# Cell Reports
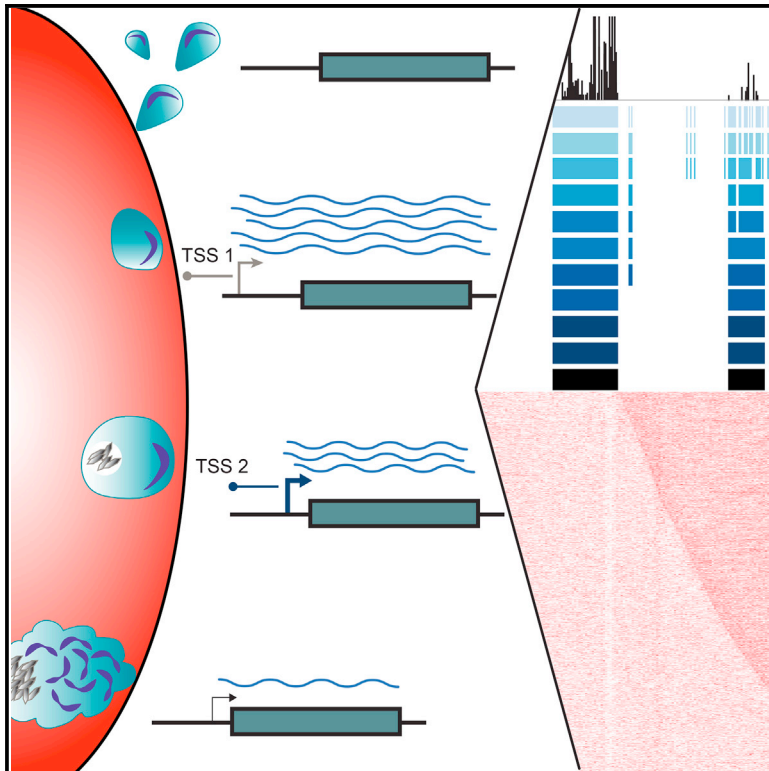
# Landscape and Dynamics of Transcription Initiation in the Malaria Parasite *Plasmodium falciparum*

## Graphical Abstract



## Highlights

- A genome-wide map of transcription start sites (TSSs) for *P. falciparum*

- Sequence and chromatin features for core promoters of malaria parasites are defined

- Dynamic study of TSS usage identifies developmentally regulated transcript isoforms

## Authors

Sophie H. Adjalley,
Christophe D. Chabbert, Bernd Klaus,
Vicent Pelechano, Lars M. Steinmetz

## Correspondence

larsms@embl.de

## In Brief

The genome of *P. falciparum* is characterized by an extremely AT-biased nucleotide composition. Adjalley et al. conducted a survey of transcription initiation events throughout the malaria parasite's blood cycle. They detect coordinate and stage-specific transcription initiation and extract sequence and chromatin signatures from promoter regions.

## Accession Numbers

GSE68982

CrossMark

CellPress

# Landscape and Dynamics of Transcription Initiation in the Malaria Parasite *Plasmodium falciparum*

Sophie H. Adjalley,[1,4,5] Christophe D. Chabbert,[1,4,6] Bernd Klaus,[1] Vicent Pelechano,[1,7] and Lars M. Steinmetz[1,2,3,*]

[1]Genome Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
[2]Stanford Genome Technology Center, Palo Alto, CA 94304, USA
[3]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Co-first author
[5]Present address: Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
[6]Present address: AstraZeneca, Oncology iMed, CRUK Cambridge Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK
[7]Present address: SciLifeLab, Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 171 65 Solna, Sweden
*Correspondence: larsms@embl.de
http://dx.doi.org/10.1016/j.celrep.2016.02.025

## SUMMARY

A comprehensive map of transcription start sites (TSSs) across the highly AT-rich genome of *P. falciparum* would aid progress toward deciphering the molecular mechanisms that underlie the timely regulation of gene expression in this malaria parasite. Using high-throughput sequencing technologies, we generated a comprehensive atlas of transcription initiation events at single-nucleotide resolution during the parasite intra-erythrocytic developmental cycle. This detailed analysis of TSS usage enabled us to define architectural features of plasmodial promoters. We demonstrate that TSS selection and strength are constrained by local nucleotide composition. Furthermore, we provide evidence for coordinate and stage-specific TSS usage from distinct sites within the same transcription unit, thereby producing transcript isoforms, a subset of which are developmentally regulated. This work offers a framework for further investigations into the interactions between genomic sequences and regulatory factors governing the complex transcriptional program of this major human pathogen.

## INTRODUCTION

Elucidation of the 23-Mb genome sequence of *P. falciparum*, the protozoan parasite responsible for the most lethal form of human malaria, revealed that its global AT content is greater than 80% and rises to 90% in introns and intergenic regions (Gardner et al., 2002). This extreme base composition raised questions on the genome organization and the mechanisms enabling gene regulation in such an AT-rich environment. Transcriptome analyses using microarrays, and later confirmed by RNA sequencing (RNA-seq) methods, showed that gene expression is tightly controlled during the parasite intra-erythrocytic developmental cycle (IDC) (Bozdech et al., 2003a; Le Roch et al., 2004; Otto et al., 2010; Sorber et al., 2011). Indeed, the majority of the ~5,500 genes vary significantly in steady-state mRNA levels between the different intra-erythrocytic stages, resulting in a developmentally linked cascade of gene expression. Genome surveys showed conservation of the basal eukaryotic transcriptional machinery but also suggested a paucity of transcription-associated factors (Bischoff and Vaquero, 2010; Callebaut et al., 2005; Coulson et al., 2004). Despite major advances comprising the identification of transcription factors that govern *P. falciparum* stage-specific gene expression (Balaji et al., 2005; De Silva et al., 2008), the lack of a comprehensive map of transcription initiation events has hampered advances in decrypting the molecular basis of transcriptional control. For instance, the mechanisms enabling the transcription machinery to cope with the exceptionally skewed nucleotide composition of the parasite genome or the basic organization of transcription units have yet to be determined. More particularly, the processes underlying the regulation of transcription initiation, such as the recruitment of RNA polymerase II (RNA PolII) or transcription start site (TSS) selection, still remain to be deciphered. Generally, characterizing the transcriptome architecture of *P. falciparum* has been technically challenging, given the low complexity of its genome and the potential for reverse-transcription-derived artifacts (Siegel et al., 2014). Nevertheless, several single-gene studies have attempted to identify transcription initiation sites (Horrocks et al., 2009), while the Full-Malaria project mapped TSS positions for ~25% of the *P. falciparum* genes by cloning and, subsequently, sequencing full-length cDNA molecules (Watanabe et al., 2002). However, the intricacy and dynamics of transcription initiation in this parasite have not yet been addressed. The establishment of next-generation sequencing technologies now provides powerful means for the comprehensive and systematic analysis of gene expression and regulation. Genome-wide approaches such as CAGE (cap analysis of gene expression) have contributed to a better understanding of eukaryotic promoter structures and the impact of local nucleotide content on transcriptional processes (Shiraki et al., 2003). Recent characterization of the eukaryotic landscape of transcription initiation by RNA PolII has, for instance, revealed the existence of "broad" and "sharp"

CrossMark

promoter classes (Lenhard et al., 2012) and shed some light on the mechanisms underlying the selection of transcription start sites (TSSs) (Haberle et al., 2014).

Here, we report an in-depth analysis of the dynamics of transcription initiation during the *P. falciparum* IDC, giving insight into the parasite's transcriptome architecture and gene expression regulation. Using 5′ cap sequencing, a modified version of the CAGE approach that alleviates some of the artifacts previously mentioned, we systematically characterized *P. falciparum* transcript 5′ ends at different stages of the 48-hr IDC, thereby generating a genome-wide TSS map at single-nucleotide resolution. We demonstrated that transcription initiation occurs across broad genomic regions and that multiple clusters of TSSs may co-exist within a gene locus to generate mRNA isoforms. Strikingly, examination of the temporal data revealed that transcription initiation from separate sites within the same transcription unit may be coordinated or stage specific, leading to the production of developmentally regulated and, possibly, differentially translated transcript isoforms for a subset of *P. falciparum* genes. Analysis of the nucleotide content and chromatin organization around the identified TSS allowed us to outline features of *P. falciparum* promoter architecture. In particular, we found that, even in the parasite's extremely AT-biased genome, promoter selection and activity are governed by a locally skewed base composition. Given the great complexity of the *P. falciparum* transcriptional program, this report constitutes a highly valuable resource for further investigations into the mechanisms directing TSS selection, including the links between *P. falciparum* promoter architecture and regulatory elements. Our study also provides a framework for the development of therapeutics that interfere with gene expression regulation in this deadly pathogen, to which half of the world population remains exposed (World Health Organization, 2015).

