

# ***Skeptical review: Comparative Single-Cell Transcriptomics Reveals Divergent Stage Transition Dynamics and Regulatory Strategies in Lab-Adapted and Field Isolates of Plasmodium falciparum\*\****

---

## **Summary**

The manuscript presents a comparative single-cell RNA-seq analysis of 45,1691 *Plasmodium falciparum* blood-stage parasites from continuously cultured laboratory strains and field isolates collected from asymptomatic donors. Using Scanpy-based preprocessing and visualization (HVG selection, PCA, UMAP), graph abstraction/trajectory inference (PAGA) and diffusion pseudotime (DPT), the authors reconstruct developmental progressions separately for lab and field parasites (Sec. 2.1–2.4, Sec. 3.1–3.3). They identify pseudotime-dynamic genes, cluster them into co-expression modules, and propose a cross-correlation/lag framework to nominate low-abundance, sharply peaking candidate regulators whose expression precedes module activation (Sec. 2.5, Sec. 3.4–3.5). The main biological claim is that lab parasites exhibit short-lag “just-in-time” regulation whereas field parasites exhibit long-lag “priming,” particularly for gametocyte-related programs (Sec. 3.5.1–3.5.3, Sec. 4). The question (lab adaptation vs in-host biology) is important and the dataset is valuable, but the current draft has substantial barriers to reproducibility (corrupted/duplicated Methods text in Sec. 2.1–2.3 and Sec. 2.5), and the central timing comparison is heavily confounded by major differences in stage composition and trajectory coverage between lab and field datasets (Sec. 3.1–3.3). Stronger controls for batch/technical effects, apples-to-apples comparisons on matched biological lineages, and formal robustness/statistical validation are needed before the “just-in-time vs priming” conclusion can be considered well-supported.

## **Strengths**

- Addresses an important and timely biological question: how laboratory adaptation may distort transcriptional programs relative to circulating field parasites, with implications for transmission biology and interpretation of in vitro studies (Sec. 1, Sec. 4).
- Large single-cell dataset (45,1691 cells; 5,1274 genes after QC as stated) enabling trajectory- and module-level analyses rather than purely descriptive snapshots (Sec. 2.1–2.2, Sec. 3.1).
- Appropriate use of established single-cell tooling (Scanpy; PCA/UMAP; PAGA; DPT) to summarize and interpret developmental structure (Sec. 2.3–2.4, Sec. 3.2–3.3).
- Moves beyond atlas-building by attempting to relate candidate regulators to downstream co-expression modules via timing relationships, which—if validated—could be a useful conceptual contribution (Sec. 2.5, Sec. 3.5).

- Figures 1–10 are generally well-conceived for communicating embeddings, trajectories, and pseudotime trends, and could become very effective with clearer parameterization/annotation.

## Major issues

1. **Core confounding: lab and field datasets cover different biological processes and developmental windows, undermining the interpretability of the proposed “lab vs field regulatory strategy” contrast (Sec. 3.1–3.3, Sec. 3.5).** Field isolates are strongly enriched for gametocytes and late rings and largely lack trophozoites/schizonts, while lab data span the full asexual IDC with only limited sexual branching. Because trajectories/pseudotime are inferred separately within each subset, regulator→module lags can differ simply due to (i) different underlying lineages (IDC vs gametocytogenesis), (ii) different root choices (early ring vs late ring), (iii) missing intermediate states and uneven sampling density, and (iv) pseudotime warping/scaling differences—rather than reflecting a true mechanistic shift from “just-in-time” to “priming.”

*Recommendation:* Reframe the main comparative claim to avoid over-attributing differences to “lab vs field” per se unless you can compare the same biological transition. Concretely: (i) perform an apples-to-apples timing analysis restricted to a matched lineage present in both sources (e.g., late ring → early gametocyte and/or gametocyte maturation only, using the lab sexual branch if sufficient cells exist); (ii) report results both on the full datasets and on the matched subset(s), explicitly stating which stages are included; (iii) where matched analyses are not feasible due to limited overlap, temper the claim to: “field dataset (gametocytogenesis-enriched) shows longer inferred regulator–module separations than lab IDC,” and clearly label this as a hypothesis about priming that requires validation.

