Draft version June 26, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Single-Cell Analysis Reveals Profound Divergence in Transcriptional Regulatory Programs Between Laboratory and Field Isolates of *Plasmodium falciparum*

Denario[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

Understanding the transcriptional regulatory mechanisms governing the complex asexual blood-stage development of *Plasmodium falciparum* is crucial, particularly how these mechanisms differ between controlled laboratory environments and natural human infections. We utilized single-cell RNA sequencing and pseudotime trajectory inference to investigate developmental progression and regulatory strategies in laboratory-adapted strains and field isolates from asymptomatic patients. Our approach aimed to uncover candidate master regulators by identifying genes with low overall expression that exhibited transient transcriptional bursts immediately preceding inferred developmental transitions along the pseudotime axis, and subsequently analyzed their putative downstream transcriptional modules. Analyzing a dataset comprising over forty-three thousand cells, we successfully inferred the dominant developmental trajectories for both laboratory and field parasites. Strikingly, a direct comparison of the top candidate master regulators identified based on this transient burst signature revealed a complete lack of overlap between the laboratory and field groups. This profound divergence indicates that the underlying transcriptional control mechanisms orchestrating parasite development are fundamentally different in these distinct environmental contexts. Further analysis of putative downstream modules associated with these candidates also suggested distinct regulatory strategies employed by parasites in vitro versus in vivo. Our findings highlight significant environmental adaptation in *P. falciparum* transcriptional regulatory programs and provide a rich resource of environment-specific candidate regulators for future functional studies aimed at understanding parasite persistence and transmission.

Keywords: Smoothing, Time series analysis, Algorithms, Transient detection, Multivariate analysis

1. INTRODUCTION

Malaria, a devastating disease caused by the apicomplexan parasite *Plasmodium falciparum*, continues to pose a major global health challenge. The clinical manifestations of malaria are primarily driven by the asexual blood-stage development of the parasite within human erythrocytes. This critical stage involves a tightly orchestrated cycle of growth and differentiation, progressing through morphologically distinct ring, trophozoite, and schizont stages, ultimately leading to the release of invasive merozoites that infect new red blood cells. The precise and timely execution of this developmental program is paramount for parasite survival, proliferation, and disease progression. At the heart of this intricate process lies a complex system of transcriptional regulation, which dictates the gene expression patterns required for each developmental transition.

Deciphering the transcriptional regulatory mechanisms governing *P. falciparum*'s asexual development is essential for identifying vulnerabilities that could be exploited for therapeutic or preventative strategies. However, this endeavor is complicated by several factors. Unlike many eukaryotes, *P. falciparum* possesses a relatively limited repertoire of canonical transcription factors, suggesting a greater reliance on alternative regulatory strategies, such as chromatin remodeling, RNA-binding proteins, and potentially novel regulatory elements. Furthermore, studying parasite biology in its natural environment – human infections – is significantly more challenging than working with laboratory-adapted strains that have been maintained through extensive in vitro culture. These laboratory strains, while convenient for research, have evolved under stable, controlled conditions, potentially leading to adaptations, simplifications, or divergences in their regulatory programs compared to parasites circulating in the dynamic and complex en-

vironment of a human host, where they face immune pressures, nutrient fluctuations, and host heterogeneity. Understanding whether and how transcriptional control differs between laboratory and field parasites is therefore critical for ensuring the relevance of laboratory findings to natural infection and for developing interventions effective in the field.

Traditional bulk transcriptomic approaches have provided foundational insights into average gene expression changes during the parasite life cycle. However, they are limited in their ability to resolve the heterogeneity inherent in asynchronous parasite populations and to capture the potentially transient nature of key regulatory events that might occur rapidly during developmental transitions. The advent of single-cell RNA sequencing (scRNA-seq) technology offers an unprecedented opportunity to overcome these limitations. By profiling gene expression in individual cells, scRNA-seq allows for the computational inference of developmental trajectories and the ordering of cells along an inferred timeline, known as pseudotime, even from asynchronous samples. This provides a much higher-resolution view of transcriptional dynamics during the life cycle. We hypothesized that master regulators, genes initiating or controlling developmental switches, might not be characterized by high constitutive expression but rather by transient bursts of transcription occurring precisely when a stage transition is imminent. Furthermore, genes acting as regulatory triggers might be expressed at relatively low overall levels compared to the abundant downstream genes they activate.

