciparum* genome using BedTools (<https://bedtools.readthedocs.io/en/latest/#>).

to intersect the aligned reads with the genome annotation. Differential expression analysis was done by use of R package DESeq2 (<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>) to call up- and down-regulated genes with an adjusted *P*-value cutoff of 0.05. Gene ontology enrichment was done using PlasmoDB (<https://plasmodb.org/plasmo/app>). Volcano plots were generated using R package EnhancedVolcano (<https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html>).



### Statistical analysis

Descriptive statistics were calculated with GraphPad Prism version 9.1.2 (GraphPad Software, San Diego, CA, USA) for determining mean, percentages, standard deviation and plotting of graphs. Excel 2013 and GraphPad Prism 9.1.2 were used for the calculation of gametocytemia of microscopic data.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The GRO-seq data used in this study are available in the Gene Expression Omnibus database under accession code [GSE85478](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85478). The steady-state RNA-seq data used in this study are available in the NCBI Sequence Read Archive under accession code [SRP026367](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRP026367), [SRS417027](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRS417027), [SRS417268](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRS417268), and [SRS417269](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRS417269).

WGS, ChIRP-seq and RNA-seq data generated in this study (66 libraries) have been deposited in the NCBI Sequence Read Archive with accession PRJNA869073 and are available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA869073>. All other data generated in this study are provided in the Supplementary Information, Supplementary data and Source Data file. Source data are provided with this paper.

### References

1. WHO. *World malaria report 2021*, <https://www.who.int/publications/i/item/9789240040496> (2021).
2. Le Roch, K. G. et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
3. Bozdech, Z. et al. Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.* **4**, R9 (2003).
4. Bozdech, Z. et al. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5 (2003).
5. Young, J. A. et al. The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol. Biochem. Parasitol.* **143**, 67–79 (2005).
6. Silvestrini, F. et al. Genome-wide identification of genes upregulated at the onset of gametocytogenesis in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **143**, 100–110 (2005).
7. Lemieux, J. E. et al. Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in ex vivo culture. *Proc. Natl. Acad. Sci. USA* **106**, 7559–7564 (2009).
8. Lu, X. M. et al. Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite *Plasmodium falciparum*. *Nucleic Acids Res.* **45**, 7825–7840 (2017).
9. Reid, A. J. et al. Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *Elife* **7**, <https://doi.org/10.7554/eLife.33105> (2018).
10. Balaji, S., Babu, M. M., Iyer, L. M. & Aravind, L. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006 (2005).
11. Shang, X. et al. Genome-wide landscape of ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network during *Plasmodium falciparum* blood-stage development. *Nucleic Acids Res.* **50**, 3413–3431 (2022).
12. De Silva, E. K. et al. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc. Natl. Acad. Sci. USA* **105**, 8393–8398 (2008).
13. Gomez-Diaz, E. et al. Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in *Anopheles gambiae*. *Sci. Rep.* **7**, 40655 (2017).
14. Gupta, A. P. et al. Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite *Plasmodium falciparum*. *PLoS Pathog* **9**, e1003170 (2013).
15. Bunnik, E. M. et al. The mRNA-bound proteome of the human malaria parasite *Plasmodium falciparum*. *Genome Biol.* **17**, 147 (2016).
16. Lacsina, J. R., LaMonte, G., Nicchitta, C. V. & Chi, J. T. Polysome profiling of the malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **179**, 42–46 (2011).
17. Mair, G. R. et al. Regulation of sexual development of *Plasmodium* by translational repression. *Science* **313**, 667–669 (2006).
18. Shock, J. L., Fischer, K. F. & DeRisi, J. L. Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol.* **8**, R134 (2007).
19. Freitas-Junior, L. H. et al. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* **121**, 25–36 (2005).
20. Petter, M. et al. H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*. *Mol. Microbiol.* **87**, 1167–1182 (2013).
21. Lopez-Rubio, J. J. et al. 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol. Microbiol.* **66**, 1296–1305 (2007).
22. Lopez-Rubio, J. J., Mancio-Silva, L. & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5**, 179–190 (2009).
23. Tonkin, C. J. et al. Sir2 paralogs cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biol.* **7**, e84 (2009).
24. Ukaegbu, U. E. et al. Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in *P. falciparum*. *PLoS Pathog* **10**, e1003854 (2014).
25. Saxena, H. & Gupta, A. *Plasmodium falciparum* PfRUVBL proteins bind at the TARE region and var gene promoter located in the subtelomeric region. *Pathog Dis* **80**, <https://doi.org/10.1093/femspd/ftac018> (2022).
26. Lemieux, J. E. et al. Genome-wide profiling of chromosome interactions in *Plasmodium falciparum* characterizes nuclear architecture and reconfigurations associated with antigenic variation. *Mol. Microbiol.* **90**, 519–537 (2013).
27. Ay, F. et al. Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.* **24**, 974–988 (2014).
28. Bunnik, E. M. et al. Changes in genome organization of parasite-specific gene families during the *Plasmodium* transmission stages. *Nat. Commun.* **9**, 1910 (2018).
29. Read, D. F., Cook, K., Lu, Y. Y., Le Roch, K. G. & Noble, W. S. Predicting gene expression in the human malaria parasite *Plasmodium falciparum* using histone modification, nucleosome positioning, and 3D localization features. *PLoS Comput. Biol.* **15**, e1007329 (2019).
30. Abdi, E. & Latifi-Navid, S. Emerging long noncoding RNA polymorphisms as novel predictors of survival in cancer. *Pathol Res. Pract.* **239**, 154165 (2022).

31. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
32. Corcoran, A. E. The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation. *Semin Immunol* **22**, 353–361 (2010).
33. Gupta, R. A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
34. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
35. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **10**, 925–933 (2013).
36. Ransohoff, J. D., Wei, Y. & Khavari, P. A. The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* **19**, 143–157 (2018).
37. Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.* **17**, 756–770 (2016).
38. Quinodoz, S. & Guttman, M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* **24**, 651–663 (2014).
39. Nakagawa, S. & Kageyama, Y. Nuclear lncRNAs as epigenetic regulators-beyond skepticism. *Biochim. Biophys. Acta* **1839**, 215–222 (2014).
40. Maclary, E., Hinten, M., Harris, C. & Kalantry, S. Long noncoding RNAs in the X-inactivation center. *Chromosome Res.* **21**, 601–614 (2013).
41. Schoeftner, S. & Blasco, M. A. Chromatin regulation and non-coding RNAs at mammalian telomeres. *Semin Cell Dev. Biol.* **21**, 186–193 (2010).
42. Feuerhahn, S., Iglesias, N., Panza, A., Porro, A. & Lingner, J. TERRA biogenesis, turnover and implications for function. *FEBS Lett.* **584**, 3812–3818 (2010).
43. Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M. & Wirth, D. F. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol. Biol. Cell* **12**, 3114–3125 (2001).
44. Gunasekera, A. M. et al. Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. *Mol. Biochem. Parasitol* **136**, 35–42 (2004).
45. Raabe, C. A. et al. A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*. *Nucleic Acids Res.* **38**, 608–617 (2010).
46. Sorber, K., Dimon, M. T. & DeRisi, J. L. RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Res.* **39**, 3820–3835 (2011).
47. Broadbent, K. M. et al. A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol.* **12**, R56 (2011).
48. Wei, C. et al. Deep profiling of the novel intermediate-size non-coding RNAs in intraerythrocytic *Plasmodium falciparum*. *PLoS One* **9**, e92946 (2014).
49. Liao, Q. et al. Genome-wide identification and functional annotation of *Plasmodium falciparum* long noncoding RNAs from RNA-seq data. *Parasitol Res.* **113**, 1269–1281 (2014).
50. Siegel, T. N. et al. Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC Genomics* **15**, 150 (2014).
51. Broadbent, K. M. et al. Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics* **16**, 454 (2015).
52. Lee, V. V. et al. Direct nanopore sequencing of mRNA reveals landscape of transcript isoforms in apicomplexan parasites. *mSystems* **6**, <https://doi.org/10.1128/mSystems.01081-20> (2021).
53. Chappell, L. et al. Refining the transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-free RNA-seq. *BMC Genomics* **21**, 395 (2020).
54. Mourier, T. et al. Genome-wide discovery and verification of novel structured RNAs in *Plasmodium falciparum*. *Genome Res.* **18**, 281–292 (2008).
55. Yang, M. et al. Full-Length Transcriptome Analysis of *Plasmodium falciparum* by Single-Molecule Long-Read Sequencing. *Front Cell Infect Microbiol* **11**, 631545 (2021).
56. Hoshizaki, J. et al. Correction: A manually curated annotation characterises genomic features of *P. falciparum* lncRNAs. *BMC Genomics* **24**, 189 (2023).
57. Epp, C., Li, F., Howitt, C. A., Chookajorn, T. & Deitsch, K. W. Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. *RNA* **15**, 116–127 (2009).
58. Amit-Avraham, I. et al. Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* **112**, E982–E991 (2015).
59. Sierra-Miranda, M. et al. Two long non-coding RNAs generated from subtelomeric regions accumulate in a novel perinuclear compartment in *Plasmodium falciparum*. *Mol. Biochem. Parasitol* **185**, 36–47 (2012).
60. Simantov, K., Goyal, M. & Dzikowski, R. Emerging biology of non-coding RNAs in malaria parasites. *PLoS Pathog* **18**, e1010600 (2022).
61. Rovira-Graells, N. et al. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum* clag3 genes. *Nucleic Acids Res.* **43**, 8243–8257 (2015).
62. Jing, Q. et al. *Plasmodium falciparum* var Gene Is Activated by Its Antisense Long Noncoding RNA. *Front Microbiol* **9**, 3117 (2018).
63. Barcons-Simon, A., Cordon-Obras, C., Guizzetti, J., Bryant, J. M. & Scherf, A. CRISPR Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General Repression of var Genes in *Plasmodium falciparum*. *mBio* **11**, <https://doi.org/10.1128/mBio.03054-19> (2020).
64. Luke, B. & Lingner, J. TERRA: telomeric repeat-containing RNA. *EMBO J.* **28**, 2503–2510 (2009).
65. Filarsky, M. et al. GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* **359**, 1259–1263 (2018).
66. Gomes, A. R. et al. A transcriptional switch controls sex determination in *Plasmodium falciparum*. *Nature* **612**, 528–533 (2022).
67. Bunnik, E. M. et al. DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite *Plasmodium falciparum*. *BMC Genomics* **15**, 347 (2014).
68. Zhang, M. et al. Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360**, <https://doi.org/10.1126/science.aap7847> (2018).
69. Sun, Q., Hao, Q. & Prasanth, K. V. Nuclear long noncoding RNAs: Key regulators of gene expression. *Trends Genet* **34**, 142–157 (2018).
70. Noh, J. H., Kim, K. M., McClusky, W. G., Abdelmohsen, K. & Gorospe, M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **9**, e1471 (2018).
71. Sun, M., Gadad, S. S., Kim, D. S. & Kraus, W. L. Discovery, annotation, and functional analysis of long noncoding RNAs controlling cell-cycle gene expression and proliferation in breast cancer cells. *Mol. Cell* **59**, 698–711 (2015).