2. **Methods are incomplete/inconsistent and contain clear text corruption in key pipeline steps, preventing reproducibility and making it difficult to assess whether lab–field differences reflect biology or processing artifacts (Sec. 2.1–2.3, Sec. 2.5; inconsistencies with Sec. 3.1).** Sec. 2.1 has garbled input/metadata descriptions; Sec. 2.3 and Sec. 2.5 include duplicated/truncated sentences and do not cleanly specify HVG selection, PCA/UMAP/PAGA settings, the dynamic-gene model, module clustering, or the regulator cross-correlation procedure. Additionally, QC thresholds and outcomes are described inconsistently (Sec. 2.2 vs Sec. 3.1), and it is unclear whether the input is raw counts or already-normalized expression (risking double-normalization).

*Recommendation:* Rewrite Sec. 2.1–2.3 and Sec. 2.5 from the original source (not OCR) into a precise, stepwise, parameterized Methods description. At minimum include: (i) data provenance (platform, mapping/quantification, whether matrices are counts vs normalized values), and starting cells/genes per origin/sample; (ii) exact

QC thresholds and the number removed at each step (per origin and per donor/strain), reconciling Sec. 2.2 with Sec. 3.1; (iii) exact normalization/log1p/scaling steps and parameters; (iv) HVG selection function/flavor and the final HVG count used in each analysis; (v) neighbors graph, PCA/UMAP settings, clustering algorithm/resolution, and PAGA/DPT settings including root selection; (vi) the statistical model used for pseudotime-dynamic genes and how the “top 500” were selected with multiple-testing correction; and (vii) the full candidate-regulator/lag pipeline with explicit formulas and thresholds. Remove duplicated/corrupted phrases so an independent group could reproduce the analysis.

3. **Batch effects and integration across donors/isolates/experiments are not addressed sufficiently, yet origin-driven separation in UMAP/PAGA is interpreted biologically (Sec. 3.1–3.2).** Without explicit assessment/mitigation, lab–field separation can be driven by technical factors (library chemistry, capture method, ambient RNA, sequencing depth, processing site/time) or donor/strain effects rather than biological adaptation. The manuscript does not state whether any batch correction was attempted, nor does it quantify whether origin separation persists within the same annotated stage after controlling for donor/strain and technical covariates.

*Recommendation:* Provide a clear batch/integration plan and report its impact. Specifically: (i) describe batch variables available (donor ID for field; strain/replicate/run/day-in-culture for lab) and how they were used; (ii) quantify origin separation within matched stages (e.g., late rings, gametocytes) before/after correction (e.g., kBET/LISI, classifier accuracy with cross-validation, or variance partitioning); (iii) apply and justify an integration approach suitable for scRNA-seq (e.g., Harmony on PCA, BBKNN, Scanorama, scVI), then re-check key qualitative conclusions (UMAP separation, PAGA topology, pseudotime trends) under integrated vs non-integrated processing; and (iv) if you choose not to correct, justify why (same technology, same pipeline, etc.) and explicitly present evidence that technical effects are minor.

4. **The central “just-in-time vs priming” timing result is not statistically or metrically validated, and the lag metric (“bins out of 100”) is not comparable across separately inferred pseudotime trajectories without additional controls (Sec. 2.5, Sec. 3.5; Figs. 7–10).** The manuscript does not define the cross-correlation computation, the binning scheme, smoothing details, lag extraction rule, or uncertainty; nor does it test whether observed lags (e.g.,  $\sim 76$ – $77$  bins in field) exceed null expectations. Because bin-based lag depends on pseudotime scaling/warping and trajectory length, the headline quantitative contrast could be an artifact of discretization and trajectory coverage.