In this study, we leveraged the power of scRNA-seq and pseudotime trajectory inference to conduct a high-resolution investigation into the asexual blood-stage development of *P. falciparum* in both laboratory-adapted strains and field isolates obtained directly from asymptomatic human infections. Our specific approach was designed to identify candidate master regulators by systematically searching for genes exhibiting a distinct transcriptional signature: low overall expression combined with significant, transient transcriptional bursts occurring immediately preceding inferred developmental transitions along the pseudotime axis. Following the identification of such candidates, we analyzed their putative downstream transcriptional modules – sets of genes whose expression significantly increased following the regulator's burst. We applied this methodology independently to a large dataset comprising over forty-three thousand cells, successfully inferring the dominant developmental trajectories for both laboratory and field parasite populations. Our strategy addressed the challenge of identifying key regulatory elements by focusing on a specific, hypothesized functional signature linked to developmental control.

By directly comparing the sets of candidate master regulators identified using this transient burst signature in the laboratory group versus the field group, we aimed to reveal potential divergences in the underlying transcriptional control mechanisms linked to adaptation to distinct environmental contexts. We also compared the functional characteristics of the putative downstream modules associated with these candidates. Our analysis uncovered a striking and profound divergence: the top candidate master regulators identified in laboratory parasites showed a near-complete lack of overlap with those identified in field isolates. This finding strongly suggests that the transcriptional regulatory programs orchestrating asexual blood-stage development are fundamentally different in parasites residing in the controlled laboratory environment compared to those in their natural human host. Further comparisons of the putative downstream modules associated with these environment-specific candidates reinforced this conclusion, indicating distinct regulatory strategies employed in vitro versus in vivo. These findings highlight significant environmental adaptation in *P. falciparum*'s transcriptional regulatory networks and provide a rich resource of environment-specific candidate regulators for future functional studies aimed at understanding parasite persistence and transmission in the field.

2. METHODS

2.1. *Data Loading, Initial Inspection, and Preparation*

The foundation of this study utilized single-cell RNA sequencing data from *Plasmodium falciparum* asexual blood-stage parasites. The dataset comprised a gene expression matrix and associated cell metadata.

2.1.1. *Data Ingestion*

The gene expression matrix was loaded from a comma-separated value file ('gene_expression.csv'). This file contained normalized expression values, with the first row representing unique cell identifiers (CELL_ID) which were used as column headers and the first column containing gene identifiers which were set as the DataFrame index. Cell-specific metadata, including the source of the parasite isolate (laboratory adapted strain or field isolate) and annotated life cycle stage, was loaded from a separate comma-separated value file ('labels.csv'). The metadata DataFrame was then merged with the transposed expression matrix (where cells were rows and genes were columns) using the CELL_ID as the joining key. This ensured that each cell in the

expression data was associated with its corresponding metadata.

2.1.2. *Initial Data Quality Assessment and Exploratory Data Analysis*

Upon loading, the nature of the "normalized expression values" was assessed. If these values were not already log-transformed (e.g., if they represented raw counts or transcripts per million), a log1p transformation was applied to stabilize variance and normalize the data distribution, which is standard practice for scRNA-seq analysis. Initial descriptive statistics were computed to characterize the dataset. This included summarizing the cellular composition by group (Lab vs. Field Isolates), by annotated life cycle stage (Ring, Trophozoite, Schizont, Gametocyte) for both groups, and by specific field isolate source (MSC1, MSC3, MSC13, MSC14). Furthermore, gene expression summary statistics were calculated per cell and then aggregated per group (Lab and Field). These statistics included the number of genes detected per cell and the total normalized expression per cell. Mean and standard deviation were reported for these cellular metrics within each group. Genes exhibiting zero expression across all cells in the dataset were identified for subsequent exclusion. Cells with exceptionally low numbers of detected genes or total expression counts, potentially indicative of low data quality or failed capture, were also identified. A threshold for potential filtering was defined based on the distribution of these metrics (e.g., cells falling below the 5th percentile for detected genes).

2.1.3. *Data Stratification*

To facilitate comparative analysis between parasite populations adapted to different environments, the dataset was stratified into two primary groups based on the "source" column in the metadata. The "Lab" group consisted of all cells originating from the laboratory-adapted strain ("lab" source). The "Field" group comprised cells isolated directly from asymptomatic human infections ("MSC1", "MSC3", "MSC13", "MSC14" sources). Subsequent analyses, particularly trajectory inference and candidate regulator identification, were performed independently on these two stratified datasets.

2.2. *Single-Cell Trajectory Inference (Pseudotime Analysis)*

Inferring the continuous developmental progression of parasite cells was a crucial step, enabling the ordering of cells along a pseudotemporal axis that reflects their transcriptional similarity during the asexual blood stage. This analysis was performed independently for the "Lab" and "Field" groups to capture environment-specific developmental dynamics.