## RESULTS

### Characterization of *P. falciparum* Transcription Initiation Patterns Reveals TSS Distributions over Broad Promoter Regions

An in-depth survey of the transcription initiation events occurring across the *P. falciparum* developmental stages was conducted from two tightly synchronized biological replicate cultures harvested at six equally spaced time points during the 48-hr IDC (Supplemental Experimental Procedures).

Strand-specific libraries for deep sequencing were constructed from the 5′ ends of capped mRNAs captured using a biotinylated 5′ adaptor containing unique molecular identifiers, as reported by Pelechano et al. (2015). In addition to calf intestine phosphatase (CIP) and tobacco acid phosphatase (TAP) treatments, enrichment for capped mRNA molecules was further ensured by enzymatic treatment of the total RNA samples with a 5′ P-dependent exonuclease. Omitting the TAP treatment resulted in a very small number of sequencing reads mapping to the *P. falciparum* genome, indicative of a negligible level of experimental noise (Supplemental Experimental Procedures). Some modifications were made to take into consideration the exceptional AT richness of the *P. falciparum* genome (Figure 1A;

Figures S1A and S1B; Supplemental Experimental Procedures). For instance, the mRNA fragmentation and the reverse transcription reaction were performed using conditions previously established for RNA-seq and DNA microarrays in *P. falciparum* (Bozdech et al., 2003b; Hoeijmakers et al., 2013a). Additionally, the second-strand synthesis and the amplification of the cDNA libraries were performed using the KAPA DNA polymerase enzyme, which has a very low AT bias and is now commonly used when preparing *P. falciparum* sequencing libraries (Oyola et al., 2012; Siegel et al., 2014). About 600,000 unique 5′-long tags (>90 nt) were recovered, on average, per independent library (Figure S1C; Table S1) and collapsed on their first base to define the position of the TSSs at single-nucleotide resolution. Compilation of the transcription initiation events detected on both positive and negative strands showed that more than 3 million nucleotide bases constituted a TSS, indicative of an extensive transcription initiation activity across the *P. falciparum* genome (Table S1). The use of in vitro transcripts to control for the robustness of the protocol confirmed that the 5′ end of these transcripts were properly identified using the 5′ cap capture (Figure S1D; Supplemental Experimental Procedures).

Transcription initiation events in *P. falciparum* appeared to occur at multiple closely spaced sites, across relatively wide regions (Figure 1B), indicative of a broad promoter architecture similar to other eukaryotic organisms (Lenhard et al., 2012). Therefore, we implemented an analytical method based on mathematical morphology operations and filtering tools (Heijmans and Ronse, 1989) that adapts to the various types of captured signals (dispersed or dense distribution, for instance). Instead of clustering the TSSs within an empirically chosen window size across the genome, it allowed the definition of separate regions of transcription initiation, or "TSS blocks," within each gene locus in a systematic way (Figures 1C and S2; Supplemental Experimental Procedures). The use of alternating sequential filters (Heijmans and Ronse, 1989) on the pooled data (i.e., combining replicates and developmental stages) led to the definition of more than 44,000 potential TSS blocks genome-wide. These distinct regions of transcription initiation were associated, where possible, to one of the annotated transcription units (Figure S2; Supplemental Experimental Procedures) to constitute a high-resolution atlas of TSSs for the *P. falciparum* transcriptome. Biological replicates exhibited an excellent correlation (Spearman correlation > 0.9), thereby confirming the robustness of our protocol (Figure S1E). In total, between the two replicates, transcription initiation events were detected for 90% of the *P. falciparum* annotated genes (4,955 out of 5,510). The vast majority (92%) of these contained more than one cluster of transcription initiation sites (median of ~6 TSS blocks), thereby expanding on previous observations from the partial analysis of full-length cDNAs (Watanabe et al., 2002). This striking finding suggests that the majority of the genes not only use multiple sites, as evidenced by the dispersed patterns of transcription initiation, but may also utilize multiple promoter regions to initiate transcription (Figure 1D). Importantly, this confirms the additional level of resolution provided by our approach in comparison to that of current RNA-seq technologies.
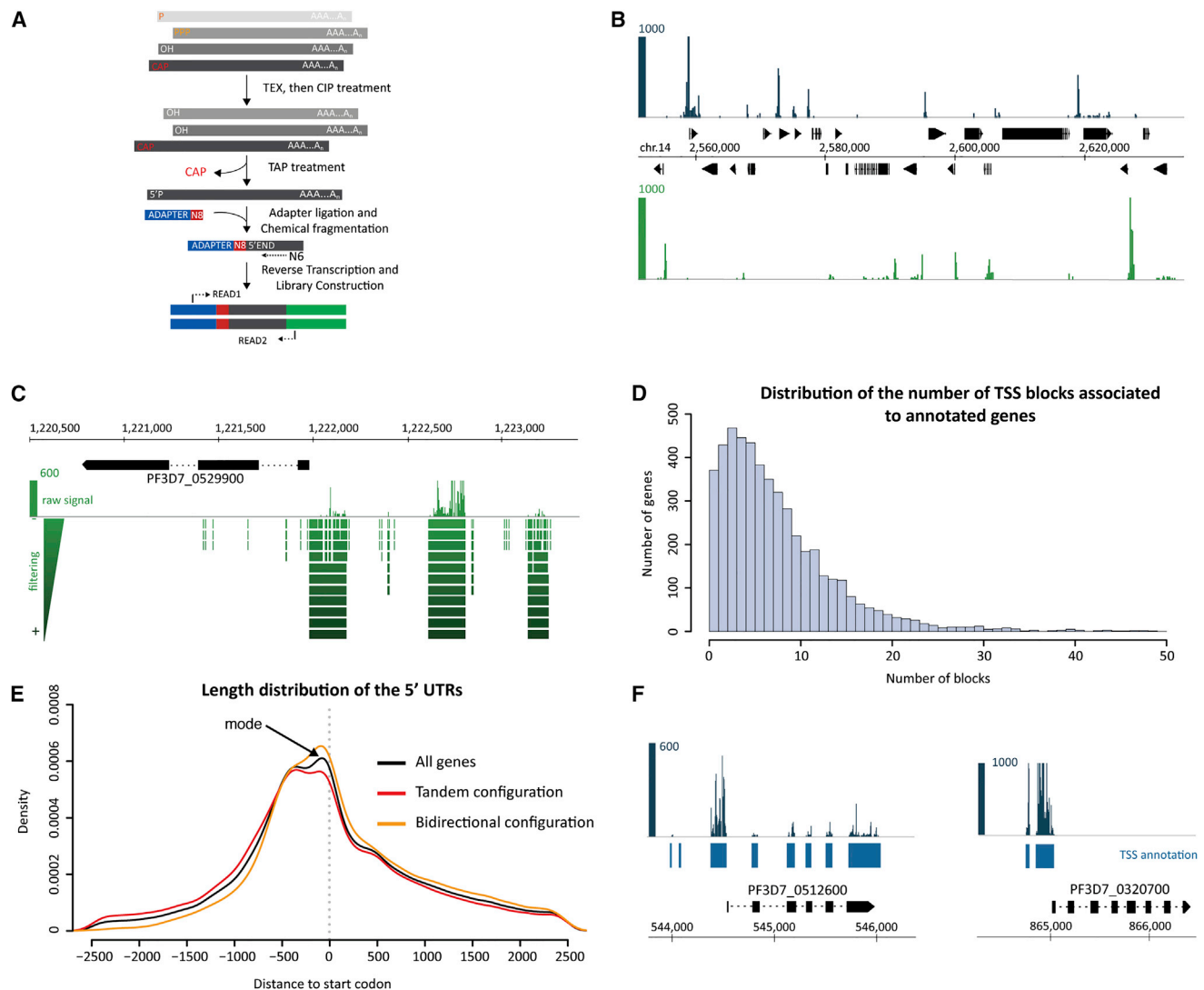
**Figure 1. Genome-wide Identification of TSS during the *P. falciparum* Asexual Cycle**

(A) Capped mRNA molecules are captured using direct ligation of barcoded RNA adapters to transcript 5′ ends. Total RNA samples are sequentially treated with a 5′ P-dependent exonuclease (TEX) and an alkaline phosphatase (CIP) to enrich for capped mRNA molecules. Illumina libraries are then amplified after reverse transcription of the tagged mRNA fragments.

(B) Coverage track showing the TSSs mapped to a locus on chromosome 14. Filtered reads from all time points and biological replicates were pooled and collapsed on their first base pair to visualize sites of initiation. The scale indicates the read counts at each genomic position.

(C) Coverage track illustrating the annotation obtained using alternate sequencing filters. Read counts for the raw signal are indicated. Green boxes show the TSS block annotation at each step of the incremental filtering.

(D) Distribution of the number of TSS blocks associated with previously annotated transcription units.

(E) Length distribution of the 5′ UTRs derived from our annotation. While most TSSs are located upstream of the start codon, a large number of transcripts have very short (<100 bp) or no 5′ UTRs due to initiation within gene bodies (black). Length distribution of the 5′ UTRs for genes in a tandem (red) and bi-directional configuration (orange) is also shown.