72. Clark, M. B. et al. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
73. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
74. Quinn, J. J. et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol.* **32**, 933–940 (2014).
75. Chu, C., Quinn, J. & Chang, H. Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.*, <https://doi.org/10.3791/3912> (2012).
76. Fan, Y. et al. Rrp6 Regulates Heterochromatic Gene Silencing via ncRNA RUF6 Decay in Malaria Parasites. *mBio* **11**, <https://doi.org/10.1128/mBio.01110-20> (2020).
77. Bunnik, E. M. et al. Comparative 3D genome organization in apicomplexan parasites. *Proc. Natl. Acad. Sci. USA* **116**, 3183–3192 (2019).
78. Winter, G. et al. SURFIN is a polymorphic antigen expressed on Plasmodium falciparum merozoites and infected erythrocytes. *J. Exp. Med.* **201**, 1853–1863 (2005).
79. Pei, X. et al. The ring-infected erythrocyte surface antigen (RESA) of Plasmodium falciparum stabilizes spectrin tetramers and suppresses further invasion. *Blood* **110**, 1036–1042 (2007).
80. Zhang, Y., Lin, Y. H., Johnson, T. D., Rozek, L. S. & Sartor, M. A. PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. *Bioinformatics* **30**, 2568–2575 (2014).
81. Silvestrini, F. et al. Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol. Cell Proteomics* **9**, 1437–1448 (2010).
82. Gardiner, D. L. et al. Implication of a Plasmodium falciparum gene in the switch between asexual reproduction and gametocytogenesis. *Mol. Biochem. Parasitol* **140**, 153–160 (2005).
83. Lasonder, E. et al. Integrated transcriptomic and proteomic analyses of P. falciparum gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res.* **44**, 6087–6101 (2016).
84. Tiburcio, M. et al. A 39-Amino-Acid C-Terminal Truncation of GDV1 Disrupts Sexual Commitment in Plasmodium falciparum. *mSphere* **6**, <https://doi.org/10.1128/mSphere.01093-20> (2021).
85. Usui, M. et al. Plasmodium falciparum sexual differentiation in malaria patients is associated with host factors and GDV1-dependent genes. *Nat. Commun.* **10**, 2140 (2019).
86. Tripathi, A. K., Mlambo, G., Kanatani, S., Sinnis, P. & Dimopoulos, G. Plasmodium falciparum Gametocyte Culture and Mosquito Infection Through Artificial Membrane Feeding. *J. Vis. Exp.*, <https://doi.org/10.3791/61426> (2020).
87. Guil, S. & Esteller, M. Cis-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* **19**, 1068–1075 (2012).
88. Orom, U. A. et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
89. Ng, S. Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
90. Prensner, J. R. et al. The long noncoding RNA SCHLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet* **45**, 1392–1398 (2013).
91. Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
92. Mele, M. & Rinn, J. L. “Cat’s Cradling” the 3D Genome by the Act of LncRNA Transcription. *Mol. Cell* **62**, 657–664 (2016).
93. Rinn, J. & Guttman, M. RNA Function. RNA and dynamic nuclear organization. *Science* **345**, 1240–1241 (2014).
94. Guizzetti, J., Barcons-Simon, A. & Scherf, A. Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res.* **44**, 9710–9718 (2016).
95. Abel, S. & Le Roch, K. G. The role of epigenetics and chromatin structure in transcriptional regulation in malaria parasites. *Brief Funct. Genomics* **18**, 302–313 (2019).
96. Cerase, A., Pintacuda, G., Tattermusch, A. & Avner, P. Xist localization and function: new insights from multiple levels. *Genome Biol.* **16**, 166 (2015).