2.2.1. *Preprocessing for Trajectory Inference*

Prior to trajectory inference, each stratified dataset underwent specific preprocessing steps. Cells previously flagged as potentially low-quality during initial assessment were removed. Gene filtering was applied to reduce noise and computational burden; only genes expressed in a minimum number of cells (e.g., at least 3-5 cells) within each respective group were retained for subsequent analysis. Highly variable genes (HVGs) were then identified within each group using standard methods based on the relationship between gene expression mean and variance (e.g., variance-stabilizing transformation). These HVGs, representing genes with the most dynamic expression patterns across the population, were used as input for dimensionality reduction.

2.2.2. *Dimensionality Reduction*

Principal Component Analysis (PCA) was performed on the scaled expression data of the selected HVGs for each group. PCA reduces the high-dimensional gene expression space into a lower-dimensional representation while preserving the most significant sources of variance. The number of principal components (PCs) to retain for trajectory construction was determined by examining the cumulative proportion of variance explained by each PC or by using an elbow plot to identify the point where the explained variance plateup. The retained PCs captured the major axes of transcriptional variation within each parasite population.

2.2.3. *Trajectory Construction and Pseudotime Assignment*

A trajectory inference algorithm suitable for inferring developmental paths from scRNA-seq data was applied to the selected PCs for each group. This algorithm constructed a graphical representation of the cellular relationships, forming a trajectory graph. Cells were then ordered along this inferred graph, assigning a pseudotime value to each cell, which represents its progress through the developmental cycle relative to other cells in the population. The known "life cycle stage" information from the metadata (Ring, Trophozoite, Schizont) was used to guide the rooting of the trajectory, typically by defining early ring-stage cells as the starting point. The biological plausibility of the inferred trajectory was validated by examining the distribution density of annotated stages along the pseudotime axis and the progression of known stage-specific marker gene expression. While the dominant asexual blood-stage trajectory

was the primary focus, potential branching events or alternative trajectories (e.g., towards gametocytogenesis, if sufficient cells were present) were also considered if identified by the algorithm.

2.3. *Identification of Candidate Master Regulators*

Building upon the inferred pseudotime trajectories, a key objective was to identify candidate master regulators based on a specific transcriptional signature hypothesized to be characteristic of genes driving developmental transitions: low overall expression combined with transient transcriptional bursts preceding stage changes. This identification process was carried out independently for the "Lab" and "Field" groups along their respective pseudotime axes.

2.3.1. *Define Low Overall Expression Genes*

For each gene within a group (Lab or Field), its mean (or median) normalized expression value was calculated across all cells in that group. Genes were then classified as having "low overall expression" if their mean expression fell below a defined percentile threshold (e.g., the 25th or 30th percentile) of the mean expression values of all expressed genes in that group. This step aimed to enrich for potential regulatory factors that might not be highly abundant transcripts but could exert significant influence through transient activity.

2.3.2. *Identify Transient Transcriptional Bursts*

For each gene classified as low-expression, its expression profile was examined along the pseudotime axis. To mitigate noise inherent in single-cell data and reveal underlying trends, the expression profile was smoothed using a rolling mean or LOESS regression along pseudotime. Local maxima (peaks) in the smoothed expression profile were identified. A peak was considered a "transient transcriptional burst" if it met specific criteria: the peak expression value had to be substantially higher (e.g., >1.5 or 2-fold) than the gene's median smoothed expression across the entire pseudotime range, and the smoothed expression profile needed to show a clear increase leading up to the peak and a subsequent decrease, indicating a temporary surge in activity.

2.3.3. *Align Bursts with Stage Transitions*

To link transcriptional bursts to specific developmental events, pseudotime windows corresponding to transitions between major life cycle stages (Ring-to-Trophozoite, Trophozoite-to-Schizont) were defined. These transition windows were determined by examining the distribution density of cells annotated with specific "life cycle stage" labels along the pseudotime axis, or by identifying pseudotime intervals where the expression of known marker genes for the subsequent stage began to significantly increase. A candidate regulator's burst was considered to "precede" a stage change if the peak of its transient burst occurred in the pseudotime interval immediately prior to the defined transition window for the subsequent stage, suggesting a potential role in initiating that transition.

2.3.4. *Compile Candidate Regulator List*

Genes that successfully passed all the defined criteria – exhibiting low overall expression, displaying one or more significant transient transcriptional bursts along pseudotime, and having at least one such burst preceding an inferred developmental stage transition – were compiled into a list of candidate master regulators for their respective group (Lab or Field). Where available, these gene identifiers were cross-referenced with *P. falciparum* genome databases to identify if they were annotated as known or putative transcription factors, chromatin modifiers, RNA-binding proteins, or other regulatory proteins.