(F) Examples of TSS distributions in multi-exon genes. Transcription widely initiates within the body of each exon of PF3D7_0512600, whereas TSS clusters are only detected upstream of the start codon for PF3D7_0320700. TSS block annotation is shown in blue.

See also Figures S1 and S2.

## Transcription in *P. falciparum* May Initiate Close to the Coding Sequence of Genes

Given that the majority of genes appeared to use multiple sites to initiate transcription, we examined their global distribution relative to the coding region (CDS) of the associated genes. This analysis showed that 81% of the total TSS blocks were positioned less than 1,000 bp upstream of the start codon, and more than 65% of these located within a distance of 500 bp or less upstream of the CDS (Figure 1E). Interestingly, the mode of the TSS block distribution was within 50 bp upstream of the
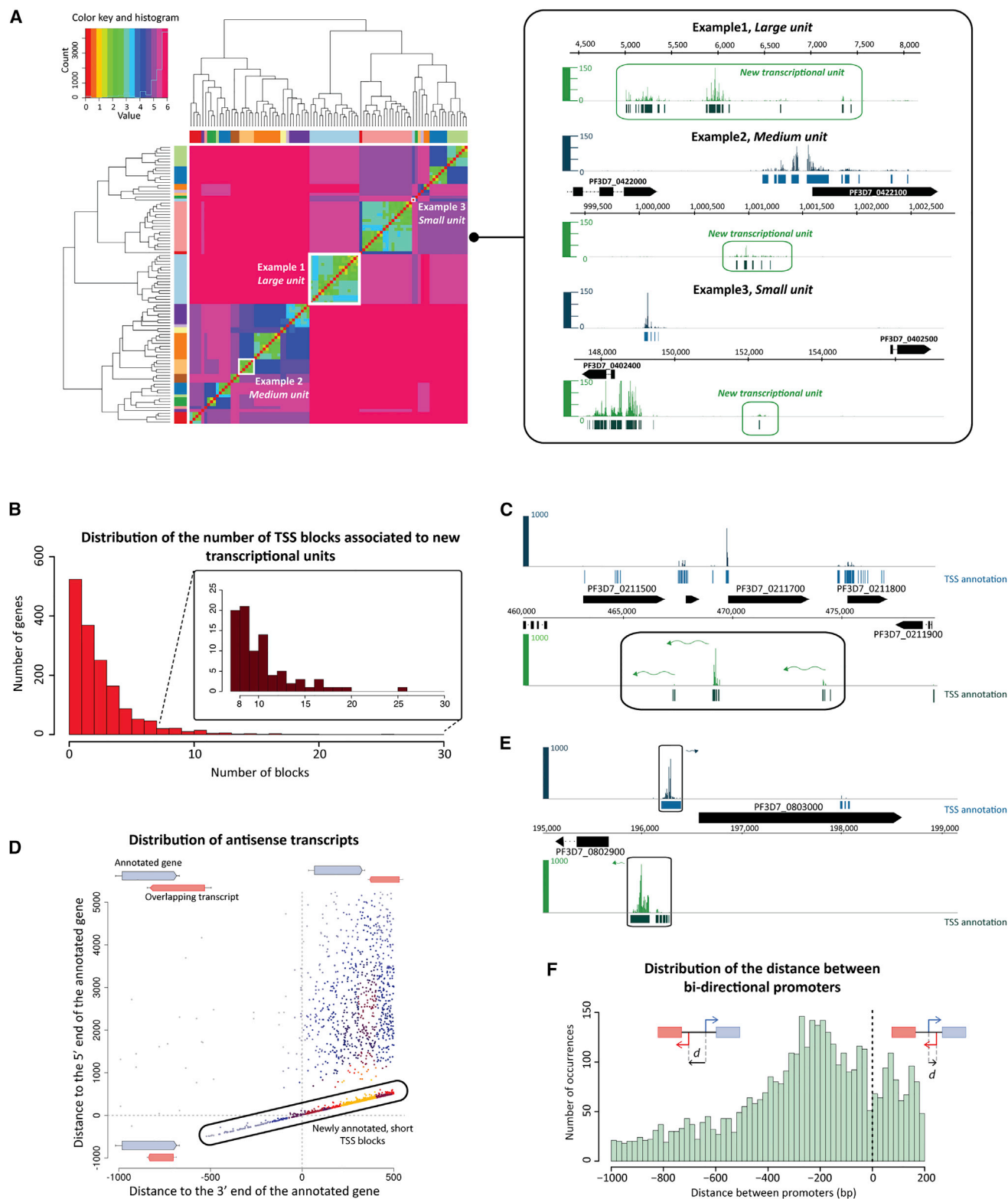
**Figure 2. Global Organization of the TSS along the *P. falciparum* Genome**

(A) Clustering of new transcription units on the (−) strand of chromosome 4. TSS blocks were clustered using the distance to their neighbors and the final dendogram was segmented after manually evaluating the number of relevant clusters. Snapshots of coverage tracks on the right show examples of large, medium, and small transcription units.

start codon (Figure 1E), suggesting transcription initiation in regions very close to the start of the CDS that may produce leader-less transcripts, i.e., with little or no 5′ UTR.

Our analysis further identified transcription initiation in non-conventional sites, such as within the CDS of single-exon genes, but also in the exons of multi-exon genes (Figure 1F). Individual instances of such configurations were also observed in the Full-Parasites dataset (Tuda et al., 2011) (Figures S3A and S3B). 49% of all identified TSS blocks were actually located downstream of the start codon. This indicates a significant amount of transcription initiation events within the CDS that potentially leads to truncated protein products as observed in yeast (Fournier et al., 2012). The number of internal TSS blocks that are directly followed by an ATG codon in frame with the annotated coding sequence (∼90%) suggests that it might also be the case in *P. falciparum*. However, some of the sequencing tags recovered in these regions may also result from the recapping of mRNA degradation products or non-coding RNAs (Schoenberg and Maquat, 2009).

Further analysis showed that the distribution of the TSS blocks was similar whether genes displayed a tandem or a bidirectional configuration (Figure 1E) and was not markedly affected by the presence of introns (Figure S3C). Increased length of the intergenic region was associated with a wider spread of the TSSs upstream of the start codon, with the TSSs closer to the CDS for genes separated by a short distance (Figure S3D). Annotated transcription units for which no TSS was identified corresponded, for the vast majority, to antigenic variant families and to genes that are expressed in the other stages of the parasite's infectious cycle (Table S2A). Nonetheless, there were also genes reported to be expressed during the IDC for which we did not detect any transcription initiation events. Among these, some may be genes that produce uncapped transcripts that our approach would not permit to capture or that are expressed at a level below our detection threshold.

### *P. falciparum* Transcriptome Exhibits Widespread Intergenic and Bidirectional Promoter Activities

An appreciable number of TSS blocks (4,600; i.e., 10%, across all 14 chromosomes) could not be assigned to any of the gene loci found in the current annotation of the *P. falciparum* genome. These were, therefore, categorized as "new blocks of transcription initiation" and manually clustered (Figure 2A; Supplemental Experimental Procedures) to isolate more than 1,500 "potential transcription units" (Figure 2B; Table S2B). Most of these clusters are located in non-coding regions, or within very large

intergenic regions between annotated genes, suggesting that a large fraction of the *P. falciparum* non-coding genome is actively transcribed (Raabe et al., 2010). Such non-coding transcription could play a role in regulating gene expression, for instance, via recruitment of chromatin modifiers or by strengthening transcriptional activities through interactions with the transcriptional machinery (Barrandon et al., 2008). This postulate extends to the telomeric and subtelomeric regions of the parasite genome, in which a few non-coding transcripts had been detected and linked to transcriptional control and silencing (Broadbent et al., 2011; Epp et al., 2009; Kyes et al., 2007). We generalized these findings with the identification of transcripts emerging from the telomeric regions of 11 out of all 14 chromosomes and within introns of at least 80% of all *var* genes (Figures S3E and S3F; Tables S3A and S3B).

Further pervasive transcription was observed with 31% of the 4,955 examined genes displaying antisense transcription initiation (Figure 2C). A large fraction (47%) of the ∼2,000 antisense TSS blocks identified genome-wide resulted from non-coding transcription. Interestingly, a third of these TSS blocks directly overlapped with the 3′ end of the sense annotated genes (Figure 2D), thereby plausibly interfering with the regulation of the derived sense transcripts. Further analysis showed that 33% of the antisense TSS blocks were located in promoter regions where transcription may initiate bidirectionally (Figure 2C). Such regions were prevalent, with the presence of TSS blocks in a divergent arrangement being observed for about half of all annotated transcription units (Figure 2E). In more than 70% of such regions, the pairs of divergent TSS blocks were separated by a distance of 400 bp or less from each other, arguing for the likely presence of bidirectional promoters with shared regulatory elements (Figure 2F). This type of configuration may allow for the co-expression of adjacent genes or, alternatively, the directional control of gene expression through the binding of regulators to separate promoter elements.

