

Skeptical review: Divergent Transcriptional Programs and Regulatory Networks Govern Plasmodium falciparum Development in Laboratory-Adapted Strains and Field Isolates

Summary

This manuscript compares *Plasmodium falciparum* scRNA-seq profiles from lab-adapted parasites (37,624 cells) and Malian field isolates (8,067 cells) to assess how laboratory adaptation affects intraerythrocytic asexual development and gametocytogenesis (Sec. 2; Sec. 3.1). The authors apply a standard single-cell workflow (QC, normalization/log transform, HVG selection, PCA/UMAP), stage-stratified differential expression, trajectory reconstruction (PAGA/diffusion pseudotime are shown in Results; Monocle 3 is also mentioned in Methods), pseudotime-smoothed expression modeling (GAM), co-expression module identification, and a heuristic pseudotime peak-based procedure to nominate “candidate master regulators” and infer putative regulator–target links (Sec. 2.3–2.10; Sec. 3.3–3.6).

The dataset is valuable and the motivating question is important: field isolates plausibly capture biological programs not present in long-term culture. The Results emphasize strong transcriptional divergence between lab and field parasites, especially in sexual stages; field-specific presence of late-stage gametocytes; and differences in inferred regulatory connectivity (more links in lab asexual early development and more links in field sexual development), which are interpreted as developmental and regulatory “rewiring” (Sec. 3.3–3.6; Conclusions Sec. 4.1).

However, several central conclusions are currently difficult to interpret robustly because the lab and field datasets differ strongly in stage composition and coverage (including missing field asexual stages and lab under-representation of late gametocytes), the field data come from only four donors (risk of pseudo-replication), and integration/batch handling and statistical modeling choices are under-specified. In addition, the trajectory and regulatory-network inference descriptions are internally inconsistent across Methods vs Results and lack sufficient statistical definition, normalization, and validation/negative controls to support strong causal language. Addressing these issues—primarily through donor-aware and stage-matched analyses, explicit integration/QC reporting, coherent and parameter-complete trajectory definitions, and more rigorous network benchmarking—would substantially strengthen the paper’s credibility and improve the “big picture” claims about lab adaptation.

Strengths

- Important and timely biological question with clear implications: how well lab-adapted *P. falciparum* reflects natural infections (Introduction; Conclusions Sec. 4.1).

- Large scRNA-seq compendium (45,691 cells) including both lab strains and genetically diverse field isolates, enabling high-resolution stage-stratified analyses (Sec. 2; Sec. 3.1).
- Conceptually rich end-to-end analysis (QC → embedding → DE → trajectory → pseudotime modeling → modules → putative regulatory relationships) that generates concrete hypotheses for follow-up (Sec. 2.3–2.10; Sec. 3.3–3.6).
- Clear qualitative evidence that field data include later gametocyte states than the presented lab dataset and that sexual trajectories differ in extent/topology (Sec. 3.1; Sec. 3.4).
- Figures convey major biological structure and are generally aligned with the narrative, and reported cell-count totals appear internally consistent (Sec. 3.1; Table 2).

Major issues

1. **Strong confounding from dataset composition (stage coverage/mixtures), donor structure (only four field donors), and potential batch/processing effects is not adequately controlled, yet conclusions are framed as broad consequences of “laboratory adaptation” (Sec. 3.1–3.6; Conclusions Sec. 4.1).** Field asexual coverage is incomplete (e.g., missing early rings/schizonts) and field samples are enriched for sexual/late gametocyte stages, while lab samples span the asexual cycle but may not cover late gametocytogenesis (Sec. 3.1; Table 2). Treating cells as independent replicates risks pseudo-replication and inflated significance, and technical differences (library chemistry/batches) could drive global shifts (e.g., “skew toward upregulation in field isolates”).