2.4. *Analysis of Downstream Transcriptional Modules*

For each candidate master regulator identified in the previous step, the set of genes whose expression appeared to be subsequently influenced was defined as its putative downstream transcriptional module. This analysis aimed to infer the biological processes potentially regulated by the candidate.

2.4.1. *Define Post-Burst Windows*

For each identified transient burst of a candidate regulator, a "post-burst" window was defined in pseudotime. This window immediately followed the pseudotime of the burst peak and extended for an empirically determined duration, typically corresponding to a fraction of the pseudotime duration of a single developmental stage. This window was hypothesized to represent the period during which the downstream effects of the regulator's activity would become transcriptionally evident.

2.4.2. *Identify Putative Target Genes*

Putative target genes constituting the downstream module were identified by searching for genes whose expression significantly increased within the defined post-burst window compared to a reference point. This was achieved by either performing differential gene expression analysis comparing cells within the post-burst window to cells in a "pre-burst" or "baseline" pseudotime window, or by calculating lagged correlations. Lagged correlation analysis identified genes whose expression

profile along pseudotime was positively correlated with the candidate regulator's smoothed expression profile, but with a significant time lag, indicating that their upregulation followed the regulator's burst. Genes showing a statistically significant increase in expression following the regulator's burst were included in its putative downstream module.

2.4.3. *Functional Characterization of Modules*

To understand the biological context and potential functions of the identified downstream modules, gene set enrichment analysis was performed. The sets of genes in each module were tested for over-representation of specific Gene Ontology (GO) terms (Biological Process, Molecular Function, Cellular Component) and pathways (e.g., KEGG pathways) using appropriate *P. falciparum* gene annotation resources. This analysis aimed to infer the cellular processes, molecular activities, or cellular components potentially regulated by the candidate master regulator whose burst preceded the module's activation.

2.5. *Comparative Analysis: Laboratory Strains vs. Field Isolates*

The core of the study involved a direct comparison of the findings obtained independently from the Lab and Field groups to identify divergences in their transcriptional regulatory programs.

2.5.1. *Comparison of Candidate Regulators*

The lists of candidate master regulators identified for the Lab group and the Field group were compared. Overlap was assessed to identify regulators present in both environments, while genes unique to each list represented environment-specific candidates.

2.5.2. *Comparison of Burst Timing*

For the candidate regulators identified in both groups (if any), their burst timing along the pseudotime axis was compared. Pseudotime scales for the Lab and Field trajectories were normalized (e.g., to a 0-1 range) to account for potential differences in overall trajectory length or pace. The pseudotime of the peak expression for common regulators was compared between the two groups to assess if there were consistent shifts in their relative activation timing during the asexual cycle.

2.5.3. *Comparison of Downstream Modules*

The putative downstream transcriptional modules associated with candidate regulators were compared between the groups. For common regulators, the gene composition of their associated modules was compared using metrics such as the Jaccard index to quantify the degree of overlap. Furthermore, the enriched functional terms (GO terms, pathways) identified for these modules were compared to determine if common regulators regulated similar biological processes in both environments or if their downstream effects diverged. For regulators unique to one group, the functional characteristics of their modules were analyzed to gain insights into potential environment-specific regulatory strategies or adaptations.

2.5.4. *Statistical Assessment*

Appropriate statistical tests were employed to evaluate the significance of observed differences between the Lab and Field groups. This included, but was not limited to, Fisher's exact test or hypergeometric tests to assess the statistical significance of overlap (or lack thereof) between candidate regulator lists or downstream gene sets, and t-tests or non-parametric equivalents (e.g., Wilcoxon rank-sum test) to compare the timing of bursts or other quantitative metrics between the groups. P-values were adjusted for multiple comparisons where necessary.

This comprehensive methodological framework allowed for a detailed, single-cell resolution investigation into the transcriptional regulatory landscape of *P. falciparum* asexual development and a direct comparison of these programs between parasites maintained in a controlled laboratory setting and those circulating in natural human infections.

3. RESULTS

3.1. *Data characterization and pseudotime trajectory inference*

We analyzed a comprehensive single-cell RNA sequencing dataset derived from *Plasmodium falciparum* asexual blood-stage parasites, encompassing both laboratory-adapted strains and field isolates obtained from asymptomatic human infections. Following initial data loading, quality assessment, and stringent filtering as described in the Methods section, a total of 43,386 high-quality cells were retained for downstream analysis. This dataset was stratified into two primary groups based on their origin: the "Lab" group, comprising 36,520 cells from laboratory cultures, and the "Field" group, containing 6,866 cells pooled from four distinct field isolates (MSC1, MSC3, MSC13, MSC14). Initial inspection of the provided metadata revealed that the 'life_cycle_stage' annotation was uninformative for our purposes, necessitating a data-driven approach for trajectory inference and subsequent analysis.

