

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

Summary

This manuscript uses single-cell RNA-seq with pseudotime trajectory inference (DPT + PAGA) to compare transcriptional dynamics during the asexual intraerythrocytic developmental cycle (IDC) of *Plasmodium falciparum* in laboratory-adapted strains versus field isolates from asymptomatic human infections (Sec. 2.1–2.2, Sec. 3.1). The central methodological idea is a burst-based heuristic for “candidate master regulators”: genes with low overall expression that exhibit sharp, transient expression bursts shortly before inferred developmental transitions along pseudotime; putative downstream “modules” are then defined using post-burst differential expression and/or lagged correlations (Sec. 2.3–2.4, Sec. 3.2–3.3). Using $\sim 43,000$ cells (36,520 lab; 6,866 field), the headline comparative result is that the top 100 burst-defined candidates in lab and field show zero overlap (Sec. 3.4, Table 4), which is interpreted as evidence for profoundly divergent regulatory programs in vitro vs in vivo. The biological question and dataset are compelling, but key elements required to evaluate and trust the main conclusion—auditable method specification, trajectory/transition validation, robustness and statistical assessment of the zero-overlap finding, treatment of confounders (batch, cell-number imbalance, isolate pooling), and functional/biological grounding of candidates and modules—are currently insufficiently developed, and some descriptions are internally inconsistent (notably around stage labels/transition definition; Sec. 2.2.3, Sec. 2.3.3, Sec. 3.1).

Strengths

- Addresses an important biological and translational question: how in vitro adaptation vs in vivo conditions may reshape asexual blood-stage transcriptional programs in *P. falciparum* (Introduction, Sec. 3.4, Conclusions).
- Leverages a large scRNA-seq dataset with clear lab vs field stratification and multiple field isolates, enabling comparative analysis (Sec. 2.1, Sec. 3.1, Table 1).
- Uses contemporary trajectory inference tooling (PAGA + DPT) that is appropriate for asynchronous parasite populations and for interrogating pseudotime dynamics (Sec. 2.2, Sec. 3.1).
- Proposes an explicit, testable heuristic (low-expression + transient pre-transition burst) and a clear pipeline (filtering \rightarrow smoothing \rightarrow peak detection \rightarrow candidate ranking \rightarrow module definition) that could be valuable if fully specified and validated (Sec. 2.3–2.4).
- Presents visually interpretable trajectory plots and burst-expression panels that make the core hypothesis easy to grasp (Figures 1–3).

- The observation of zero overlap between top-ranked candidate lists (Table 4) is striking and, if robust to confounders and parameter choices, could motivate substantial follow-up work.

Major issues

1. **Reproducibility is currently limited because key algorithmic and parameter details are missing or described hypothetically** (e.g., “if not already log-transformed...”, “e.g.” thresholds, “empirically determined”), even though thresholding, smoothing, and peak-calling drive the main result (Sec. 2.1–2.4). Critical unspecified items include: the exact normalization used (and whether applied jointly or separately to Lab/Field); whether/which log transform was applied; concrete cell/gene QC thresholds and outcomes; HVG method and number of HVGs; number of PCs retained and selection criterion; neighborhood graph parameters; clustering resolution; PAGA pruning/thresholding; DPT settings; trajectory rooting rule; smoothing method and bandwidth/span; peak-finding algorithm and its parameters (prominence, height, min distance/width); handling of multiple peaks per gene; and the precise rule(s) for module membership, statistical tests, and multiple-testing corrections (Sec. 2.3–2.4).