Recommendation: Across Sec. 3.3–3.6, restrict direct lab-vs-field comparisons to clearly shared and matched segments of development (same stages and, ideally, matched substates along within-stage pseudotime). Re-run key analyses with donor-aware inference: (i) pseudobulk DE per donor×stage (field) and per strain/batch×stage (lab, if available), using edgeR/DESeq2/limma-voom (or muscat), and report consistency across MSC donors; (ii) sensitivity analyses via subsampling/downsampling lab cells to match field cell counts and stage distributions before trajectory/network reconstruction; (iii) explicitly report whether lab and field were processed/sequenced together and, if not, include batch correction/integration (e.g., scVI/Harmony/BBKNN/Scanorama) and show that core findings persist. Add an explicit limitations paragraph in Sec. 3.6 or Conclusions Sec. 4.1 stating what can and cannot be generalized given missing field asexual stages, donor count, and compositional imbalance.

2. **Batch correction/integration, QC, and covariate handling are insufficiently described to assess whether observed lab–field differences are biological vs technical (Sec. 2.1–2.3; Sec. 2.6; Sec. 3.1–3.3).** Critical details are missing/frag-

mented: filtering thresholds (min/max genes per cell, other QC metrics), doublet/ambient-RNA handling, normalization units, HVG selection strategy, and whether integration across strains/patients/batches was performed and evaluated.

Recommendation: Add a dedicated Methods subsection (e.g., Sec. 2.3.x) listing exact QC thresholds and tools: min/max genes per cell, UMI/count thresholds, any mitochondria/rRNA metrics used (or justify why not), doublet detection method and cutoffs, ambient RNA mitigation (if any), normalization target (e.g., counts per 10,000), and the exact data layer used thereafter. If integration/batch correction is used, specify tool and parameters and provide quantitative diagnostics (batch mixing, variance explained). If not used, provide evidence that batch effects are small relative to stage/biology (e.g., embeddings colored by batch/donor; stage composition per batch) and temper claims accordingly (Sec. 3.1–3.3).

- 3. Differential expression (including the reported strong skew toward upregulation in field isolates) is not statistically calibrated for donor/strain structure and technical covariates, risking inflated significance and misinterpretation (Sec. 2.5; Sec. 3.3).** Stage labels are coarse; within-stage maturation differences between in vivo and in vitro could dominate “within-stage” DE, and the manuscript does not demonstrate that lab and field cells represent comparable substates within each broad stage category.

Recommendation: In Sec. 2.5, state the exact DE method used (e.g., Scanpy `rank_genes_groups` settings, test type, correction) and which matrix (log1p vs scaled) was used. Re-run the headline DE results using pseudobulk (donor/strain as the replicate unit) and include covariates (library size; detection rate; batch) as appropriate. In Sec. 3.3, add diagnostics per stage: distributions of QC metrics (UMIs/genes detected), detection fractions, and within-stage pseudotime/subcluster composition by condition. Where within-stage heterogeneity is large (notably gametocytes), perform within-stage subclustering and compare matched substates (or compare along a shared within-stage pseudotime axis) rather than only coarse labels.

- 4. Trajectory inference and pseudotime construction are described inconsistently (Monocle 3 in Methods vs PAGA/diffusion pseudotime in Results) and lack key parameterization (root choice, neighbor graph parameters, clustering resolution, joint vs separate inference), creating uncertainty about all downstream GAM/module/regulatory timing analyses (Sec. 2.6–2.7 vs Sec. 3.4).**

Recommendation: Reconcile Methods and Results by explicitly stating, for each trajectory (asexual lab, asexual field, sexual lab, sexual field): (i) the method used to build the graph (PAGA/Monocle 3/other), (ii) how pseudotime was assigned (DPT/Monocle pseudotime), (iii) how the root was selected (root cluster definition and rationale), and (iv) the key parameters (kNN neighbors, resolution, filtering/subsetting). If multiple methods were tried, report concordance (e.g., correlation of pseu-

dotime orderings; stability of branchpoints) and use one consistent pseudotime variable for all timing claims (Sec. 2.7; Sec. 3.4; Sec. 2.9–2.10). Restrict asexual “progression dynamics” claims to the segments actually present in field data (Sec. 3.4; Sec. 3.6).