### Core Promoter Regions Are Defined by a Specific Chromatin Organization that Includes a Precise Positioning of the +1 Nucleosome

Generating a genome-wide map of transcription initiation events enabled us to assess the structural properties that may define TSS choice in the malaria parasite. Among these, local chromatin organization plays an important role in the regulation of eukaryotic transcription initiation through the association of specific histone marks and variants with active promoters. Therefore, we examined the dynamics of H3K4me3-, H3K9ac-,

---

(B) Distribution of the number of TSS blocks associated with the clustered transcription units. The enlargement illustrates the presence of transcription units with a high number of TSS blocks (>7 per unit).
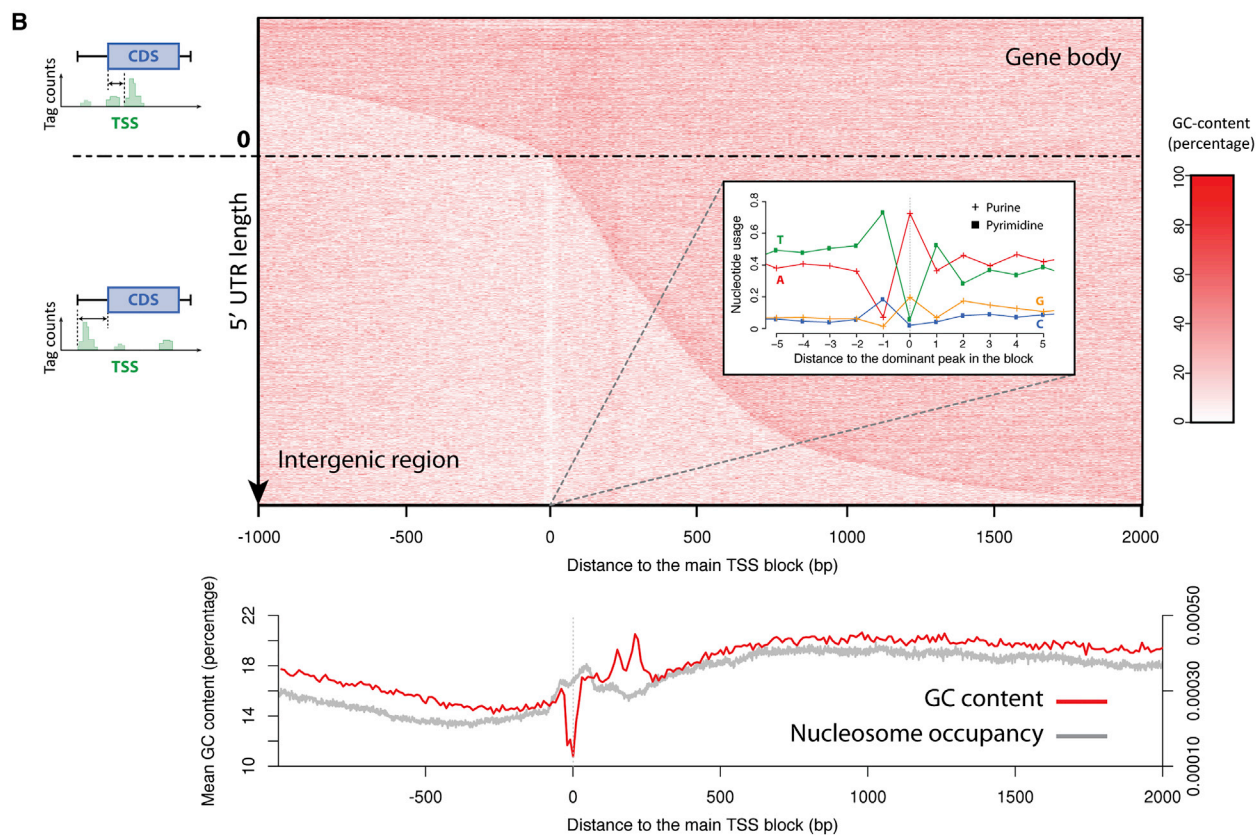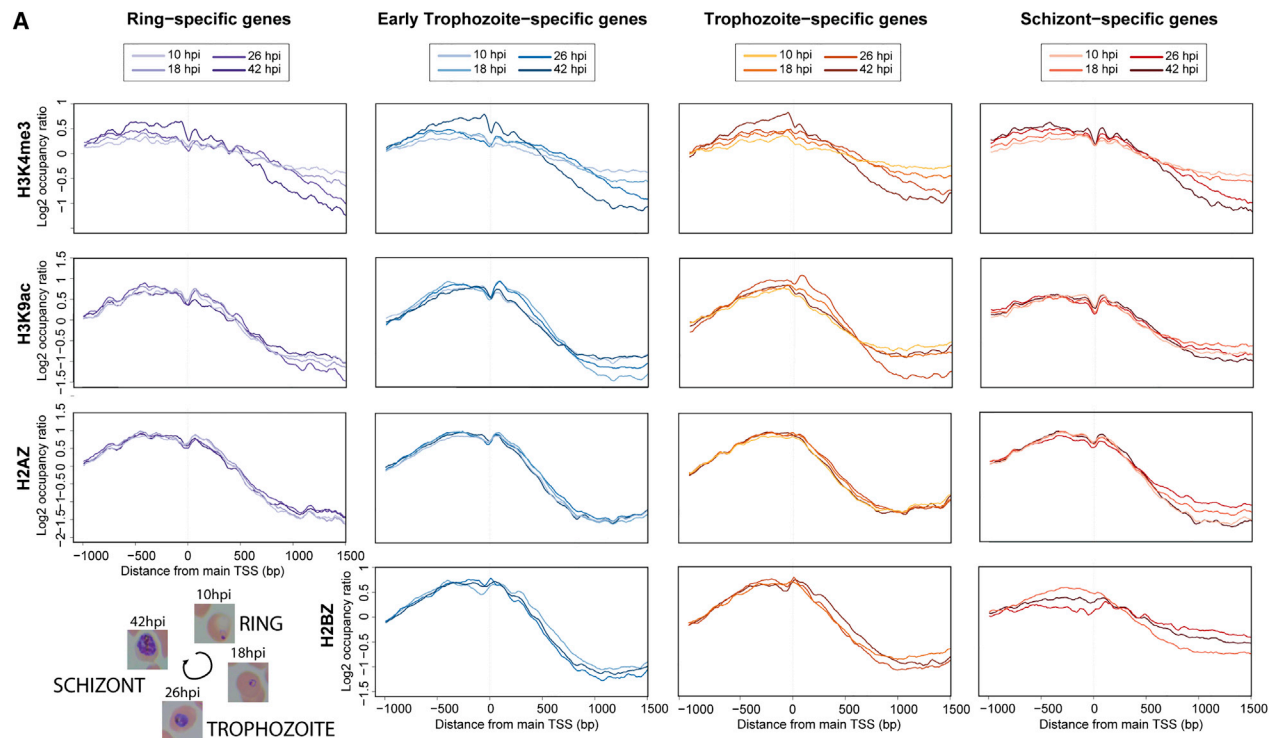
(C) Coverage track of a genomic locus that contains several newly annotated antisense transcripts potentially overlapping with the 3′ end of genes located on the opposite strand. TSS block annotations are indicated in blue (+ strand) and green (− strand).

(D) Distribution of the various classes of antisense transcripts. The distance from the end of the antisense transcriptional unit (i.e., group of antisense TSS blocks assumed to derive from the same transcriptional unit) to the 3′ end of the CDS of annotated sense transcripts is plotted against the distance from the start of the antisense transcriptional unit to the 3′ end of the CDS of annotated sense transcripts. This distribution reveals that most antisense transcripts initiate close to the 3′ end of annotated genes.

(E) Coverage track depicting a bidirectional promoter in which transcription initiates in the opposite direction from sites located a few hundred base pairs apart.

(F) Distribution of the distance (d) between the 5′ ends of bidirectional promoter elements. Negative distances correspond to TSSs located upstream of the 5′ reference point, while positive values correspond to TSSs located downstream (and, therefore, potentially resulting in overlapping transcripts).

See also Figure S3.

(legend on next page)

H2A.Z-, and H2B.Z occupancies around the most active TSSs, using publicly available datasets of chromatin immunoprecipitation sequencing (ChIP-seq) (Bártfai et al., 2010; Hoeijmakers et al., 2013b) (Figure 3A; Supplemental Experimental Procedures).