Recommendation: Expand Sec. 2.1–2.4 into an auditable, deterministic specification (main text + supplement) and add a single consolidated parameter table. At minimum: (i) state the exact normalization (e.g., size-factor/CPM/TPM/Scanpy `\text{normalize_total}`) and whether Lab/Field were normalized separately or together; (ii) state explicitly whether values are linear or $\log 1p$ and carry that notation consistently; (iii) report concrete QC thresholds (min genes/counts per cell, min cells per gene, rRNA/mitochondrial thresholds if used, doublet handling) and the number/% cells/genes removed; (iv) specify HVG selection method, number of HVGs, scaling, and PCs used; (v) list PAGA/DPT parameters (k NN k , metric, clustering method/resolution, PAGA connectivity threshold, DPT settings) and whether identical settings were used for Lab and Field; (vi) define smoothing (method + bandwidth) and peak calling (algorithm + all parameters) and how multiple peaks are handled; (vii) define module calling with explicit statistical tests, effect sizes, FDR correction, thresholds, and the exact Boolean rule combining DE and lagged correlation (union vs intersection). Include software packages and versions.

2. **Internal inconsistency around stage labels/marker genes and the definition of “pre-transition” bursts undermines the conceptual core.** Methods describe transition-aware burst selection and alignment using life-cycle stage labels/marker genes (Sec. 2.2.3, Sec. 2.3.3), but Results state that provided `\text{life_cycle_stage}` labels were “uninformative” and canonical IDC marker genes were absent from the gene list (Sec. 3.1). If transitions cannot be defined/validated, it is unclear how “immediately preceding a de-

velopmental transition” is operationalized, how the trajectory is rooted, and whether the method in practice reduces to detecting peaks anywhere along pseudotime.

Recommendation: Resolve the discrepancy explicitly. Either: (A) provide a label-free, fully specified method for rooting and defining transition windows (e.g., graph/geodesic landmarks, change-point detection on pseudotime density or cluster boundaries, or external reference projection), including numeric pseudotime ranges for transitions and pre-/post-burst windows (Sec. 2.3.3, Sec. 2.4.1); or (B) if robust transition definition is not possible with available genes/labels, reframe throughout (Abstract, Sec. 2.3, Sec. 3.2–3.4, Conclusions) as “transient bursts along pseudotime” rather than “pre-transition master regulators,” and temper causal/transition-specific language accordingly.

3. **The central comparative claim (“profound divergence” / “fundamentally different regulatory architectures”)** is driven almost entirely by the observation of zero overlap between the top 100 burst-defined candidates in Lab vs Field (Sec. 3.4, Table 4), but the manuscript does not quantify (i) whether zero overlap is statistically unexpected given the tested gene universe and filtering, or (ii) whether it is robust to parameter choices, list size, or sampling variability. Given the heavy dependence on ranking by peak prominence/height and on low-expression filtering, the zero-overlap result could be fragile.

Recommendation: In Sec. 3.4 (and Sec. 2.5), add a dedicated robustness + significance section: (i) report the total shared gene universe evaluated and the number of genes passing each filter in Lab and Field; (ii) compute expected overlap under a null model (hypergeometric conditioned on the shared universe and candidate counts; and/or permutations that preserve marginal properties such as mean expression and dropout), reporting p -values and confidence intervals; (iii) show overlap as a function of list size (top 10/25/50/100/200/500) and as a function of key thresholds (low-expression percentile; burst height/prominence cutoffs; smoothing bandwidth); (iv) assess stability via bootstrap/subsampling of cells within each group; (v) report rank correlations (or lack thereof) of burst scores genome-wide. Based on these results, either strengthen the claim or revise wording to “divergent burst-defined candidate lists under the chosen procedure.”

4. **Potential confounders are not adequately controlled: Lab vs Field differ substantially in cell number (36,520 vs 6,866; Table 1), may differ in sequencing depth/dropout and batch structure, and Field aggregates multiple genetically distinct isolates into one pooled trajectory.** These factors can change HVG selection, manifold geometry, pseudotime density, smoothing behavior, peak prominence ranking, and perceived burst “sharpness,” and therefore could drive apparent divergence independent of environment (Sec. 2.1–2.2, Sec. 3.1–3.4).