*Recommendation:* Make the timing analysis mathematically explicit and statistically supported. Concretely: (i) define pseudotime discretization (equal-width vs equal-cell-count bins; bin edges), smoothing (window size, boundary handling), and the exact cross-correlation formula (normalization/mean subtraction; handling of missing bins);

(ii) replace or complement “lag in bins” with a continuous-time measure (e.g., difference between regulator peak location and module activation/peak in continuous pseudotime from spline/GAM fits); (iii) provide uncertainty (bootstrap over cells and/or isolates; sensitivity to bin number and smoothing window); (iv) include null models (pseudotime permutation, matched mean/variance random genes, shuffled module assignments) and correct for multiple testing across gene–module pairs; and (v) explicitly test whether lag distributions differ between lab and field within matched lineage(s) (see confounding issue) using permutation tests or appropriate two-sample comparisons with confidence intervals.

5. **Dynamic gene selection, module construction, and the “candidate master regulator” definition are heuristic and under-specified, risking selection of technical artifacts or early markers rather than plausible regulators (Sec. 2.5, Sec. 3.4–3.5).** The manuscript does not specify the pseudotime-dynamics model/test (GAM vs other), covariates (batch/donor, library size), multiple-testing correction, or the rationale for exactly six modules. The regulator selection emphasizes low-abundance and sharp peaks, which can enrich for dropout/noise and may be sensitive to smoothing choices. The term “master regulator” implies causality not supported by correlation/lag alone.

*Recommendation:* Strengthen biological plausibility and methodological rigor of the regulator/module framework: (i) specify the exact model and hypothesis test used to rank/select dynamic genes (including FDR control), and report robustness to alternative model/parameters; (ii) justify six modules quantitatively (e.g., silhouette/gap statistic; stability under resampling) and report module sizes and functional enrichment (GO/KEGG where possible; curated malaria gene sets); (iii) constrain regulator candidates to plausible regulatory classes (e.g., ApiAP2 TFs, chromatin modifiers, RNA-binding proteins) or at least report enrichment of these classes among selected candidates versus background; (iv) control for technical covariates (library size/complexity; potential cell-cycle effects for asexual stages) in dynamic/regulator detection; (v) apply multiple-testing correction across regulator–module associations; and (vi) consistently rename to “candidate regulators” and present “master regulator” only as a hypothesis, ideally supported by overlap with known regulators (e.g., AP2-G/GDV1 axis; known gametocyte regulators) and/or external datasets.

6. **Comparative lab–field differential expression (especially within matched stages) is described but not convincingly presented, yet the Discussion/Conclusions make broad origin-specific functional claims (Sec. 2.6–2.7, Sec. 3.6/Results coverage, Sec. 4).** Given the confounding and potential batch effects, within-stage DE for overlapping stages (late rings and gametocytes) is essential to support claims about stress responses, immune evasion, sexual commitment, etc. The current Results do not provide counts of DE genes, effect sizes, FDR thresholds, donor-stratified consistency, or enrichment analyses supporting these statements.

*Recommendation:* Add (or expand) a dedicated Results subsection for origin comparisons on matched stages/lineages. Minimum expectations: (i) define stage labels and matching criteria; (ii) report DE method (e.g., pseudo-bulk per donor/strain, mixed models, or single-cell tests with donor as covariate), thresholds, and effect sizes; (iii) present donor-stratified consistency for field isolates; (iv) provide functional enrichment for robust DE sets; and (v) if trajectory-based DE (e.g., tradeSeq) is used, clearly specify the model (knots, covariates, contrasts) and summarize the major findings. If such analyses cannot be done robustly, soften/remove broad functional generalizations in Sec. 4 so every claim is traceable to a documented analysis.