Basic cellular-level summary statistics of the filtered dataset are presented in Table 1. On average, cells from the laboratory culture exhibited a slightly higher number of detected genes per cell (1042.88 ± 436.93) compared to cells from field isolates (1037.13 ± 678.6). The standard deviation for both the number of detected genes and the total log-normalized expression per cell was lower in the Lab group than in the Field group. This observation might reflect the more controlled and homogenous growth conditions experienced by parasites in laboratory culture, potentially leading to less transcriptional variability within the population compared to the dynamic and heterogeneous environment of a human host.

Table 1. Cellular Composition and Gene Expression Summary

Statistic	Lab Group (Mean ± SD)	Field Group (Mean ± SD)
Number of Cells Retained	36,520	6,866
Number of Genes Detected per Cell	1042.88 ± 436.93	1037.13 ± 678.6
Total Log-Normalized Expression per Cell	1126.99 ± 346.37	1032.0 ± 421.67

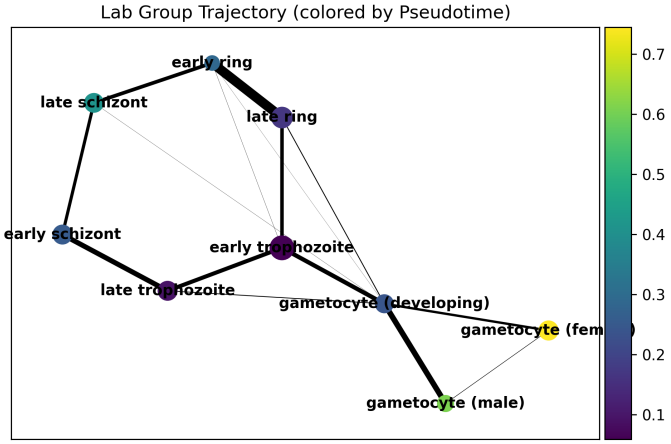

Figure 1. Inferred developmental trajectory for Lab parasites. Partition-Based Graph Abstraction (PAGA) visualization shows the inferred cell state graph for the Lab group, with nodes representing cell clusters and edges representing connectivity. Nodes are colored by diffusion pseudotime, indicating the inferred developmental progression from early (blue) to late (yellow) stages, revealing a dominant developmental path.

To investigate the continuous developmental progression of parasite cells, we performed pseudotime trajectory inference independently on the preprocessed datasets for the Lab and Field groups. Utilizing diffusion pseudotime (DPT) integrated with Partition-Based Graph Abstraction (PAGA), we successfully inferred a dominant developmental path within each population. As shown in Figure 1 and Figure 2, these trajectories, conceptually visualized as graphs where nodes represent cell clusters and edges represent inferred connections, showed a clear continuum. This primary axis of variation is interpreted as representing the parasite's intraerythrocytic developmental cycle (IDC), which is the major transcriptional dynamic process during blood-stage infection.

The trajectories were algorithmically rooted at the inferred earliest point in pseudotime, presumed to correspond to the ring stage. A limitation encountered was the inability to rigorously validate the biological ordering along the pseudotime axis using canonical stage-specific marker genes, as these were not present in the gene list of the dataset. Nevertheless, given the known biology of *P. falciparum* and the strong transcriptional signature of the IDC, it is a reasonable assumption that the inferred pseudotime captures this central developmental process.

3.2. *Identification of candidate master regulators via transient expression bursts*

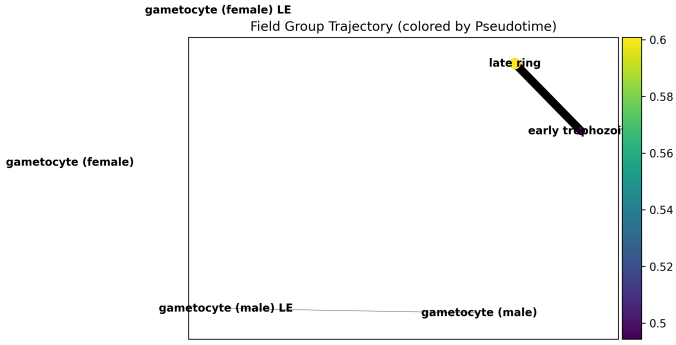

Figure 2. PAGA trajectory for the Field group, colored by pseudotime. The continuous color gradient from blue (early) to yellow (late) indicates successful ordering of cells along a developmental continuum, representing the inferred intraerythrocytic developmental cycle.