Recommendation: In Sec. 2.1–2.2 and Sec. 3.1–3.4: (i) explicitly describe batch structure (runs/libraries/lanes; field-vs-lab processing differences) and whether any integration/batch-correction was applied; if not, add an analysis using an established approach (e.g., Harmony/Seurat integration/scVI) and reassess trajectories and candidate lists; (ii) control for cell-number imbalance by downsampling Lab to match Field and repeating candidate calling and overlap analyses; (iii) analyze Field isolates separately where feasible (per-isolate trajectories and candidate lists), quantify within-field overlap/heterogeneity, and test whether pooling reduces peak sharpness or alters rankings. If per-isolate power is limited, present it as a limitation and avoid over-interpreting pooled results.

5. **Trajectory validity (IDC interpretation) is not convincingly established, yet it is central to biological interpretation of burst timing and transitions (Sec. 3.1). The manuscript notes that canonical marker genes are absent and stage annotations are uninformative, raising the risk that pseudotime reflects technical gradients (library size, rRNA content, stress) rather than developmental progression.**

Recommendation: Strengthen Sec. 3.1 with quantitative validation: (i) project cells onto external IDC references (e.g., published bulk time-course IDC transcriptomes) via correlation/nearest-neighbor mapping using the available gene set; and/or test enrichment of IDC phase gene sets along pseudotime even if single canonical markers are missing; (ii) show correlations of pseudotime with technical covariates (UMIs, detected genes, rRNA/mitochondrial fraction where applicable) to rule out dominant technical gradients; (iii) verify robustness of global ordering with at least one alternative trajectory method (e.g., Slingshot/Monocle3) and/or parameter perturbations of k NN/clustering/PAGA pruning. If validation remains weak, explicitly qualify that IDC stage mapping is approximate and that results are hypothesis-generating.

6. **“Candidate master regulators” are defined purely by expression dynamics (low overall expression + transient burst), but the manuscript provides little evidence that these candidates are plausible regulators rather than low-count/noisy genes. Candidates are mostly reported as PF3D7 IDs with limited functional annotation, and there is no enrichment test for known regulatory families (ApiAP2 TFs, chromatin regulators, RNA-binding proteins), nor basic checks that bursts are supported by broad cell-level signal rather than a small number of outlier cells (Sec. 3.2, Tables 2–3).**

Recommendation: In Sec. 3.2 and supplement: (i) annotate candidates using PlasmoDB (gene names, domains, predicted localization, known phenotypes) and explicitly flag known/predicted regulatory proteins; (ii) test enrichment of regulatory classes among candidates vs the expressed-gene background (report odds ratios and adjusted p -values); (iii) report per-candidate expression support metrics (fraction of cells ex-

pressing, mean/median counts in peak window vs baseline, robustness to removing top-expressing cells) to reduce the chance that bursts reflect sparse dropout noise; (iv) adjust claims to “candidate regulators” unless additional evidence is provided.

- 7. Downstream module definition is under-specified and the resulting module sizes (e.g., >1,000 genes for some candidates; Sec. 3.3) raise concerns that modules may reflect broad pseudotime progression rather than specific regulator–target relationships. It is unclear how pre-/post-burst windows are set per gene, what DE model/test is used, how multiple testing is corrected, what lagged-correlation metric/lag range is used, and whether DE and lag criteria are combined (union/intersection). Extremely large reported $\log_2\text{FC}$ values (e.g., >24; Sec. 3.3) also suggest potential baseline/pseudocount artifacts.**

Recommendation: In Sec. 2.4 and Sec. 3.3: (i) precisely define module membership rules, windows, tests (including covariates), effect sizes, and FDR procedures; (ii) add controls to assess specificity (e.g., modules from random “burst times,” or from non-regulatory genes with similar peak times/mean expression); (iii) report module coherence metrics (average within-module correlation; reproducibility across subsamples; functional enrichment consistency); (iv) investigate and explain extreme $\log_2\text{FC}$ values (pseudocount choice, linear vs log scale, near-zero baseline) and cap/regularize where appropriate. If modules remain large, explicitly interpret them as broad stage-associated programs rather than specific regulatory targets.