5. **Regulatory network / “master regulator” inference is under-specified, not normalized for differing trajectory lengths/search spaces, and insufficiently validated; yet the manuscript uses strong language (“rewiring,” “master regulators”) based largely on link counts and heuristic precedence rules (Sec. 2.8–2.10; Sec. 3.5–3.6; Conclusions Sec. 4.1).** Multiple-testing control, null/negative controls, and robustness to thresholding/binning are not presented, and link-count comparisons are not directly comparable across conditions without normalization.

Recommendation: Expand Sec. 2.8–2.10 to provide a precise, reproducible definition of: (i) the input representation used for modules and regulator inference (e.g., GAM-fitted curves sampled on B pseudotime bins), (ii) binning (B , boundaries, equal-width vs equal-mass), smoothing parameters, and peak-calling criteria on a clearly defined scale (see also log/2-fold issue below), (iii) the statistical test used to call regulator→target links (including time windowing/lag definition), and (iv) multiple-testing correction (FDR) and counts passing each filter. In Sec. 3.5, normalize network comparisons (edges per regulator; per module; per target; per pseudotime length) and add robustness analyses (vary thresholds/bins; subsample to matched cell counts). Add negative controls (permuted pseudotime; shuffled peaks/targets) to estimate expected edge counts under null. Benchmark against known regulators (ApiAP2 family; AP2-G/AP2-FG and other established gametocyte regulators) and, if feasible, motif enrichment in promoters of inferred targets. If rigorous validation is not feasible, reframe networks as hypothesis-generating co-expression/timing associations and soften causal terminology in Sec. 3.6 and Conclusions Sec. 4.1.

6. **Interpretation of sexual-stage findings needs clearer separation of biology vs sampling/annotation: (i) “late-stage gametocytes absent from lab strains” may reflect the lab dataset design (timepoints/induction conditions) rather than an intrinsic absence in lab-adapted parasites; (ii) reported co-expression of male/female markers and conclusions about relaxed lineage separation may be confounded by misannotation, doublets, or ambient RNA (Sec. 3.3.3; Sec. 3.4.2; Sec. 3.6).** The unstructured report also flags a likely marker interpretation/mapping error: Pfs25 is typically a female/sexual-stage marker rather than male-specific.

Recommendation: In Methods (Sec. 2.1–2.2) and Results (Sec. 3.1; Sec. 3.4), describe lab gametocyte induction and sampling (strains used; induction protocol; duration/timepoints) and explicitly rephrase the claim to “not captured in this lab dataset” unless you can show that late stages were expected but absent. Validate late-stage gametocyte identity with a panel of maturation markers and show monotonic

trends along sexual pseudotime (Sec. 3.4). For sex annotations, verify gene ID \leftrightarrow name mapping and marker sex-specificity (including correcting any Pfs25 misstatement), compute quantitative male/female scores per cell using multiple canonical markers, report fractions of ambiguous/co-expressing cells per donor, and apply doublet/ambient-RNA checks targeted at mixed marker expression (Sec. 2.4; Sec. 3.3.3). Present alternative technical explanations alongside biological hypotheses and temper conclusions in Sec. 3.6/Conclusions unless reproducible across donors and markers.

Minor issues

1. Figure set is difficult to audit in places due to missing/placeholder references (e.g., “Figure ??”), ambiguous terminology (e.g., “FA plot”), inconsistent legends/abbreviations, and limited reporting of sample sizes, thresholds, and uncertainty in statistical panels (Sec. 3.3–3.5; multiple figures including 1–4, 5–6, 8, 12–17, 18–21). Overplotting/small fonts and non-colorblind-safe palettes reduce accessibility.