7. **The manuscript appears not to be a careful final draft in several places, which materially impedes review and risks undermining confidence in the work (Abstract/Sec. 1; Sec. 2–3).** Examples include an unrelated physics keyword list, placeholder figure references (“Figure ??”), and other formatting/OCR artifacts; these issues occur alongside core methodological corruption (Sec. 2.3, Sec. 2.5).

*Recommendation:* Conduct a full pass to ensure the submitted version is publication-ready: remove irrelevant keywords and placeholder author/affiliation artifacts if present; replace all “Figure ??” with correct references; ensure all figures and captions are included and legible; and correct OCR-induced corruption throughout, prioritizing Sec. 2.1–2.5. Consider re-exporting directly from the source manuscript (LaTeX/Word) rather than via OCR to avoid reintroducing corruption.

## Minor issues

1. QC reporting is not sufficiently stratified by origin/sample, and Figure 1 does not clearly show thresholds or per-sample variability; gene-level filtering is not justified with cells-per-gene distributions (Fig. 1; Sec. 2.2; Sec. 3.1).

*Recommendation:* Update Figure 1 to include per-origin and ideally per-donor/strain facets or overlays. Add explicit threshold lines (e.g.,  $\geq 200$  genes,  $\geq 500$  counts) with counts/percentages below thresholds. Add a cells-per-gene distribution (or ECDF) with the chosen cutoff (3 vs 5 cells/gene) and report genes retained.

2. Parameter reporting and robustness for UMAP/PAGA/DPT are limited despite interpretive reliance on embedding/trajectory topology (Sec. 2.3–2.4; Sec. 3.2–3.3).

*Recommendation:* Report key parameters (neighbors graph  $k$ , UMAP  $n_{\text{neighbors}}/\text{\min\{\_}\text{dist}/metric}$ ; Leiden/Louvain resolution; PAGA edge threshold; DPT settings). Briefly summarize robustness checks (alternative UMAP/PAGA parameters; alternative roots) and whether the high-level conclusions persist.

3. Pseudotime rooting choices are described qualitatively but not operationally, and sensitivity to alternative roots is not shown (Sec. 2.4, Sec. 3.3).

*Recommendation:* Describe how the root cluster/cells were selected computationally (marker genes, cluster annotation, earliest node). Provide a sensitivity analysis showing how key lag results change under plausible alternative roots, especially for the field trajectory rooted at late rings.

4. Four field cells with infinite pseudotime are briefly noted but not characterized (Sec. 3.3).

*Recommendation:* Report their QC metrics, UMAP/PAGA positions, and cluster/stage labels; state whether they form a disconnected component. Justify exclusion and confirm that inclusion/exclusion does not change downstream module/lag conclusions.

5. Inconsistencies in reported pipeline settings: HVGs reported as 2,1000 in Methods vs 2,1500 in Results; gene filtering threshold as 5 vs 3 cells/gene; statements about removing low-quality cells conflict with “all cells and genes passed” (Sec. 2.2–2.3 vs Sec. 3.1–3.2).

*Recommendation:* Audit the manuscript for parameter consistency and align all sections/captions with the actual run configuration. If different settings were used for different analyses, explicitly list them in a single table (analysis → parameters) and reference it from Sec. 2.

6. Figures 2–3 have labeling/legend clarity issues (e.g., explained variance depiction, cluttered labels, unclear categorical legends), making it difficult to evaluate PCA choices and origin separation (Figs. 2–3; Sec. 3.2).

*Recommendation:* Standardize axis labels (raw explained variance ratio plus cumulative curve), annotate the selected PC cutoff, reduce label clutter, and ensure all color encodings have readable legends. Where origin separation is interpreted, add a quantitative separation metric within matched stages (pre/post integration if applicable).

7. Figures 5–6 and 7–10 do not consistently annotate module identities, scaling/normalization choices for plotted curves, bin counts, or biological landmarks along pseudotime (Sec. 3.4–3.5).

*Recommendation:* Add explicit module labels (and gene counts), clarify whether curves are mean( $\log 1p$ ),  $z$ -scored, or min–max scaled, display per-bin cell counts, and overlay stage landmarks/transition windows to make timing interpretations more transparent.