- 8. Functional enrichment analysis is described (Sec. 2.4.3) but not presented in a way that supports the manuscript’s narrative about “distinct regulatory strategies” in Lab vs Field. Without enrichment tables/plots and clear background/universe definitions, module interpretation remains anecdotal (Sec. 3.3–3.4).**

Recommendation: Either present the promised enrichment results or remove/scale back the claims. Concretely: (i) for representative modules (and/or aggregated across top candidates), report GO/KEGG enrichment with adjusted p -values, effect sizes, and the exact background universe (e.g., all expressed genes in that group); (ii) compare Lab vs Field at the level of enriched terms/pathways (e.g., Jaccard similarity or correlation of $-\log_{10}$ FDR across terms), which can reveal convergent biology even when gene-level regulator lists differ; (iii) provide full enrichment tables in supplementary files and cite them in Sec. 3.3–3.4.

Minor issues

1. Quality control is described using conditional/example language and ranges (e.g., “e.g., below 5th percentile”, “3–5 cells”), making it unclear what was actually applied and whether thresholds were identical across Lab and Field (Sec. 2.1.2, Sec. 2.1.3).

Recommendation: Rewrite Sec. 2.1.2–2.1.3 to state the exact QC and filtering thresholds used (single values, not ranges), separately for Lab and Field if they differ, and report the number/% of cells and genes removed at each step. Add basic QC plots (counts/genes distributions, dropout rates) to the supplement and reference them.

2. Normalization/scaling comparability across Lab and Field is unclear. Differences in peak heights between groups (e.g., Field peaks >1 while Lab peaks $\sim 0.7\text{--}0.8$) suggest either different scaling or different preprocessing, which complicates direct comparison of burst metrics (Sec. 3.2, Tables 2–3).

Recommendation: Clarify whether expression was normalized/log-transformed jointly or separately by group, and whether burst metrics are computed on comparable scales. If computed separately, state this explicitly and avoid direct numerical comparisons of peak height across groups, or rerun preprocessing jointly to enable comparability.

3. Fold-change and “ $>1.5\text{--}2x$ ” burst criteria are described alongside “log-normalized” expression without clarifying whether criteria are computed in linear space or log space; this affects interpretability and reproducibility (Sec. 2.3.1–2.3.3).

Recommendation: State explicitly the space used for burst detection (linear normalized counts vs $\log 1p$). If using log space, define thresholds additively ($\Delta \log$ expression) or specify the back-transformation used before computing fold-change.

4. Figures 1–2 (PAGA/trajectory) do not fully specify visual encodings (node size, edge width/strength, color scale), root selection, or directionality. Figure 3 lacks clear y-axis labeling/units and does not clearly communicate smoothing/binning choices (Figures 1–3; Sec. 3.1–3.2).

Recommendation: Expand captions to define node/edge encodings, pseudotime color mapping (0–1), rooting, and progression direction (mark root and add arrows). For Figure 3, add y-axis label (including transformation), describe smoothing/bins, and optionally add confidence bands or cell-density along pseudotime.

5. Stage labels (ring/trophozoite/schizont) are presented in figures despite limited validation, and Figure 2 appears to include gametocyte labels while the analysis focuses on the asexual IDC, creating potential interpretive confusion (Sec. 3.1; Figures 1–2).

Recommendation: Qualify stage labels as putative (or move detailed staging to text with supporting evidence). If gametocytes are present, explicitly state whether they were included/excluded from trajectory inference and, if included, show branching or separate plots/analyses to avoid conflating lineages.

6. The manuscript’s framing (“master regulators”, “fundamentally different programs”) likely overstates what can be concluded from one burst-based heuristic without experimental validation and given current uncertainty about transitions and confounders (Abstract, Sec. 1, Sec. 3.4, Conclusions).