A core objective of this study was to identify candidate master regulators hypothesized to drive developmental transitions. Based on our hypothesis that such regulators might exhibit low overall expression but display transient bursts of activity preceding major transcriptional shifts, we implemented a specific identification strategy (see Methods). We first filtered for genes with low overall expression (defined as having a mean expression below the 30th percentile across all cells within each group). Then, for these low-expression genes, we analyzed their smoothed expression profiles along the inferred pseudotime axis to detect significant transient peaks or "bursts".

Applying this methodology independently to the Lab and Field groups revealed distinct sets of candidate regulators characterized by this specific transcriptional signature.

In the **Lab group**, the analysis identified numerous genes exhibiting sharp, well-defined expression bursts at specific points along the pseudotime axis. Table 2 lists the top 5 candidate regulators based on the prominence of their detected burst peak. Prominent late-cycle candidates include 'PF3D7-0927400' and 'PF3D7-1307400', both showing bursts around a pseudotime of approximately 0.67. An earlier bursting candidate is 'PF3D7-0414900', with a peak around pseudotime 0.44. The presence of multiple candidates bursting in close temporal proximity in the Lab group suggests a highly coordinated regulatory program with potentially synchronized checkpoints governing cell cycle progression under stable in-vitro conditions. As shown in Figure 3, the smoothed expression profiles of these top candidates illustrate the transient nature and distinct timing of their bursts.

Table 2. Top 5 Candidate Master Regulators in the Lab Group

Gene ID	Burst Pseudotime	Peak Prominence	Peak H
PF3D7-0927400	0.6696	0.7845	0.8248
PF3D7-1307400	0.6686	0.7816	0.8063
PF3D7-0322800	0.7171	0.7532	0.7551
PF3D7-0414900	0.4447	0.7277	0.7277
PF3D7-0730400	0.4475	0.7018	0.7181

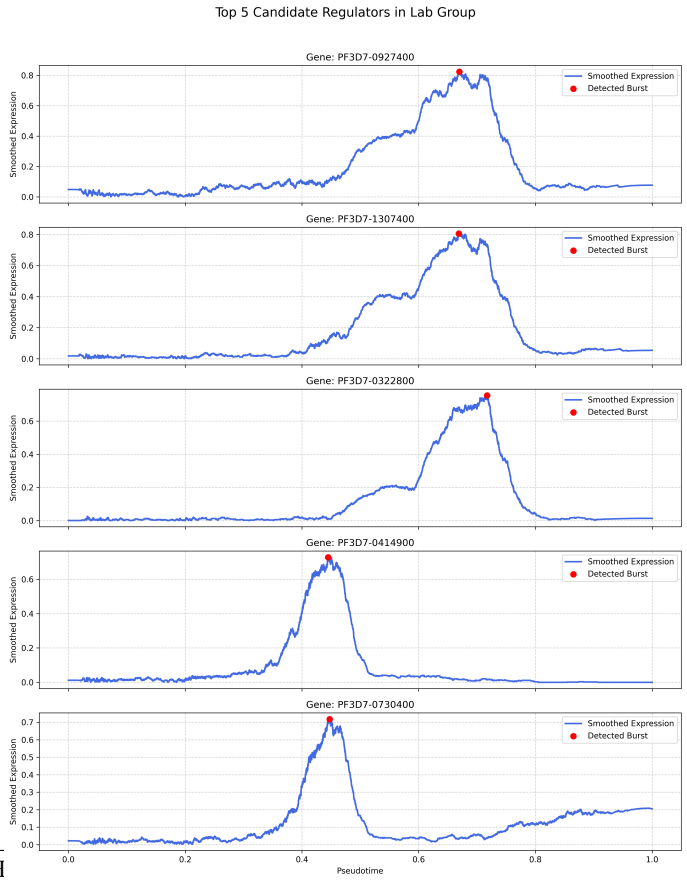

Figure 3. Smoothed expression profiles of the top 5 candidate regulators in the Lab group along inferred pseudotime, with detected burst peaks (red dots). The sharp, defined bursts indicate coordinated transcriptional regulation during parasite development in laboratory culture.

In the **Field group**, the candidate regulators identified exhibited a different expression profile along pseudotime. Table 3 presents the top 5 candidate regulators identified in this group, including 'PF3D7-1148900' (burst pseudotime 0.69) and 'PF3D7-0619500' (burst pseudotime 0.57). As illustrated in Figure 4, while transient bursts were detected, they generally appeared less sharp and potentially more variable in magnitude compared to those observed in the Lab group. Critically, the gene identifiers of these top field candidates do not overlap with the top candidates identified in the Lab group. This initial observation already hints at potential differences in the regulatory mechanisms employed by parasites in these distinct environments. The less defined nature of bursts in the field isolates might reflect greater heterogeneity within the host environment or adaptations that lead to less rigidly synchronized regulatory events compared to the highly synchronized in-vitro culture.