97. Hanly, D. J., Esteller, M. & Berdasco, M. Interplay between long non-coding RNAs and epigenetic machinery: emerging targets in cancer? *Philos. Trans. R Soc. Lond. B Biol. Sci.* **373**, <https://doi.org/10.1098/rstb.2017.0074> (2018).
98. Vance, K. W. & Ponting, C. P. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet* **30**, 348–355 (2014).
99. Wierzbicki, A. T. The role of long non-coding RNA in transcriptional gene silencing. *Curr. Opin. Plant Biol.* **15**, 517–522 (2012).
100. Reininger, L. et al. A NIMA-related protein kinase is essential for completion of the sexual cycle of malaria parasites. *J. Biol. Chem.* **280**, 31957–31964 (2005).
101. Reininger, L. et al. An essential role for the Plasmodium Nek-2 Nima-related protein kinase in the sexual development of malaria parasites. *J. Biol. Chem.* **284**, 20858–20868 (2009).
102. Tomas, A. M. et al. P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. *EMBO J.* **20**, 3975–3983 (2001).
103. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
104. Joshi NA, F. J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. [Software]. (Version 1.33) (2011).
105. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
106. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
107. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
108. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
109. Howick, V. M. et al. The Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. *Science* **365**, <https://doi.org/10.1126/science.aaw2619> (2019).
110. Jiang, L. et al. PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. *Nature* **499**, 223–227 (2013).
111. Sonoiki, E. et al. A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue. *Nat. Commun.* **8**, 14574 (2017).
112. Ribeiro, J. M. et al. Guide RNA selection for CRISPR-Cas9 transfections in Plasmodium falciparum. *Int. J. Parasitol* **48**, 825–832 (2018).
113. Ribaut, C. et al. Concentration and purification by magnetic separation of the erythrocytic stages of all human Plasmodium species. *Malar J.* **7**, 45 (2008).
114. Adjalley, S. H., Lee, M. C. & Fidock, D. A. A method for rapid genetic integration into Plasmodium falciparum utilizing mycobacteriophage Bxb1 integrase. *Methods Mol. Biol.* **634**, 87–100 (2010).

## Acknowledgements

This work was supported by NIH grants to KGLR (nos. 1R01 AI136511 and R21 AI142506-01) and by the University of California, Riverside to KGLR (no. NIFA-Hatch-225935). We also thank the parasitology and insectary core facilities at the JHMRI and Bloomberg Philanthropies for their support of these facilities.

## Author contributions

Conceptualization was the responsibility of KGLR that also supervised the project together with ML, WSN and PS. Methodology was performed by GB, XML, SA, ZC, TH, DW, TW, AC, TL, JP, AKT, GX, JC, TAT and SD; GB, XML, SA, BH and KGLR carried out the analysis. Software, formal analysis and data curation were provided by SA, BH and SD. GB, ZC, SA and KGLR wrote the original draft. All authors reviewed and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40883-w>.

**Correspondence** and requests for materials should be addressed to Karine G. Le Roch.

**Peer review information** : *Nature Communications* thanks Elena Gómez-Díaz, José Quilles Junior and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023