Recommendation: Resolve all “Figure ??” placeholders and define terms like “FA plot” consistently (caption + main text). Standardize figure legends to include: color mappings (stage/module/condition), n cells per panel (and n donors/strains where relevant), statistical thresholds (FDR/log₂FC), and directionality/progression indicators for pseudotime (arrows/gradients). Improve readability (vector export, larger fonts, alpha blending) and adopt colorblind-safe palettes. For regulatory/trajectory claims, add uncertainty visualization (e.g., bootstrapped confidence bands for smoothed curves; replicate-level summaries).

2. Module detection and functional enrichment are mentioned but not presented with sufficient detail to evaluate module robustness, biological interpretation, and cross-condition correspondence (Sec. 2.8–2.8.2; Sec. 3.5).

Recommendation: Provide module summary outputs (main or Supplement): number of modules, size distributions, top genes per module, module expression (eigengene/mean curve) vs pseudotime, and how “activation” is defined. For enrichment (gprofiler2/PlasmoDB), specify background gene set, databases/ontologies, and FDR correction, and include full enrichment tables with representative terms/FDRs referenced in Sec. 3.5.

3. Statistical reporting is inconsistent/vague across analyses (e.g., “p-values approaching zero”), and thresholds are not consolidated (Sec. 2.5; Sec. 2.8–2.10; Sec. 3.3–3.5).

Recommendation: Add a compact “Analysis thresholds” table (main or Supplement) listing, for each analysis type (DE, enrichment, peak calling, edge calling), the test used, FDR cutoff, and any effect-size filters. Replace qualitative phrases with explicit counts/ranges (e.g., number of significant genes per stage; min/median adjusted p -values).

4. Dataset composition and replication structure are not summarized in one place, complicating interpretation of which comparisons are supported by how many donors/strains/timepoints (Sec. 2.2; Sec. 2.4; Sec. 3.1; Table 2).

Recommendation: Add a comprehensive table listing, for each stage/sex label, counts per field donor (MSC1/3/13/14) and per lab strain/culture (including independent cultures/timepoints if applicable). Reference this table explicitly in Sec. 3.1 when discussing coverage and limitations.

5. HVG selection and scaling choices may bias downstream embeddings/trajectories toward lab–field differences rather than within-stage biology, especially under strong compositional imbalance; it is also unclear which data layer (scaled vs log) is used for each downstream analysis (Sec. 2.3.1; Sec. 2.5; Sec. 2.7–2.10).

Recommendation: Clarify, for each analysis block (PCA/UMAP, DE, GAM smoothing, module clustering, link inference), the exact matrix used (normalized, log1p, scaled). Consider HVG selection strategies that reduce confounding (e.g., HVGs within stage, or integration-aware HVGs) and report sensitivity of main qualitative conclusions to HVG choice.

6. Data/code availability and study metadata are not clearly linked in the text (unstructured report notes GitHub is mentioned without link/accessions), limiting reproducibility (Methods/Supplement references).

Recommendation: Provide explicit dataset accession IDs (raw + processed matrices), a permanent code repository link (commit/tag), environment/package versions, and a minimal “run order” for reproducing figures. If some data are restricted (human subjects), specify access procedure.

Very minor issues

1. Internal ambiguity about whether log transform is applied once or twice: Sec. 2.1.3 states $\log(X + 1)$; Sec. 2.3.1 again states log1p on normalized data. This also affects interpretability of “fold change” and peak thresholds (Sec. 2.9.2).

Recommendation: State explicitly the exact transformation pipeline and ensure log is applied once. For every downstream analysis, specify the data layer used. Ensure any “fold change”/“2-fold” criteria are computed on an appropriate (linear) scale or rewritten correctly for log scale.

2. Peak-calling criterion “> 2-fold increase over its own median expression” is undefined on log-transformed values and pseudotime binning is not formally specified (Sec. 2.9.2; Sec. 3.5.1).

Recommendation: Define bin count and binning rule and specify whether fold-change is computed after back-transform to linear space. If remaining on log1p, use an additive threshold consistent with the scale.