Recommendation: Revise the Abstract/Introduction/Conclusions to consistently use “candidate/putative regulators” and to qualify that regulatory roles are inferred from expression dynamics and require experimental validation. Replace absolute language (“fundamentally different architecture”) with claims tied to the operational definition and robustness results.

7. Table 1 summarizes depth/genes detected mainly via $\text{mean} \pm \text{SD}$; given likely differences in dropout and heavy-tailed distributions, this may obscure comparability between Lab and Field (Table 1; Sec. 3.1).

Recommendation: Add median and IQR (and/or quantiles) for key QC metrics and include distribution plots (violin/histograms). Consider reporting gene-level dropout summaries to contextualize low-expression/burst calling.

8. Ethics/provenance information for field isolates is insufficiently documented (Sec. 2.1.3).

Recommendation: Add an Ethics/Compliance statement (IRB/ethics approvals, consent, de-identification), and provide accession IDs / original-study citation if data are re-used.

9. Presentation/metadata issues reduce manuscript professionalism: the author/affiliation line in the unstructured version (“Anthropic, Gemini & OpenAI servers. Planet Earth.”) and overly generic keywords are not appropriate for a scientific submission.

Recommendation: Replace with correct author names/affiliations (or omit in a blinded draft) and update keywords to include malaria/Plasmodium, IDC, scRNA-seq, trajectory inference/pseudotime, ApiAP2/chromatin regulation as relevant.

Very minor issues

1. Typographical and formatting artifacts (broken words/line breaks such as “environment”, stray markdown/LaTeX like “>!1.5”, inconsistent heading styles with stray “#”, malformed superscripts/bold, and confusing phrases such as “0.7556th detected burst peaks”) reduce readability (Sec. 1–3; Figure 3 caption).

Recommendation: Proofread and clean the manuscript for consistent typography and rendering. Standardize headings, remove stray markdown/LaTeX, and rewrite unclear caption text (e.g., specify “peak height = 0.7556” rather than “0.7556th peak”).

2. Some figure/table references appear as placeholders (e.g., page8_img*.jpg) and are not well integrated into the Results narrative (Sec. 3.1–3.3).

Recommendation: Ensure all figures are properly embedded, numbered, and explicitly described in the text (what each panel shows and why it matters). Remove placeholder references.

3. Terminology varies between “normalized”, “log-normalized”, and “log₂ fold-change” without consistent notation for base and transformation, which complicates interpretation (Sec. 3.3).

Recommendation: Introduce consistent notation (e.g., x = linear normalized; y = $\log_2 \text{FC}(x)$; $\log_2 \text{FC}$ computed from x with specified pseudocount) and use it throughout.

Key statements and references

- • Utilizing diffusion pseudotime (DPT) integrated with Partition-Based Graph Abstraction (PAGA), the authors inferred a dominant developmental path for both laboratory and field *Plasmodium falciparum* populations, interpreting the primary axis of variation as representing the intraerythrocytic developmental cycle (IDC), consistent with prior applications of DPT and PAGA to scRNA-seq developmental data.
- *Reference(s):* [11]
- • The trajectory inference was rooted at the inferred earliest pseudotime point, presumed to correspond to the ring stage, and its biological plausibility was assessed by examining the distribution of annotated stages and the progression of known stage-specific marker gene expression along pseudotime, following standard scRNA-seq trajectory validation practices described in earlier work.
- *Reference(s):* [11]
- • Highly variable genes (HVGs) used for dimensionality reduction were identified within each group using standard methods based on the mean-variance relationship and variance-stabilizing transformation, approaches that have been widely adopted in single-cell RNA-seq analyses to capture dynamic transcriptional variation.
- *Reference(s):* [11]
- • Gene set enrichment analysis of putative downstream transcriptional modules was performed using Gene Ontology (GO) terms and KEGG pathways with *Plasmodium falciparum*-specific annotations, following established pathway-enrichment methodologies to infer the biological processes regulated by candidate master regulators.
- *Reference(s):* [11]