8. R-based tools (e.g., tradeSeq via rpy2) are mentioned ambiguously; it is unclear what was actually run versus proposed (Sec. 2.6–2.7).

*Recommendation:* State explicitly which external tools were used in the final results, with versions and key parameters, and point to the exact Result subsection/figure where those outputs appear. If not used, rewrite as future work.

9. Reproducibility and data/code availability are not clearly stated despite a multi-step pipeline (Sec. 2.7, Sec. 4).

*Recommendation:* Add a clear Data/Code Availability statement with repository links (GitHub/Zenodo), environment specification (e.g., conda YAML), and data accessions (GEO/ENA). Provide processed artifacts (AnnData objects, module assignments, pseudotime values) sufficient to reproduce figures.

10. Abstract/Introduction wording implies network inference and variability analyses that are not clearly presented as dedicated quantitative results (Abstract; Sec. 1; Sec. 3.4–3.5).

*Recommendation:* Align claims with presented analyses: either add explicit variability/network quantification or soften the Abstract/Intro to describe module/timing correlations rather than inferred regulatory networks.

## Very minor issues

1. Numerous typographical/formatting/OCR artifacts reduce readability (split words, duplicated fragments, inconsistent code formatting, inconsistent species name formatting, placeholder figure references) across Sec. 1–3, especially Sec. 2.3 and Sec. 2.5.

*Recommendation:* Proofread against the source manuscript and re-export cleanly. Standardize formatting for *P. falciparum*, percentages, gene IDs, and code (backticks). Remove duplicated sentences and fix all placeholder references.

2. Irrelevant physics keywords appear in the Abstract/early sections and should not be present in a malaria/scRNA-seq manuscript (Abstract, Sec. 1).

*Recommendation:* Replace with relevant keywords (e.g., *Plasmodium falciparum*, malaria, scRNA-seq, gametocytogenesis, trajectory inference, pseudotime, PAGA).

3. Candidate regulator gene IDs are presented with limited functional context, reducing accessibility for non-specialists (Sec. 3.5.1–3.5.2).

*Recommendation:* When first introducing highlighted candidates (e.g., PF3D7\_XXXXXXX), add short functional annotations (TF family/chromatin/RBP/unknown) and citations where known; ensure consistent gene-ID formatting.

4. Variability metrics and smoothing steps are mentioned without precise definitions (e.g., CV formula; handling zeros/log<sub>1p</sub>; smoothing window and order of transformations), which affects interpretation of “sharp transient peaks” (Sec. 2.5).

*Recommendation:* Define CV (and any alternative metrics) explicitly, state whether computed on raw counts, normalized counts, or log<sub>1p</sub> values, and specify smoothing window/algorithm and the exact transformation sequence used for peak calling and correlation.

## Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

**Maths relevance:** light

The paper contains very little explicit mathematics (no numbered equations or derivations). The quantitative content is primarily methodological descriptions ( $\log 1p$  transform,  $z$ -score scaling, min-max scaling, cross-correlation lag, coefficient of variation/variance comparisons) and pseudotime bin/lag statements. The main internal-consistency checks therefore focus on definition consistency and whether stated quantitative procedures are sufficiently specified to support the paper's central timing claims.

### Checked items

#### 1. ✓ **Log transform definition** (Sec. 2.2, p.3)

- **Claim:** Data were normalized and log-transformed using  $\log 1p(X)$  to stabilize variance.
- **Checks:** definition consistency, domain/sanity check
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:**  $X$  denotes the (normalized) expression matrix,  $\log 1p(X)$  means elementwise  $\log(1 + X)$
- **Notes:** As stated,  $\log 1p$  is a standard elementwise transform and is mathematically well-defined for nonnegative  $X$  (typical for normalized counts). The paper does not contradict this elsewhere.