3. A stated “four-fold increase” in links for Sexual Field vs Sexual Lab (1,917 vs 404) does not match the ratio (~ 4.75) under the reported counts and the manuscript’s stated tolerance (Sec. 3.5).

Recommendation: Correct the numeric claim (or clarify an alternative definition) so text matches the reported counts.

4. Numerous typographical/OCR/LaTeX artifacts, inconsistent heading formatting, and corrupted strings reduce readability (e.g., broken words in Introduction Sec. 1; garbled file/function text in Sec. 2.1.1; truncated table headers; inconsistent species/gene formatting).

Recommendation: Thoroughly proofread and clean OCR/LaTeX artifacts; standardize headings; italicize *P. falciparum* consistently; standardize gene ID formatting (PF3D7_XXXXX); ensure tables render cleanly.

5. Non-biological/incorrect content appears in front matter (astronomy-related keywords; unstructured report also notes placeholder-like authorship/affiliation text).

Recommendation: Remove/replace incorrect keywords with malaria/scRNA-seq terms and delete any placeholder authorship/affiliation content to match journal standards.

6. Ethics/informed-consent statement for human-derived field isolates is not clearly present in the main text (Methods/Conclusions).

Recommendation: Add an explicit Ethics/Human Subjects statement in or near Methods, including IRB approvals and consent procedures, or cite where it appears in Supplement.

Mathematical consistency audit

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

Maths relevance: light

The paper is primarily methodological/biological and contains very little explicit mathematics or derivations (no numbered equations). The key analytic content consists of data transforms (log_{1p}), percentile thresholds, differential-expression effect sizes described as log-fold changes, pseudotime-based modeling (GAM smoothing), and heuristic definitions for peak detection and regulator-target timing. The main internal-consistency problems are definitional: unclear or conflicting statements about which transforms/trajectory algorithms were used and on which scale fold-change/peaks are computed.

Checked items

1. ✓ **Log transform stated in QC step** (Sec. 2.1.3, p.2)

- **Claim:** A log-transformation $\log(X + 1)$ is applied to normalized expression values.
- **Checks:** definition consistency, notation
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** X denotes normalized expression values (nonnegative).
- **Notes:** The transform $\log(X + 1)$ is mathematically well-defined for $X \geq 0$. No base specified, but that is not internally inconsistent at this point.

2. ✘ Second log transform statement (possible double-logging) (Sec. 2.3.1, p.3)

- **Claim:** A log-transformation (log1p) was applied to the normalized expression data.
- **Checks:** definition consistency, pipeline consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** The same normalized matrix as in Sec. 2.1.3 is being referenced.
- **Notes:** This duplicates Sec. 2.1.3 and creates ambiguity: either the log transform is applied twice ($\log(1 + \log(1 + X))$) or only once. This affects interpretability of downstream ‘log-fold change’ and ‘2-fold peak’ criteria.

3. ⚠ Scaling step vs later effect-size language (Sec. 2.3.1 (scaling), p.3; Sec. 3.3 (LFC), p.6-7)

- **Claim:** Data are scaled to zero mean/unit variance per gene; later results report log-fold change (LFC) between conditions.
- **Checks:** units/scale consistency, definition consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** Scaled values may be used for some downstream tasks (unspecified).
- **Notes:** If LFC is computed on scaled data, the term ‘fold change’ becomes conceptually inconsistent (dimensionless ratio vs standardized units). The paper does not specify which data layer LFC uses, so internal consistency cannot be verified.

4. ⚠ DGE test + LFC definition gap (Sec. 2.5.1, p.3; Sec. 3.3, p.6-7)

- **Claim:** Wilcoxon rank-sum test is used with BH-FDR correction; results report genes with LFC values.
- **Checks:** symbol/definition consistency, missing definitions
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** LFC is some log ratio/difference summary between groups.