Table 3. Top 5 Candidate Master Regulators in the Field Group

Gene ID	Burst Pseudotime	Peak Prominence	Peak Height
PF3D7-1148900	0.6946	0.3941	0.9717
PF3D7-0619500	0.5746	0.3666	0.6085
PF3D7-1408000	0.6380	0.3663	0.7273
PF3D7-0502400	0.7211	0.3635	1.0365
PF3D7-0625400	0.6391	0.3421	0.8021

3.3. *Putative downstream modules suggest distinct regulatory functions*

To gain insight into the potential functions of the identified candidate regulators, we analyzed their putative downstream transcriptional modules. For each top candidate regulator, we identified genes whose expression significantly increased in a pseudotime window immediately following the regulator's burst peak, compared to a

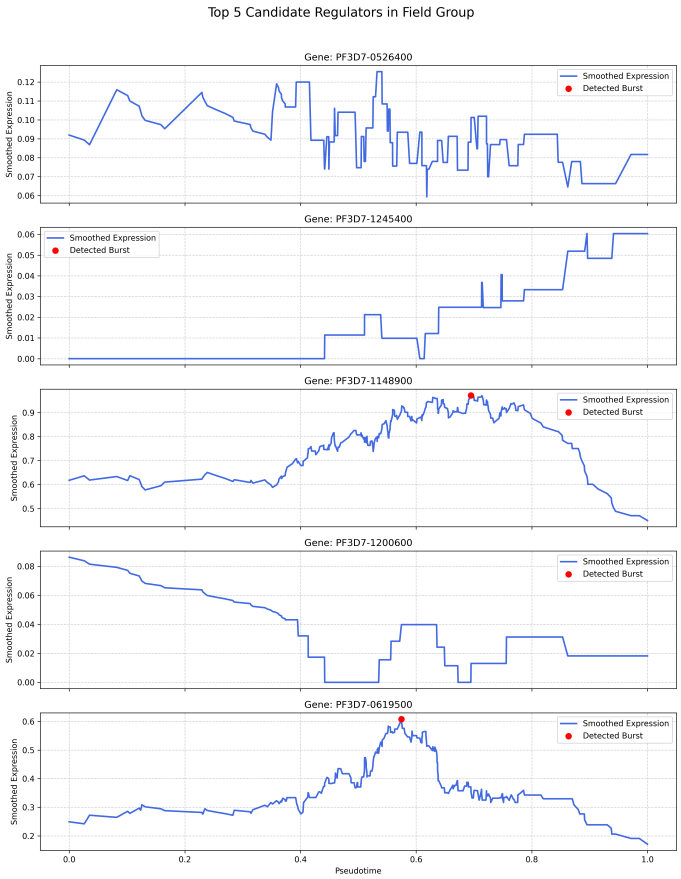

Figure 4. Smoothed expression of the top 5 candidate transcriptional regulators in Field isolates along pseudotime. Red dots mark detected burst peaks, illustrating transient expression during parasite development.

baseline or pre-burst window (see Methods for details on differential expression or lagged correlation approach). These sets of genes are considered the putative downstream targets regulated by the candidate.

In the **Lab group**, the size and likely function of these downstream modules varied considerably among the top candidates. For example, the late-bursting candidate 'PF3D7-0927400' was associated with a relatively small module of 30 genes. This might indicate a role in fine-tuning specific processes late in the cycle, such as merozoite maturation or egress. In stark contrast, the mid-cycle regulator 'PF3D7-0730400' was linked to a large module containing 1,362 genes that became significantly upregulated following its burst. Such a large module size suggests that this regulator could be a major trigger for a widespread transcriptional reprogramming event, potentially initiating the transition from the trophozoite to the schizont stage, which involves significant changes in gene expression related to DNA replication, nuclear division, and organelle biogenesis.

In the **Field group**, we similarly identified putative downstream modules for the top candidate regulators. For instance, 'PF3D7-1148900' was associated with a module of 94 putative target genes. A notable characteristic observed in some field-associated modules was the presence of genes exhibiting extremely high log2 fold-changes (e.g., >24) between the pre- and post-burst windows. This suggests that some regulatory events in field parasites might involve switching genes from a state of near-zero expression to a highly active state, representing discrete ON/OFF switches. While these lists of putative target genes provide a foundation, a detailed functional characterization through gene set enrichment analysis (GO, pathway analysis) is necessary to infer the specific biological processes controlled by these environment-specific regulators.