Alignment of all transcripts at the start position of their most active TSS block showed predominant enrichment in these histone post-translational modifications and variants at the 5′ end of genes (Figure 3A), as in other eukaryotic species. H3K4me3- and H3K9ac-marked nucleosomes were precisely positioned at the +1 position, i.e., the first nucleosome downstream of the TSS block, suggesting that the strong relationship between TSS and position of the +1 nucleosome is maintained in the malaria parasite (Struhl and Segal, 2013). Nevertheless, both marks displayed highly discernible occupancy profiles. Globally, H3K4me3 enrichment increased steadily, as the parasite progressed through the IDC, with a stronger marking of the +1 and +2 nucleosomes and a more pronounced nucleosome-free region (NFR) by the end of the cycle (Figure 3A). In contrast, H3K9ac enrichment at the +1 nucleosome position increased progressively before dropping at the schizont stage to a level equivalent to, or lower than, that observed at early developmental stages (Figure 3A). Enrichment for the acetylation mark around the TSSs was more pronounced with a prominent NFR at the trophozoite stage, during which the bulk of transcriptional activity occurs (Sims et al., 2009).

Interestingly, the overall profiles of H3K4me3 enrichment were highly similar between groups of genes expressed at different developmental stages, confirming the disconnect between H3K4me3 occupancy and genes' transcriptional activity (Figure 3A) (Bártfai et al., 2010). Genes specifically expressed late in the parasite IDC, on the other hand, exhibited very little variation in occupancy for H3K9ac-marked nucleosomes around the TSSs between time points (Figure 3A). This trend was also observed for nucleosomes containing the histone variants H2A.Z and H2B.Z (Figure 3A). Their occupancy profiles appeared constant and similar between genes expressed at distinct stages of the parasite life cycle, with a characteristic enrichment at the +1 nucleosome (Mavrich et al., 2008) and a stronger depletion toward the 3′ end of transcripts. In contrast, genes preferentially active during schizogony exhibited an increased enrichment for H2A.Z in the 5′ region of the transcripts near the TSS, in comparison to regions further downstream in the gene body, especially at the beginning and end of the IDC. This

differed from the more even distribution of H2B.Z on either side of the TSS, which produced a generally flatter profile.

Altogether, these observations indicate that, although chromatin organization around the TSS displays expected features such as a marked positioning of the +1 nucleosome downstream of the TSS, the presence of the examined histone marks and variants do not tightly associate with transcriptional activation as observed in other eukaryotes.

### Regions of Transcription Initiation in *P. falciparum* Are Characterized by a Markedly Skewed Nucleotide Composition at and Downstream of the TSS

In addition to examining the chromatin landscape around the sites of transcription initiation, we also assessed whether *P. falciparum* core promoters are characterized by a particular nucleotide signature. Analysis of the base composition in the core promoter region after aligning all genomic sequences at the most active TSS block revealed a strong decline in GC content around the position of transcription initiation (Figure 3B). Such a decline was also visible, albeit within a narrower region, around the TSSs detected in the body of mono- or multi-exonic genes, suggesting that many of the internal TSSs are genuine sites of transcription initiation rather than biological or experimental artifacts (Figure S3G). When looking at the dominant TSS peak within the block, we observed that transcription in the parasite preferentially initiates with the pyrimidine-purine di-nucleotide T-A at position −1, 0 (Figure 3B). This trend was independent of the activity of the considered TSS block (Figure S3H). In contrast, no specific di-nucleotide composition was observed when randomly selecting a genomic position within the blocks (Figure S3H). Surprisingly, we also detected the presence of two well-defined peaks corresponding to a local increase in G/C about 150 bp and 210 bp downstream of the TSSs, respectively, equally visible downstream of the internal exonic TSSs (Figures 3B and S3G). Alignment of the profile of GC content and that of nucleosome positioning derived from Bunnik et al. (2014) revealed that these two peaks coincided with a dip in nucleosome occupancy. Nucleotide composition around the TSSs did not vary strikingly with the strength of the promoter, except for a greater GC content right at the borders of the TSS blocks and, in particular, downstream of those with a lower activity (i.e., less frequent usage) (Figures S3I and S3J). This change was accompanied by a shift of the +1 nucleosome

---

**Figure 3. Architecture of Active Promoters during the *P. falciparum* Asexual Cycle**

(A) Occupancy of histone post-translational modifications and variants around stage-specific promoters. The occupancy profiles are organized by mark/variant (rows in the table) and stage-specific genes (columns in the table; ring-specific genes are in purple, early trophozoite-specific genes are in blue, trophozoite-specific genes are in orange, and schizont-specific genes are in red). In each case, the occupancy profiles were plotted for each time point at which the mark or variant was profiled (Bártfai et al., 2010, Hoeijmakers et al., 2013b). Color shades from light (early time point) to dark (late time point) represent progression through the IDC. Only the most active TSS block was considered for each gene, and all ChIP signals were normalized using the chromatin input signal. Stage-specific promoters were called using the counts obtained from our study (Supplemental Experimental Procedures). This global view illustrates developmentally regulated changes in the distribution of histone marks and variants around *P. falciparum* promoters. hpi, hours post-invasion.

(B) Local nucleotide content and nucleosome occupancy around the most active TSS block. The heatmap illustrates the local GC content in 10-bp bins around the TSS block for each gene with an annotated TSS block. Analysis of the nucleotide composition in a symmetrical 10-bp window around the dominant TSS peak within the block shows the preferential pyrimidine (T)-purine (A) di-nucleotide used to start transcription. The average profile below the heatmap reveals the drop in GC content around the site of initiation, as well as the +1 nucleosome barrier. The additional downstream peaks enriched in GC nucleotides are also apparent and delimit a decrease in nucleosome occupancy. Nucleosome occupancy was plotted using the micrococcal nuclease sequencing (MNase-Seq) data from Bunnik et al. (2014).

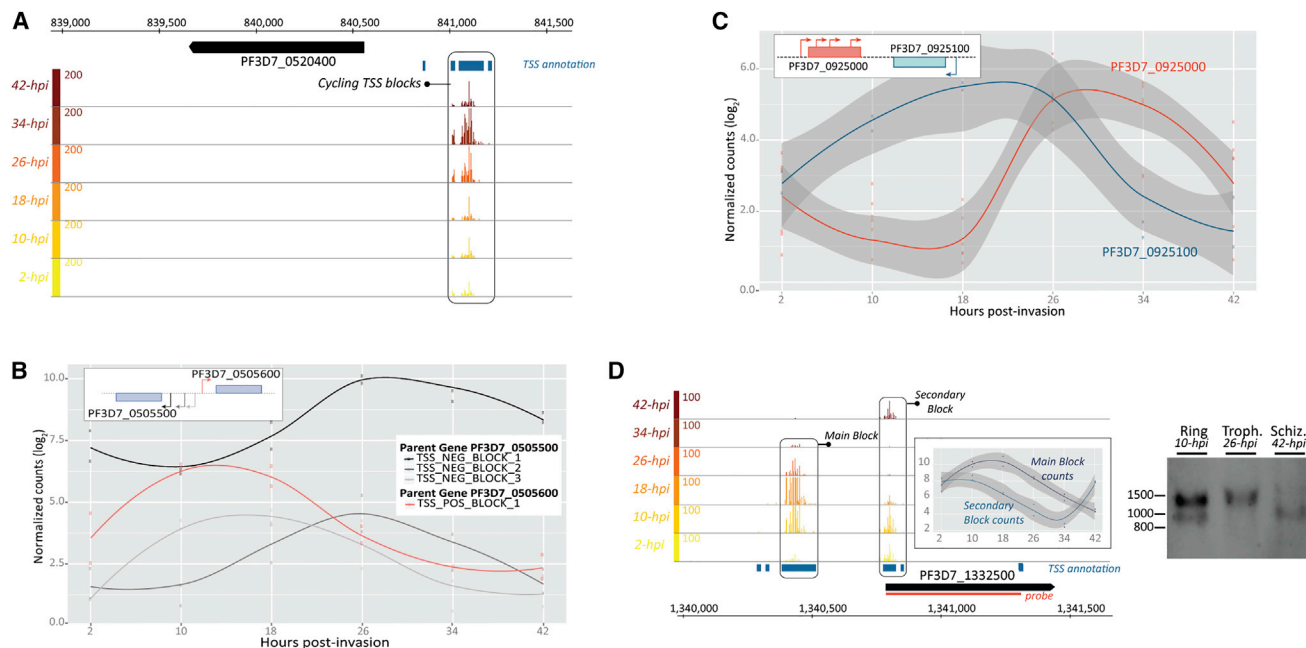See also Figure S3 and Supplemental Experimental Procedures.

**Figure 4. Dynamic Usage of Transcription Initiation Sites during *P. falciparum* Intra-erythrocytic Development**

(A) Coverage tracks illustrating the periodic transcription initiation events associated with the gene PF3D7_0520400. This cycling transcriptional activity is characteristic of the reported cascade of gene expression during the parasite IDC. hpi, hours post-invasion.

(B) Normalized counts of a set of asynchronous TSS blocks associated with a pair of bidirectional genes (PF3D7_0505500/PF3D7_0505600). The timing of maximal expression for the four blocks defines two classes of TSS clusters with different expression patterns and yet sometimes associated with the same gene.