Mathematical consistency audit

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

Maths relevance: light

The document is primarily a methods/results narrative for scRNA-seq pseudotime analysis and peak-based candidate selection. It contains essentially no explicit equations or multi-step derivations. The central quantitative logic relies on definitions (log transforms, percentile thresholds, smoothing, peak detection metrics, fold-change criteria, set overlap measures, and statistical tests) that are described verbally but not formalized; therefore most symbolic verification is blocked by missing mathematical specification.

Checked items

1. ✓ **Pseudotime normalization vs reported burst positions** (Sec. 2.5.2 (p.5), Tables 2–3 (p.7))
 - **Claim:** Lab/Field pseudotime scales are normalized to 0–1 for comparisons; burst pseudotimes are reported for top candidates.
 - **Checks:** range/sanity check, definition consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Burst pseudotime values in tables are on the same normalized scale described in Methods.
 - **Notes:** All reported burst pseudotimes in Tables 2–3 lie within $[0, 1]$, consistent with a normalized pseudotime scale.

2. ✓ **Set-overlap arithmetic in Table 4** (Table 4, Sec. 3.4 (p.8))
 - **Claim:** Top 100 Lab regulators and Top 100 Field regulators have 0 shared regulators; thus 100 are unique to each group.
 - **Checks:** set-cardinality consistency
 - **Verdict:** PASS; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Each list contains exactly 100 distinct gene IDs (no duplicates within a list).
 - **Notes:** If each list has 100 unique items, then shared= 0 implies unique-to-Lab= 100 and unique-to-Field= 100. The table’s counts are internally consistent. The only uncheckable assumption is absence of duplicates within each top-100 list (not shown).

3. △ **Log transform vs fold-change burst criterion** (Sec. 2.1.2 (p.3), Sec. 2.3.2 (p.4), Sec. 3.3 (p.8))
 - **Claim:** Data may be \log_2 -transformed; bursts are defined by a peak being > 1.5 or 2-fold above a median baseline; downstream effects discuss \log_2 fold-change.
 - **Checks:** definition/scale consistency, dimensional/transform consistency
 - **Verdict:** UNCERTAIN; confidence: high; impact: critical

- **Assumptions/inputs:** “Normalized expression values” are sometimes log-transformed ($\log 1p$). Burst detection compares a peak to a baseline using multiplicative factors ('fold').
 - **Notes:** A multiplicative fold criterion is well-defined in linear expression space, but becomes ambiguous if applied directly to $\log 1p$ values (ratios in log space are not interpretable as fold-changes in the original scale without back-transform). The paper does not specify whether fold criteria/ \log_2FC are computed on linear values, on logged values, or with back-transformation. This ambiguity affects the formal correctness of the central selection rule.
4. ✘ **Use of stage labels for rooting vs reported uninformative labels** (Sec. 2.2.3 (p.3–4), Sec. 2.3.3 (p.4), Sec. 3.1 (p.5–6))
- **Claim:** Stage labels guide rooting and define transition windows; but results state `\text{life_cycle_stage}` labels were uninformative and canonical markers were absent.
 - **Checks:** logical consistency of assumptions, definition availability
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** Methods intend to use stage labels/markers for rooting and transition-window definition. Results indicate those labels/markers were not usable.
 - **Notes:** Methods rely on stage labels/marker genes to (i) root the trajectory and (ii) define stage-transition windows, but Results state stage annotation is uninformative and marker genes are absent, blocking those steps as written. The paper does not replace them with a precise alternative procedure, so the derivation/logic for ‘preceding a stage transition’ is not internally supported.
5. △ **Low-expression gene filter definition ambiguity** (Sec. 2.3.1 (p.4); Sec. 3.2 (p.6))
- **Claim:** Low-expression genes are defined by mean (or median) below a percentile threshold (30th percentile used in results).
 - **Checks:** definition consistency, edge-case sanity check
 - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Percentile is computed over genes that are 'expressed' (nonzero in at least some cells).
 - **Notes:** The reported use of the 30th percentile is consistent, but the statistic is not fixed (mean vs median) and 'expressed genes' set is not formally defined (e.g., after filtering by min cells). Without these, the filter is not mathematically reproducible.