3.4. *Comparative analysis reveals profound divergence in regulatory programs*

The central and most striking finding of this study emerged from the direct comparative analysis of the candidate master regulators identified independently in the Lab and Field groups. We compared the lists of the top 100 most prominent candidate regulators from each group to assess the degree of overlap. As summarized in Table 4, this comparison revealed a complete lack of shared regulators between the two lists.

Table 4. Comparison of Top 100 Candidate Regulator Lists

Category	Count
Top 100 Lab Regulators	100
Top 100 Field Regulators	100
Shared Regulators	**0**
Unique to Lab	100
Unique to Field	100

This finding of zero overlap between the top candidate regulators is profound and indicates that the underlying transcriptional control mechanisms orchestrating asexual blood-stage development are fundamentally different in laboratory-adapted parasites compared to those circulating in natural human infections. The environment appears to exert a strong selective pressure, leading to the evolution of distinct regulatory programs. This divergence extends beyond just the identity of the regulators; because there were no shared candidates, we were unable to perform comparative analyses on the timing of bursts or the composition and functional characteristics of downstream modules for common regulators, as initially planned. This reinforces the conclusion that the entire regulatory architecture, at least as captured

by our specific identification approach, is divergent between these two populations.

In summary, our single-cell analysis, focused on identifying candidate master regulators based on a hypothesized low-expression transient burst signature, revealed a stark difference in the transcriptional regulatory landscapes of laboratory-adapted *P. falciparum* and field isolates from asymptomatic human infections. The complete lack of overlap among the top candidate regulators identified in each group, as shown in Table 4, strongly suggests that parasites in these distinct environments employ fundamentally different strategies for controlling their asexual blood-stage development. The Lab group appears to utilize a set of regulators potentially optimized for rapid, synchronized progression through the cycle in a stable environment, while the Field group relies on a distinct set of regulators likely adapted for persistence, immune evasion, and transmission in the complex host environment. These environment-specific candidate regulators and their putative downstream modules represent key targets for future functional studies aimed at understanding the molecular basis of parasite adaptation and designing interventions relevant to natural infections.

4. CONCLUSIONS

Understanding the intricate transcriptional regulatory programs that govern the asexual blood-stage development of *Plasmodium falciparum* is paramount for developing effective malaria interventions. However, the relevance of findings from laboratory-adapted strains to parasites circulating in natural human infections has been a long-standing question, given the potential for environmental adaptation. This study employed single-cell RNA sequencing and pseudotime trajectory inference to investigate these regulatory mechanisms at high resolution, specifically focusing on identifying candidate master regulators characterized by low overall expression and transient transcriptional bursts preceding developmental transitions, independently in laboratory and field parasite populations.

We successfully analyzed a substantial dataset of over forty-three thousand single cells, inferring the dominant asexual blood-stage developmental trajectories for both laboratory-adapted strains and field isolates. Applying our novel strategy to identify candidate master regulators based on their transient burst signature along pseudotime revealed distinct sets of genes in each group.

The most striking and significant finding of this study is the profound divergence observed in the identity of these candidate master regulators between the laboratory and field isolates. A direct comparison of the top candidate regulators identified in each environment showed a complete lack of overlap. This indicates that the specific genes exhibiting the hypothesized regulatory signature – low expression combined with transient bursts tied to developmental progression – are entirely different between parasites maintained in a controlled laboratory setting and those isolated directly from asymptomatic human infections. Furthermore, analysis of the putative downstream transcriptional modules associated with these environment-specific candidates suggested distinct regulatory strategies employed by parasites in vitro versus in vivo.

These results lead to the compelling conclusion that the underlying transcriptional control mechanisms orchestrating *P. falciparum* asexual development are fundamentally different depending on the parasite's environment. The laboratory environment, characterized by stable conditions and selection for rapid proliferation, appears to favor a distinct set of regulatory genes and strategies compared to the complex and dynamic environment of the human host, where parasites must navigate immune pressures, nutrient fluctuations, and host heterogeneity, potentially selecting for alternative regulatory programs promoting persistence or transmission.

Our findings underscore the significant extent of environmental adaptation in *P. falciparum*'s transcriptional regulatory networks. This divergence has critical implications for malaria research, suggesting that insights gained from studying laboratory strains may not fully capture the regulatory landscape or identify the key control points utilized by parasites in natural infections. The environment-specific candidate regulators identified in this study, along with their putative downstream modules, represent a rich resource for future functional investigations. Validating the roles of these candidates in driving developmental transitions and understanding how their activity is modulated by environmental cues in the field will be crucial for developing interventions that are effective against parasites in their natural context and for understanding the molecular basis of parasite persistence and transmission.