- **Notes:** While Wilcoxon + BH-FDR is logically compatible, the paper never defines LFC mathematically (log base; mean/median; linear vs log scale). This blocks analytic checking of interpretability of reported ‘LFC=10.44’ style statements.
5. ✘ **Trajectory method mismatch (Monocle 3 vs PAGA/diffusion pseudotime)** (Sec. 2.6.1-2.6.2, p.3; Sec. 3.4, p.8)
- **Claim:** Trajectories are inferred with Monocle 3 (Methods), but Results state PAGA and diffusion pseudotime were used and show PAGA graphs.
 - **Checks:** method/notation consistency, dependency consistency
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** Only one pseudotime variable is used for downstream GAM/peak/regulatory inference.
 - **Notes:** The pipeline is internally inconsistent: two different trajectory/pseudotime constructions are claimed. Downstream analyses depend on pseudotime ordering, so the mathematical object ‘pseudotime’ is not uniquely defined.
6. △ **GAM smoothing along pseudotime (model unspecified)** (Sec. 2.7.2, p.3)
- **Claim:** Gene expression is modeled as a function of pseudotime using GAMs.
 - **Checks:** missing derivation/specification
 - **Verdict:** UNCERTAIN; confidence: low; impact: minor
 - **Assumptions/inputs:** There exists a response variable (expression) and predictor (pseudotime).
 - **Notes:** No explicit GAM form (link function, error distribution, smoothing basis) is given, so there is no algebra to verify. This is acceptable descriptively but prevents an analytic consistency check of what ‘modeled’ precisely means.
7. ✘ **Peak definition uses '2-fold increase' under log transform** (Sec. 2.9.2, p.4)
- **Claim:** A peak is a local maximum with expression > 2 -fold over the gene’s median expression within that trajectory.
 - **Checks:** scale consistency, definition consistency
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** Expression values used for peak detection are log-transformed per Sec. 2.1.3/2.3.1 unless otherwise stated.
 - **Notes:** On a log scale, ‘2-fold’ is not represented by multiplying the logged values by 2. The paper does not specify whether fold changes are computed after back-transforming. This makes the peak criterion mathematically ambiguous/inconsistent.

8. Δ **Definition of 'immediately prior' as 10% pseudotime window** (Sec. 2.9.3, p.4)

- **Claim:** 'Immediately prior' is defined as a window in pseudotime (e.g., 10% of the total pseudotime range).
- **Checks:** definition consistency, sanity/edge-case
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** Pseudotime is scaled to a fixed range (implicitly comparable across trajectories).
- **Notes:** The definition is mathematically clear if pseudotime is normalized to a common range, but the paper does not state the pseudotime scaling (e.g., $[0, 1]$) nor how ranges compare across lab vs field trajectories. Without this, the 10% rule may not be comparable across conditions.

9. Δ **Before/after peak testing for targets (criterion unspecified)** (Sec. 2.10.1, p.4)

- **Claim:** Genes significantly changing after a regulator's peak are considered targets, using lagged correlation or before/after testing in pseudotime windows.
- **Checks:** missing definitions, logical consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** There is a defined peak time t^* and windows $(t^* - \Delta, t^*)$ and $(t^*, t^* + \Delta)$.
- **Notes:** The approach can be mathematically coherent, but the paper omits the exact statistical definition of 'significantly changed' and the windowing rule relative to t^* . This prevents verifying internal consistency of the inferred 'putative links' as a well-defined mathematical procedure.

Limitations

- The provided PDF text contains essentially no explicit equations/derivations; the audit is limited to checking consistency of mathematical definitions, transforms, and algorithmic dependencies described in prose.
- Several key steps (exact data layer used for DGE, LFC computation, GAM specification, peak-calling computation scale) are not formally defined, forcing multiple items to be marked UNCERTAIN where verification would require missing details.
- No symbolic unit/dimensional analysis is possible because quantities are largely dimensionless (normalized/log expression) and units are not specified.