#### 2. △ **Z-score scaling step** (Sec. 2.3, p.3)

- **Claim:** Selected HVG expression values were scaled ( $z$ -score transformation) across cells, with clipping for outliers.
- **Checks:** definition clarity, internal consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** Scaling is performed per gene across cells, Clipping occurs after scaling (not fully specified)
- **Notes:**  $z$ -score scaling is conceptually clear, but the paper does not specify whether scaling is per-gene or per-cell (common practice is per-gene). This is not contradicted later, but it is not fully auditable.

#### 3. ✘ **HVG count inconsistency (2000 vs 2500)** (Sec. 2.3, p.3 (2000 HVGs) vs Sec. 3.2, p.4 (2,500 HVGs))

- **Claim:** The analysis uses a specified number of highly variable genes for dimensionality reduction.
- **Checks:** parameter consistency
- **Verdict:** FAIL; confidence: high; impact: moderate

- **Assumptions/inputs:** HVG count is a key input to PCA/UMAP
  - **Notes:** Methods states 2000 HVGs are selected; Results states UMAP used 2,1500 HVGs. These cannot both be true for the same analysis unless explicitly describing different runs, which is not stated.
4. ✘ **Gene filtering threshold inconsistency (5 vs 3 cells)** (Sec. 2.2, p.2 (fewer than 5 cells) vs Sec. 3.1, p.4 (minimum 3 cells per gene))
- **Claim:** Genes below a detection frequency threshold are removed.
  - **Checks:** definition/threshold consistency
  - **Verdict:** FAIL; confidence: high; impact: moderate
  - **Assumptions/inputs:** Gene detection means expression  $> 0$  (as earlier stated for detected genes per cell)
  - **Notes:** The paper states two different minimum detection thresholds for genes ( $\geq 5$  cells vs  $\geq 3$  cells). It is unclear which was actually applied to produce the reported results.
5. △ **Cell filtering logic vs statement 'all cells passed'** (Sec. 2.2, p.2 vs Sec. 3.1, p.4)
- **Claim:** Stringent filtering removes low-quality cells, but later all cells and genes pass initial filters.
  - **Checks:** logical consistency
  - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
  - **Assumptions/inputs:** Filtering criteria are actually executed on the dataset used downstream
  - **Notes:** Methods describes removing cells below thresholds; Results claims all cells and genes passed the initial filters. This could be consistent (filters defined but removed nothing), but the paper does not explicitly confirm post-filter counts or whether additional thresholds (e.g., 'below two SD') removed anything.
6. ✘ **Relative expression definition ambiguity (z-score vs min-max)** (Sec. 2.5, p.3 (z-score example and later min-max scaling))
- **Claim:** Candidate regulators are identified using 'relative' expression peaks preceding module activation.
  - **Checks:** definition consistency, method-specification adequacy
  - **Verdict:** FAIL; confidence: high; impact: critical
  - **Assumptions/inputs:** A single time series per gene is used for peak calling and cross-correlation
  - **Notes:** Two different normalizations are presented as 'relative expression' (z-scored within gene; min-max scaled smoothed profile). The Results/Figures (Sec. 3.5, Figs. 7–10) do not specify which normalization underlies the lag estimates, making the central timing analysis non-auditable.

7.  $\triangle$  **Cross-correlation lag claim needs explicit formula** (Sec. 2.5, p.3 and Sec. 3.5–3.5.2, pp.6–7)

- **Claim:** Lag between regulator and module activation is determined via cross-correlation along pseudotime, yielding lags of 1–2 bins (lab) and  $\sim$  76–77 bins (field).
- **Checks:** missing derivation/definition, logical adequacy for main claim
- **Verdict:** UNCERTAIN; confidence: high; impact: critical
- **Assumptions/inputs:** Pseudotime profiles are discretized into bins ('out of 100'), Cross-correlation is computed between two binned series
- **Notes:** The paper does not provide the mathematical definition of the cross-correlation used (normalization, mean subtraction, lag range, edge handling) nor the rule converting correlation output to a reported lag. Without this, the stated lag magnitudes cannot be verified from the paper's mathematics.