6. \triangle **Peak metrics (prominence/height) not defined** (Sec. 2.3.2 (p.4); Tables 2–3 (p.7))

- **Claim:** Candidates are ranked by prominence of detected burst peak; tables report 'Peak Prominence' and 'Peak Height'.
- **Checks:** notation/definition completeness, internal interpretability
- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** Peak prominence and height correspond to standard peak-detection outputs, computed from the smoothed profile.
- **Notes:** No mathematical definition is given for 'peak prominence' or 'peak height' (nor whether they are computed on the smoothed profile, detrended profile, or normalized profile). This prevents checking whether reported rankings match stated criteria or whether the two metrics are consistent.

7. \triangle **Smoothing operator along pseudotime** (Sec. 2.3.2 (p.4))

- **Claim:** Expression profiles are smoothed with a rolling mean or LOESS to mitigate noise before peak detection.
- **Checks:** definition completeness, edge-case considerations
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** Cells are ordered along pseudotime and smoothing is performed with respect to pseudotime.
- **Notes:** Rolling mean requires a window definition (in cell index space or pseudotime-width) and handling of irregular pseudotime spacing; LOESS requires span/bandwidth. These are not specified, so the operator is not mathematically defined well enough to audit downstream peak logic.

8. \triangle **Definition of 'precede a stage change' windowing** (Sec. 2.3.3 (p.4))

- **Claim:** A burst precedes a stage change if the burst peak occurs immediately prior to the transition window of the subsequent stage.
- **Checks:** logical definition precision
- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** Transition windows are defined on pseudotime. 'Immediately prior' corresponds to a specified interval.
- **Notes:** Neither 'transition window' nor 'immediately prior' is parameterized (no interval width, no rule for boundaries). Even if stage labels were available, the condition is not mathematically precise.

9. \triangle **Downstream module identification via DE vs lagged correlation** (Sec. 2.4.2 (p.5); Sec. 3.3 (p.8))

- **Claim:** Targets are genes significantly increased post-burst vs baseline, or positively correlated with a lag to the regulator profile.
- **Checks:** definition completeness, consistency of selection rule

- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** A statistical significance test exists for DE and/or for lagged correlation. A lag parameter is optimized or selected.
- **Notes:** No explicit test statistic, null hypothesis, effect-size definition, or multiple-testing correction is given; lagged correlation is not defined (Pearson/Spearman, on smoothed vs raw, how lag grid is chosen). Therefore the module membership rule cannot be audited symbolically.

10. \triangle **Jaccard index mentioned without definition** (Sec. 2.5.3 (p.5))

- **Claim:** Module overlaps are quantified using the Jaccard index.
- **Checks:** notation/definition completeness
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** Modules are treated as sets of gene identifiers.
- **Notes:** The Jaccard index formula is not stated, and it is unclear whether it is computed on raw modules, thresholded modules, or after any filtering. While standard, the audit is constrained to the paper text; the missing definition blocks a strict internal check.

11. \triangle **Statistical overlap tests and multiple-comparison correction** (Sec. 2.5.4 (p.5))

- **Claim:** Fisher’s exact/hypergeometric tests assess overlap significance; p -values are adjusted for multiple comparisons.
- **Checks:** assumption completeness, set-theoretic consistency
- **Verdict:** UNCERTAIN; confidence: high; impact: minor
- **Assumptions/inputs:** A universe/background gene set is defined for hypergeometric/Fisher tests.
- **Notes:** Hypergeometric/Fisher overlap tests require an explicit background size (gene universe) and contingency table construction; neither is defined. Multiple-comparison adjustment method is not specified. This is more about analytic completeness than algebraic correctness.