(C) Normalized counts of a pair of genes (PF3D7_0925000/PF3D7_0925100) in antisense orientation and displaying temporal usage of TSS blocks (marked as dots) in opposite phase.

(D) Example of a gene, PF3D7_1332500, harboring two types of TSS blocks with different temporal usage during the IDC. The coverage tracks and normalized counts indicate a strong increase in activity at 18 hpi for the main TSS block, while the secondary block near the start of the gene coding sequence is mostly used at the early and late stages. Northern blot analysis clearly illustrates the changes in relative abundance of the two populations of RNA molecules between stages. These include a long transcript emerging from the secondary TSS block, whose length suggests that it might fully overlap with the gene coding sequence and thus constitute a leader-less transcript. The genomic location covered by the northern blot probe is highlighted in orange. Troph., trophozoite; Schiz., schizont. See also Figure S4.

and a disappearance of the NFR around the TSSs, indicative of a direct link between nucleotide content, nucleosome positioning, and transcriptional activity.

### Dynamic Analysis of TSS Usage during *P. falciparum* IDC Reveals Alternative Transcription Initiation and Possibly Shared Regulatory Elements

The generation of our genome-wide TSS annotation enabled us to examine the dynamics of transcription initiation during the *P. falciparum* IDC and identify significant differences in TSS usage between developmental stages. Stringent filters using the replicate information were applied (Figure S4A; Supplemental Experimental Procedures) and confirmed extensive transcriptional activity during the *P. falciparum* IDC. In total, transcription initiation events conserved across replicates were detected for 74% of all 5,510 protein-coding genes (local false discovery rate [FDR] < 0.1), and 68% of all annotated transcription units (4,838 out of 7,090, FDR < 0.1) that include the new transcription units, as defined earlier.

We found that the majority of the actively transcribed genome (91% of the 4,738 transcription units to which active TSS blocks were associated) exhibited a cycling behavior throughout the

parasite developmental cycle ($\log_2$ fold change > 0.5; Figure 4A; Figure S4B). These numbers are in agreement with previous microarrays and RNA-seq studies and confirm that the parasite transcriptional cascade is recapitulated at the TSS level. Furthermore, comparison of our data with a recent RNA-seq analysis (Siegel et al., 2014) showed that the detected transcriptional initiation events correlated well with gene expression levels (Figure S4C). This suggests that the majority of these events are associated with the production of full-length transcripts.

In view of the non-random organization of eukaryotic genomes, indicative of the functional importance of gene distribution (Hurst et al., 2004), we first investigated possible interdependency in TSS usage between genes depending on their arrangement in the genome. For the vast majority (91%) of gene pairs in divergent configuration, bidirectional transcription initiated coordinately, suggesting potential co-expression (Figure S4D). However, a number of those (183 out of 2,084; Table S4A) exhibited distinct cycling behavior, as reflected by the contrasting profiles of the associated TSS blocks (Figure S4E). Intriguingly, we also identified divergent gene pairs with distinct expression patterns that comprised TSS blocks with highly similar profiles of temporal usage (Figure 4B;

Figure S4F), suggesting the possible presence of both common and separate promoter elements. Altogether, these results demonstrate that bidirectional promoters may contain multiple regulatory elements that would enable both coordinated and dissociated TSS usage within divergent transcription units. Given that many of the regions of bidirectional transcription were also sites of antisense transcription, we analyzed the temporal usage pattern of TSS blocks for cycling sense/antisense pairs (11% of all sense/antisense pairs, stringent local FDR threshold of 0.01). In our survey, we included sense/antisense pairs on opposite strands in a convergent configuration. We observed dissociated transcription for most. Indeed, only a minority of the antisense transcription initiation events (<1%) occurred in synchrony with their sense counterparts, whereas a few pairs of sense/antisense transcription units (52 out of the ~1,100 pairs surveyed) exhibited expression profiles in opposite phase (Figure 4C; Table S4B). Further investigation will be needed to assess whether these profiles are the result of an inter-dependency between sense and antisense transcription initiation whereby antisense transcription acts as a regulator of gene expression by promoting or preventing expression of the sense transcript (Wei et al., 2011).

Given that most transcription units in *P. falciparum* contain multiple TSS clusters, we additionally examined the dynamics of transcription initiation within each of those. For most of the annotated transcription units, transcription appeared to initiate coordinately at each of the associated TSS blocks, suggesting that, even when transcription starts at multiple sites within a given gene locus, all sites are concomitantly used (Figure 4A). Nonetheless, a subset of *P. falciparum* transcription units (124, 3.4%) exhibited TSS blocks with distinct cycling behavior (Figure 4D; Table S4C). Among these, 57 harbored TSS blocks with temporal usage patterns in opposite phase, indicative of a switch between TSSs in a stage-specific manner and suggesting that initiation events at these sites are mutually exclusive (Table S4D). Northern blot analysis of selected candidates (Figures 4D and S4G) confirmed that transcription initiation at alternative sites does lead to production of full-length transcript isoforms, some of which are possibly non-coding (Figure S4G) and whose expression is developmentally regulated. Interestingly, some of the TSS blocks near the start of the gene coding sequence may generate leader-less transcripts, as was observed for PF3D7_1332500 (Figure 4D). Alternative transcription initiation may lead to the occurrence of alternative translation initiation sites. These, in fact, appear to be widespread and may influence translation efficiency or enrich the pools of proteins (de Klerk and 't Hoen, 2015). Interestingly, initial comparison of our TSS map with the recently published profile of ribosome occupancy during the IDC of *P. falciparum* (Caro et al., 2014) suggests that a dynamic change in TSS usage may be accompanied by a switch between translation initiation sites (Figure S4H). Additionally, it further argues for the likely co-existence of transcript isoforms that are differentially translated (Figure S4I).

## DISCUSSION

This report constitutes a comprehensive survey of the transcription initiation events occurring across the *P. falciparum* asexual blood cycle and provides insights into promoter architecture for a complex and, thus far, poorly characterized genome. Our data, therefore, represent a powerful resource that will enable further investigation into the molecular mechanisms governing the selection of TSSs and more broadly transcriptional regulation in such an extremely base-biased environment.

The in silico survey of DNA physicochemical properties predicted core promoter regions using the partial TSS mapping from Watanabe et al. (2002) and estimated thymine-adenine to be the preferred sequence at the TSS (Brick et al., 2008). We demonstrated that TSS selection at the genome-wide level is organized around such a specific di-nucleotide composition, reflecting the general preference for a pyrimidine-purine initiation site observed in other eukaryotes (Carninci et al., 2006). Given the low-complexity sequence context associated with the *P. falciparum* genome, the question, therefore, arises of what drives the selection of a genomic site for transcription initiation in such an AT-rich environment. We showed that the frequency of usage of a TSS is actually guided by the local G/C content at precise positions downstream of the TSS. This indicates that at least part of the structure of core promoters is genomically encoded and that there might be spatial constraints for the positioning of the basal transcription machinery. Indeed, the local increase in GC content around weaker TSSs suggests the establishment of tighter boundaries for transcription initiation, possibly by way of a more defined nucleosome positioning or particular epigenetic mechanisms such as DNA methylation (Ponts et al., 2013). In fact, our observation that the diminution in TSS strength is accompanied with an increase in the distance between the TSS and the +1 nucleosome, together with a reduced NFR, argues for an association between TSS selection, nucleosome positioning, and regional nucleotide content. Altogether, these observations indicate that, despite the parasite's extraordinary AT-biased genome, promoter selection and activity are primarily defined by local base composition and chromatin structure. Intriguingly, the timing of TSS usage does not correlate with enrichment in the histone marks H3K4me3 and H3K9ac or the variants H2A.Z and H2B.Z at the 5′ end of genes, which are typically associated with promoter activity (Li et al., 2007; Zlatanova and Thakar, 2008). Instead, these seem to be associated with *P. falciparum* maturation, independently of the activation of gene expression, in agreement with previous studies reporting a rather moderate association between the dynamic changes in the histone marks/variants and transcript levels (Bártfai et al., 2010; Hoeijmakers et al., 2013b). Interestingly, the greatest changes in chromatin organization around the TSS occur at the schizont stage and for schizont-specific genes. This could reflect the replication activity that takes place during this developmental stage or a mechanism of synchronization of gene expression to prepare all future daughter cells for the next cycle.