8.  $\triangle$  **Pseudotime 'bins out of 100' undefined** (Sec. 3.5.1, p.7 ('1-2 bins (out of 100)') and Sec. 3.5.2, p.7)

- **Claim:** Pseudotime is treated as discretized into 100 bins for lag reporting.
- **Checks:** definition completeness, internal consistency
- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** Continuous DPT pseudotime is mapped to 100 discrete bins
- **Notes:** Earlier sections describe continuous diffusion pseudotime but do not define any discretization into exactly 100 bins (binning rule, per-condition comparability). The lag statements depend on this discretization.

9.  $\triangle$  **Coefficient of variation / transition-window variability screening** (Sec. 2.5, p.3)

- **Claim:** Genes with higher variability (CV or variance) in transition windows than in stable windows are prioritized.
- **Checks:** definition completeness, sanity conditions
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** Transition windows are defined by pseudotime/stage boundaries, Expression values are on a specified scale (likely  $\log 1p$ )
- **Notes:** No explicit CV formula, no statement of whether sd/mean is computed on raw, normalized, or  $\log 1p$  data, and no handling described for near-zero means (where CV can diverge). This is not an algebraic contradiction but blocks a strict audit.

10.  $\checkmark$  **Root choice for field isolates due to missing early stages** (Sec. 3.3, p.5)

- **Claim:** Field pseudotime is rooted in late ring because early asexual stages are absent.

- **Checks:** logical consistency
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** Pseudotime requires a designated root or starting set
- **Notes:** Within the narrative, this is internally consistent: lab uses early ring; field uses the earliest available group (late ring).

## Limitations

- The provided PDF text contains no explicit equations or step-by-step derivations; most quantitative procedures are described verbally, which limits the ability to audit algebraic correctness beyond checking for definitional and parameter consistency.
- Figures are referenced with placeholders (e.g., 'Figure ??') in multiple places, and figure axis labels/definitions are not fully available in the parsed text; this constrains verification of what transformations are actually plotted.
- Key algorithmic definitions (cross-correlation formula, pseudotime binning rule, smoothing window) are omitted, preventing a complete symbolic audit of the central lag-based regulatory timing claims.

## Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

10 numeric checks were run: 8 PASS, 1 FAIL, and 1 UNCERTAIN. The key numeric inconsistency identified is the mismatch between 2,!000 vs 2,!500 HVGs across Methods and Results. Several other quantitative claims (QC recomputation, sparsity recomputation, module-size summation, and pseudotime lag verification) remain unverified because they require underlying matrices/CSVs or figure-derived values.

### Checked items

1. ✓ **C1\_total\_cells\_sum** (p.4 (Results 3.1), also p.1-2 (Abstract/Intro mentions total))
  - **Claim:** Dataset comprises 45,!691 individual parasites, combining laboratory strains (37,!624 cells) and field isolates (8,!067 cells).
  - **Checks:** parts\_to\_total
  - **Verdict:** PASS
  - **Notes:** Verified  $37,!624 + 8,!067 = 45,!691$ .
2. ✓ **C2\_infinite\_pseudotime\_exclusion\_count** (p.5 (Results 3.3))
  - **Claim:** In field isolates, 4 out of 8,!067 cells were assigned infinite pseudotime values and excluded from subsequent pseudotime-dependent analyses.
  - **Checks:** ratio\_and\_remaining\_count
  - **Verdict:** PASS