Limitations

- The provided paper text contains essentially no explicit equations or step-by-step derivations to audit; most quantitative procedures are described verbally.
- Several central method components (smoothing, peak detection metrics, statistical tests) are not mathematically defined in the document, forcing many items to be marked UNCERTAIN.
- This audit does not evaluate numerical correctness of tables/figures or whether the reported candidate sets truly follow from the described pipeline; it only checks internal logical/analytic consistency of the written mathematics and definitions.

Numerical results audit

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

Six numerical consistency checks were executed and all passed (cell-count identity in Table 1; three Lab-vs-Field directional comparisons in Table 1; internal set-count consistency in Table 4; and an within-row equality in Table 2). No computation errors were reported.

Checked items

1. ✓ **C1_total_cells_sum** (Results §3.1 (page 5), Table 1)
 - **Claim:** A total of **43,386** high-quality cells were retained, stratified into **36,520** Lab cells and **6,866** Field cells.
 - **Checks:** parts_to_total
 - **Verdict:** PASS
 - **Notes:** Checked Lab + Field equals Total.
2. ✓ **C2_table1_mean_diff_detected_genes** (Results §3.1 (page 6), Table 1)
 - **Claim:** Lab mean detected genes per cell is slightly higher than Field: **1042.88** vs **1037.13**.
 - **Checks:** inequality_direction
 - **Verdict:** PASS
 - **Notes:** Checked strict $a > b$ using provided ordering; computed difference (Lab - Field) = **5.75**.
3. ✓ **C3_table1_sd_lower_lab_detected_genes** (Results §3.1 (page 6), Table 1)
 - **Claim:** The standard deviation for detected genes per cell was lower in Lab than Field: **436.93** vs **678.6**.
 - **Checks:** inequality_direction
 - **Verdict:** PASS
 - **Notes:** Checked strict $a < b$ using provided ordering; computed difference (Lab - Field) = **-241.67**.
4. ✓ **C4_table1_sd_lower_lab_total_expr** (Results §3.1 (page 6), Table 1)
 - **Claim:** The standard deviation for total log-normalized expression per cell was lower in Lab than Field: **346.37** vs **421.67**.
 - **Checks:** inequality_direction
 - **Verdict:** PASS
 - **Notes:** Checked strict $a < b$ using provided ordering; computed difference (Lab - Field) = **-75.30**.
5. ✓ **C5_table4_internal_consistency_counts** (Results §3.4 (page 8), Table 4)

- **Claim:** Top 100 Lab regulators = 100, Top 100 Field regulators = 100, Shared regulators = 0, Unique to Lab = 100, Unique to Field = 100.
 - **Checks:** set_count_consistency
 - **Verdict:** PASS
 - **Notes:** Checked unique counts equal TopN - Shared for lab and field, plus basic shared \leq TopN sanity.
6. ✓ **C6_table2_peak_height_equals_prominence_row4** (Results §3.2 (page 7), Table 2)
- **Claim:** In Table 2, PF3D7-0414900 has Peak Prominence 0.7277 and Peak Height 0.7277 (identical).
 - **Checks:** equality_within_row
 - **Verdict:** PASS
 - **Notes:** Checked equality within tolerance.

Limitations

- Only the provided PDF text was used; underlying scRNA-seq matrices, pseudotime values, smoothing outputs, and regulator/module gene lists are not included, preventing recomputation of most analysis-derived quantities.
- No plot digitization was performed; figure-based numeric validation (from curves/axes) is out of scope.
- Many numerical claims are qualitative/approximate (e.g., 'over forty-three thousand', 'around 0.67', 'e.g., $\{>\}24$ '), limiting strict numeric verification without raw results tables.