Further examination of the dynamics of TSS usage captured the cascade of gene expression characteristic of the malaria parasite, whereby all TSSs associated with the same transcription unit were coordinately used at a specific time during the parasite's life cycle. For a vast majority of such transcription units, we observed a high recurrence of synchronous transcription initiation events from multiple sites. This may lead to the

simultaneous production of transcript isoforms with alternative 5′ UTRs, as indicated by our northern blot analysis, and thus increased transcriptome diversity (Pal et al., 2011). The concomitant expression of several isoforms with varying 5′ UTRs may have important regulatory consequences (Davuluri et al., 2008), notably by influencing transcript stability or translation efficiency (de Klerk and 't Hoen, 2015). Interestingly, recent ribosome and polysome profiling studies in *P. falciparum* have reported enrichment for ribosomes along the 5′ UTR of numerous transcripts (Bunnik et al., 2013; Caro et al., 2014), including those lacking AUGs (Caro et al., 2014). This suggests the existence of upstream open reading frames (uORFs) that are actively translated from non-cognate initiation codons (Ingolia, 2014) and may affect the efficiency of translation of the full-length transcript. In addition to the numerous transcriptional events detected upstream of the translation start codon, we noticed that most of the transcriptionally active loci contained TSSs that overlap with the start of the coding sequence. Our observations suggest that these events produce, at least in some cases, leader-less transcripts that may be translated (Cortes et al., 2013). The high number of mono- and multi-exonic genes containing exonic promoters also demonstrates the general prevalence of transcription initiation events within gene bodies, as observed elsewhere (Carninci et al., 2006). Conservation of the TA di-nucleotide pattern and GC patches at exonic promoters argue for these internal TSSs being true sites of transcription initiation. Nevertheless, some of the internal TSSs may reflect recapping of degradation products (Lenhard et al., 2012) or may be the result of a moderate level of experimental noise from which other studies, such as that of Tuda et al. (2011) reporting similar observations, may not be immune either.

Analysis of the relative organization of transcription units, to isolate potential regulatory elements and co-regulated loci, led to the identification of numerous TSS positions in intergenic and other non-coding regions. Most of these TSSs could be clustered in hypothetical, previously unreported, non-coding transcription units that may carry a regulatory role, as recently suggested for the telomeric long, non-coding RNAs (lncRNAs) during parasite invasion (Broadbent et al., 2015). Many of these non-coding transcription units correspond to transcription initiation in an antisense orientation to coding genes. More generally, this configuration was widespread across the genome, regardless of the coding potential of the sense/antisense pairs, as previously observed (López-Barragán et al., 2011; Militello et al., 2005; Siegel et al., 2014). Such arrangements that presumably yield overlapping transcripts may carry a regulatory role, as shown by the expression of certain sense-antisense transcript pairs in a mutually exclusive manner. In multiple loci, antisense events originated from bidirectional transcription activity, which we detected for most of the gene pairs in a head-to-head arrangement. For many, we observed coordinate usage of the TSS in both orientations, indicative of the transcription units' co-expression. However, we also identified genomic loci for which divergent transcriptional events were asynchronous or sometimes mutually exclusive, suggesting the possibility of stage-specific regulation of the directionality of bidirectional promoters. Altogether, these observations point toward the prob-

able presence of bidirectional promoters with shared and separate regulatory elements (Trinklein et al., 2004).

In contrast to the majority of genes for which concomitant usage of the associated TSSs was detected, a minority appeared to switch from one TSS to another in a stage-specific manner. This indicates that TSS usage may be temporally controlled, resulting in the generation of developmentally regulated variants. This observation suggests the existence of a context-dependent selection of the TSS, whereby regulatory processes guide the choice for alternative promoters or sequence elements within the same promoter. Interestingly, dynamic TSS usage has been reported in the context of tissue specificity (FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014) and embryonic development (Haberle et al., 2014) and attributed to the differential activity of chromatin-defined enhancers (Andersson et al., 2014) or a switch between TSS selection mechanisms (Haberle et al., 2014). While several studies argue for the existence of *cis*-regulatory sequences in the genome of *P. falciparum* (Horrocks et al., 2009), further investigations will be needed to assess whether these or other promoter/regulatory element switching mechanisms influence this choice. Local changes in nucleosome configuration mediated by distinct nucleotide compositions (Haberle et al., 2014), altered chromatin states (Davuluri et al., 2008), or the binding of regulatory factors such as the ApiAP2 transcription factors to specific regions around the TSS (De Silva et al., 2008) may, for instance, be at play. Given the restricted number of *P. falciparum* genes that display such a behavior of dynamic TSS usage, it will also be interesting to investigate whether this particular mode of transcription regulation is linked to the biological function of these genes. Indeed, the developmentally regulated switch between TSS blocks for this subset of genes may be a mechanism to transcriptionally control the stability or translation efficiency of the associated transcripts, or even their biological function, when it is needed the most. The fact that some of the transcripts emerging from alternate TSS blocks display a distinct temporal profile from that reported for the expression of the corresponding gene would suggest that these transcripts are non-coding RNAs that are possibly non-polyadenylated and, therefore, not detected with classical transcriptomic approaches.

With more than 60% of *P. falciparum* genes for which no biological function has been assigned (Bréhélin et al., 2008), assessing the functional consequence of a developmentally regulated switch between TSSs will require further mechanistic studies.

Our precise mapping of TSSs genome-wide revealed an unexpectedly complex and dynamic transcriptomic landscape and constitutes a major advance toward deciphering the molecular basis of *P. falciparum* transcriptional control. We observed the existence of a TSS signature and postulate that spatial constraints mediated by local base composition and nucleosome occupancy may control the accessibility of the basal transcriptional machinery to promoter regions and, thus, frequency of TSS usage. Therefore, this highly valuable resource opens new avenues into the characterization of regulatory elements and the mechanisms directing TSS selection. The recent implementation of CRISPR/Cas9 approaches to edit the parasite genome (Ghorbal et al., 2014; Wagner et al., 2014) will permit further

examination of the interactions between promoters and transcription factors, as well as whether widespread non-coding transcriptional activity plays a role in regulating gene expression.

## EXPERIMENTAL PROCEDURES

Detailed protocols can be found in the Supplemental Experimental Procedures.

### Parasite Culture
3D7 parasites were synchronized by two consecutive treatments with 5% sorbitol for three or more successive generations before initiating time point samplings every 8 hr throughout the IDC.

### 5′ End Capture and Library Preparation
Single-strand ligation to the 5′ end of transcripts was performed using an RNA adaptor containing a 8-mer molecular barcode. cDNA second-strand synthesis was performed using a biotinylated primer complementary to the RNA 5′ adaptor. All samples were treated in the presence of in vitro transcripts.

### Northern Blot Analysis
Probe templates were labeled using the DIG Northern Starter Kit (Roche) as per manufacturer's instructions. Hybridizations were performed at 55°C.

### TSS Annotation, Antisense, and Bidirectional TSS Detection
To define TSS blocks, morphological operations were performed on the tag counts that were summed across time points and replicates for each position of the genome. The final annotation corresponds to the genomic coordinates of every block associated with the optimal filtering step of each genomic feature. TSS blocks identified on the opposite strand of annotated features were computed within a window of 500 bp around their 3′ end to detect antisense transcription initiation events. Pairs of annotated features in a divergent arrangement that were separated by a distance of 1 kb or less were analyzed for possible bidirectional transcriptional activity.

### Analysis of the Core Promoter Architecture
Chromatin organization around TSSs was analyzed using ChIP-seq datasets from Bártfai et al. (2010) and Hoeijmakers et al. (2013b). Bins of 10 bp within a 2,500-bp window (from −1,000 to +1,500 bp) around the start of the most active TSS block were used to estimate the coverage in the corresponding genomic locations.

The local GC content was computed in 10-bp bins around the start of the block with the highest counts. We assessed potential nucleotide biases at the site of transcription initiation by extracting the nucleotide content in a 5-bp window around the genomic coordinate corresponding to the highest peak of collapsed 5′ tags.

### Analysis of the Dynamics of TSS Usage
A likelihood ratio test was performed to identify TSS blocks for which the usage was different for at least one of the six time points in both biological replicates. TSS blocks with a local FDR < 0.01 were considered active. Transcript level at every single time point was tested against the mean value calculated across all time points. An absolute $\log_2$ fold change $\geq 0.5$ was required to identify peaking time points.

### Data Access
A temporary genome browser is accessible at http://steinmetzlab.embl.de/shiny/TSS_malaria_adjalley_chabbert/ to query for specific genes and different time points until the current integration of the data into PlasmoDB is completed.

## ACCESSION NUMBERS

The accession number for all sequencing raw data reported in this paper is GEO: GSE68982.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.02.025.

## AUTHOR CONTRIBUTIONS

S.H.A., C.D.C., and V.P. conceived the study. S.H.A. performed the experiments, with contributions from C.D.C. and V.P. S.H.A., C.D.C., and B.K. performed the analysis. C.D.C. developed the morphological mathematics approach for the *P. falciparum* TSS data. C.D.C. and B.K. established the statistical pipeline for the dynamic analysis. S.H.A. and C.D.C. interpreted the results, together with V.P. and L.M.S. S.H.A. and C.D.C. wrote the manuscript, with inputs from the other contributors. All authors read and approved the final manuscript.

## REFERENCES

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Balaji, S., Babu, M.M., Iyer, L.M., and Aravind, L. (2005). 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.