- **Notes:** Remaining cells computed as  $8,1067 - 4 = 8,1063$ ; fraction infinite computed as  $4/8,1067 \approx 0.00049585$ .
3. ✓ **C3\_sparsity\_percentage\_bounds** (p.4 (Results 3.1))
- **Claim:** Overall sparsity of the expression matrix was 80.25%.
  - **Checks:** percentage\_range\_check
  - **Verdict:** PASS
  - **Notes:** Bounds-only check: 80.25% lies within  $[0, 100]$ .
4. △ **C4\_genes\_filter\_threshold\_consistency** (p.4 (Results 3.1))
- **Claim:** Filtering thresholds: minimum 200 genes and 500 total counts per cell; minimum 3 cells per gene.
  - **Checks:** threshold\_internal\_consistency
  - **Verdict:** UNCERTAIN
  - **Notes:** This item's expected relationship references medians (937; 2,1059) that are not included in this candidate's provided values, so the comparison could not be executed for this item.
5. ✓ **C5\_medians\_vs\_thresholds** (p.4 (Results 3.1))
- **Claim:** Median detected genes per cell was 937; median sum of normalized expression counts was 2,1059; thresholds were minimum 200 genes and 500 counts per cell.
  - **Checks:** unit\_consistent\_numeric\_comparison
  - **Verdict:** PASS
  - **Notes:** Computed ratios:  $937/200 = 4.685$  and  $2,1059/500 = 4.118$  (arithmetic check).
6. ✗ **C6\_hvg\_count\_inconsistency\_2000\_vs\_2500** (p.3 (Methods 2.3) vs p.4 (Results 3.2))
- **Claim:** Methods: selecting 2000 HVGs for downstream analysis; Results: UMAP on 2,1500 most highly variable genes.
  - **Checks:** repeated\_constant\_mismatch
  - **Verdict:** FAIL
  - **Notes:** Reported HVG counts are not equal (2,1000 vs 2,1500).
7. ✓ **C7\_dynamic\_genes\_count\_consistency** (p.6 (Results 3.4))
- **Claim:** For both lab and field isolates, identified the top 500 most dynamically expressed genes.
  - **Checks:** constant\_reuse\_consistency
  - **Verdict:** PASS
  - **Notes:** Checked that 500 is a positive integer.

8. ✓ **C8\_modules\_count\_and\_size\_bounds** (p.6 (Results 3.4))
  - **Claim:** In lab isolates, 500 genes were partitioned into six modules; module sizes range from 35 to 130 genes.
  - **Checks:** range\_feasibility\_check
  - **Verdict:** PASS
  - **Notes:** Feasibility check: with 6 modules of size 35–130, total possible range is 210–780 genes; 500 lies within this range.
  
9. ✓ **C9\_pseudotime\_bins\_short\_lag\_range** (p.7 (Results 3.5.1))
  - **Claim:** In lab isolates, regulator precedes Module 2 activation by a pseudo-time lag of 1–2 bins (out of 100).
  - **Checks:** range\_within\_total\_bins
  - **Verdict:** PASS
  - **Notes:** Verified  $0 \leq 1 \leq 2 \leq 100$ .
  
10. ✓ **C10\_pseudotime\_bins\_long\_lag\_range** (p.7 (Results 3.5.2))
  - **Claim:** In field isolates, regulator peak precedes Module 5 activation by approximately 76–77 pseudotime bins.
  - **Checks:** range\_within\_total\_bins
  - **Verdict:** PASS
  - **Notes:** Verified  $0 \leq 76 \leq 77 \leq 100$  under an explicit assumption that field isolates use 100 pseudotime bins (the paper does not explicitly confirm this for field).

### Limitations

- Only PDF text was available; underlying matrices (geneexpression.csv, labels.csv) and referenced CSV outputs (e.g., gene\_modules\_lab.csv) were not provided, preventing recomputation of QC metrics, sparsity, module sizes, and pseudotime relationships.
- Several quantitative claims are only visualized in figures without explicit numeric tables; verifying them would require extracting values from image pixels, which is out of scope per instructions.
- Some checks (e.g., long-lag bins within total bins) rely on an implicit assumption that binning (100 bins) is shared across lab and field analyses; the PDF does not explicitly confirm this for field isolates.