Barrandon, C., Spiluttini, B., and Bensaude, O. (2008). Non-coding RNAs regulating the transcriptional machinery. Biol. Cell *100*, 83–95.

Bártfai, R., Hoeijmakers, W.A.M., Salcedo-Amaya, A.M., Smits, A.H., Janssen-Megens, E., Kaan, A., Treeck, M., Gilberger, T.-W., Françoijs, K.-J., and Stunnenberg, H.G. (2010). H2A.Z demarcates intergenic regions of the *plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3. PLoS Pathog. *6*, e1001223.

Bischoff, E., and Vaquero, C. (2010). In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. BMC Genomics *11*, 34.

Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., and DeRisi, J.L. (2003a). Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. Genome Biol. *4*, R9.

Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003b). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol. *1*, E5.

Bréhélin, L., Dufayard, J.-F., and Gascuel, O. (2008). PlasmoDraft: a database of *Plasmodium falciparum* gene function predictions based on postgenomic data. BMC Bioinformatics *9*, 440.

Brick, K., Watanabe, J., and Pizzi, E. (2008). Core promoters are predicted by their distinct physicochemical properties in the genome of *Plasmodium falciparum*. Genome Biol. *9*, R178.

Broadbent, K.M., Park, D., Wolf, A.R., Van Tyne, D., Sims, J.S., Ribacke, U., Volkman, S., Duraisingh, M., Wirth, D., Sabeti, P.C., and Rinn, J.L. (2011). A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. Genome Biol. *12*, R56.

Broadbent, K.M., Broadbent, J.C., Ribacke, U., Wirth, D., Rinn, J.L., and Sabeti, P.C. (2015). Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. BMC Genomics *16*, 454.

Bunnik, E.M., Chung, D.-W.D., Hamilton, M., Ponts, N., Saraf, A., Prudhomme, J., Florens, L., and Le Roch, K.G. (2013). Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. Genome Biol. *14*, R128.

Bunnik, E.M., Polishko, A., Prudhomme, J., Ponts, N., Gill, S.S., Lonardi, S., and Le Roch, K.G. (2014). DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite *Plasmodium falciparum*. BMC Genomics *15*, 347.

Callebaut, I., Prat, K., Meurice, E., Mornon, J.-P., and Tomavo, S. (2005). Prediction of the general transcription factors associated with RNA polymerase II in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. BMC Genomics *6*, 100.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Caro, F., Ahyong, V., Betegon, M., and DeRisi, J.L. (2014). Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. eLife *3*, 04106.

Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., and Young, D.B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. Cell Rep. *5*, 1121–1131.

Coulson, R.M.R., Hall, N., and Ouzounis, C.A. (2004). Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. Genome Res. *14*, 1548–1554.

Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T.H.M. (2008). The functional consequences of alternative promoter use in mammalian genomes. Trends Genet. *24*, 167–177.

de Klerk, E., and 't Hoen, P.A.C. (2015). Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet. *31*, 128–139.

De Silva, E.K., Gehrke, A.R., Olszewski, K., León, I., Chahal, J.S., Bulyk, M.L., and Llinás, M. (2008). Specific DNA-binding by apicomplexan AP2 transcription factors. Proc. Natl. Acad. Sci. USA *105*, 8393–8398.

Epp, C., Li, F., Howitt, C.A., Chookajorn, T., and Deitsch, K.W. (2009). 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.

FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Fournier, C.T., Cherny, J.J., Truncali, K., Robbins-Pianka, A., Lin, M.S., Krizanc, D., and Weir, M.P. (2012). Amino termini of many yeast proteins map to downstream start codons. J. Proteome Res. *11*, 5712–5719.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature *419*, 498–511.

Ghorbal, M., Gorman, M., Macpherson, C.R., Martins, R.M., Scherf, A., and Lopez-Rubio, J.-J. (2014). Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. Nat. Biotechnol. *32*, 819–821.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381–385.

Heijmans, H.J.A.M., and Ronse, C. (1989). The Algebraic Basis of Mathematical Morphology (Centrum voor Wiskunde en Informatica).

Hoeijmakers, W.A.M., Bártfai, R., and Stunnenberg, H.G. (2013a). Transcriptome analysis using RNA-Seq. Methods Mol. Biol. *923*, 221–239.

Hoeijmakers, W.A.M., Salcedo-Amaya, A.M., Smits, A.H., Françoijs, K.-J., Treeck, M., Gilberger, T.-W., Stunnenberg, H.G., and Bártfai, R. (2013b). H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the *Plasmodium falciparum* genome. Mol. Microbiol. *87*, 1061–1073.

Horrocks, P., Wong, E., Russell, K., and Emes, R.D. (2009). Control of gene expression in *Plasmodium falciparum* - ten years on. Mol. Biochem. Parasitol. *164*, 9–25.

Hurst, L.D., Pál, C., and Lercher, M.J. (2004). The evolutionary dynamics of eukaryotic gene order. Nat. Rev. Genet. *5*, 299–310.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Kyes, S., Christodoulou, Z., Pinches, R., Kriek, N., Horrocks, P., and Newbold, C. (2007). *Plasmodium falciparum var* gene expression is developmentally controlled at the level of RNA polymerase II-mediated transcription initiation. Mol. Microbiol. *63*, 1237–1247.

Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J., et al. (2004). Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Res. *14*, 2308–2318.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. *13*, 233–245.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

López-Barragán, M.J., Lemieux, J., Quiñones, M., Williamson, K.C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., and Su, X.-Z. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. BMC Genomics *12*, 587.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

Militello, K.T., Patel, V., Chessler, A.D., Fisher, J.K., Kasper, J.M., Gunasekera, A., and Wirth, D.F. (2005). RNA polymerase II synthesizes antisense RNA in *Plasmodium falciparum*. RNA *11*, 365–370.

Otto, T.D., Wilinski, D., Assefa, S., Keane, T.M., Sarry, L.R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., et al. (2010). New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. Mol. Microbiol. *76*, 12–24.

Oyola, S.O., Otto, T.D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D.J., Macinnis, B., Kwiatkowski, D.P., Swerdlow, H.P., and Quail, M.A. (2012). Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. BMC Genomics *13*, 1.

Pal, S., Gupta, R., Kim, H., Wickramasinghe, P., Baubet, V., Showe, L.C., Dahmane, N., and Davuluri, R.V. (2011). Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. Genome Res. *21*, 1260–1272.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400–1412.

Ponts, N., Fu, L., Harris, E.Y., Zhang, J., Chung, D.-W.D., Cervantes, M.C., Prudhomme, J., Atanasova-Penichon, V., Zehraoui, E., Bunnik, E.M., et al. (2013). Genome-wide mapping of DNA methylation in the human malaria parasite *Plasmodium falciparum*. Cell Host Microbe *14*, 696–706.

Raabe, C.A., Sanchez, C.P., Randau, G., Robeck, T., Skryabin, B.V., Chinni, S.V., Kube, M., Reinhardt, R., Ng, G.H., Manickam, R., et al. (2010). A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*. Nucleic Acids Res. *38*, 608–617.

Schoenberg, D.R., and Maquat, L.E. (2009). Re-capping the message. Trends Biochem. Sci. *34*, 435–442.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc. Natl. Acad. Sci. USA *100*, 15776–15781.

Siegel, T.N., Hon, C.-C., Zhang, Q., Lopez-Rubio, J.-J., Scheidig-Benatar, C., Martins, R.M., Sismeiro, O., Coppée, J.-Y., and Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. BMC Genomics *15*, 150.

Sims, J.S., Militello, K.T., Sims, P.A., Patel, V.P., Kasper, J.M., and Wirth, D.F. (2009). Patterns of gene-specific and total transcriptional activity during the Plasmodium falciparum intraerythrocytic developmental cycle. Eukaryot. Cell *8*, 327–338.

Sorber, K., Dimon, M.T., and DeRisi, J.L. (2011). 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.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome Res. *14*, 62–66.

Tuda, J., Mongan, A.E., Tolba, M.E.M., Imada, M., Yamagishi, J., Xuan, X., Wakaguri, H., Sugano, S., Sugimoto, C., and Suzuki, Y. (2011). Full-parasites: database of full-length cDNAs of apicomplexa parasites, 2010 update. Nucleic Acids Res. *39*, D625–D631.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*. Nat. Methods *11*, 915–918.

Watanabe, J., Sasaki, M., Suzuki, Y., and Sugano, S. (2002). Analysis of transcriptomes of human malaria parasite *Plasmodium falciparum* using full-length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs. Gene *291*, 105–113.

Wei, W., Pelechano, V., Järvelin, A.I., and Steinmetz, L.M. (2011). Functional consequences of bidirectional promoters. Trends Genet. *27*, 267–276.

World Health Organization (2015). World malaria report 2015. http://www.who.int/malaria/publications/world-malaria-report-2015/en/.

Zlatanova, J., and Thakar, A. (2008). H2A.Z: view from the top. Structure *16*, 166–179.