Numerical results audit

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

Of 14 numeric checks performed, 13 PASS and 1 FAIL. Passes include exact agreement for multiple cell-count totals and Table 2 row/column sums, plus percentage rounding consistency for lab/field composition. The sole failure is the “four-fold increase” claim for Sexual Field vs Sexual Lab links, where the computed ratio is ~ 4.75 rather than 4.

Checked items

1. ✓ **C01_cells_total_sum** (p.4 Results §3.1 (paragraph) and Table 1)
 - **Claim:** Total cells reported as 45,691, composed of 37,624 lab cells and 8,067 field cells.
 - **Checks:** parts_vs_total
 - **Verdict:** PASS
 - **Notes:** $37,624 + 8,067 = 45,691$ (exact).
2. ✓ **C02_cells_percentages** (p.4 Results §3.1 (first paragraph))
 - **Claim:** Lab cells are 82.3% and field cells are 17.7% of 45,691 total cells.
 - **Checks:** percent_of_total_and_sum_to_100
 - **Verdict:** PASS
 - **Notes:** Recomputed: lab 82.3444% and field 17.6556%; both match reported values within 0.1% rounding; reported percentages sum to 100.0.
3. ✓ **C03_table2_field_row_sum** (p.5 Table 2 (Field row))
 - **Claim:** Field row stage counts sum to the Field total of 8,067.
 - **Checks:** row_sum_equals_total
 - **Verdict:** PASS
 - **Notes:** Field stage counts sum to 8,067 (exact).
4. ✓ **C04_table2_lab_row_sum** (p.5 Table 2 (Lab row))
 - **Claim:** Lab row stage counts sum to the Lab total of 37,624.
 - **Checks:** row_sum_equals_total
 - **Verdict:** PASS
 - **Notes:** Lab stage counts sum to 37,624 (exact).
5. ✓ **C05_table2_total_row_sum** (p.5 Table 2 (Total row))
 - **Claim:** Total row stage counts sum to the Total of 45,691.
 - **Checks:** row_sum_equals_total
 - **Verdict:** PASS
 - **Notes:** Total-row stage counts sum to 45,691 (exact).
6. ✓ **C06_table2_column_consistency_early_trophozoite** (p.5 Table 2 (early trophozoite column))

- **Claim:** For early trophozoite, Total (9,757) equals Field (122) + Lab (9,635).
 - **Checks:** column_sum_equals_total
 - **Verdict:** PASS
 - **Notes:** $122 + 9,635 = 9,757$ (exact).
7. ✓ **C07_table2_column_consistency_late_ring** (p.5 Table 2 (late ring column) and p.6 §3.3.1)
- **Claim:** Late ring counts: Field 428 and Lab 5,438; Total 5,866. Text also states late ring comparison used 428 field vs 5,438 lab cells.
 - **Checks:** cross_reference_and_column_sum
 - **Verdict:** PASS
 - **Notes:** $428 + 5,438 = 5,866$ and the text matches the table counts (exact).
8. ✓ **C08_table2_column_consistency_female_gametocyte** (p.5 Table 2 (gametocyte (female) column) and p.7 §3.3.3)
- **Claim:** Female gametocyte counts: Field 1,656 and Lab 3,903; Total 5,559. Text states female gametocytes comparison used 1,656 field vs 3,903 lab cells.
 - **Checks:** cross_reference_and_column_sum
 - **Verdict:** PASS
 - **Notes:** $1,656 + 3,903 = 5,559$ and the text matches the table counts (exact).
9. ✓ **C09_table2_column_consistency_male_gametocyte** (p.5 Table 2 (gametocyte (male) column) and p.7 §3.3.3)
- **Claim:** Male gametocyte counts: Field 3,364 and Lab 1,964; Total 5,328. Text states male gametocytes comparison used 3,364 field vs 1,964 lab cells.
 - **Checks:** cross_reference_and_column_sum
 - **Verdict:** PASS
 - **Notes:** $3,364 + 1,964 = 5,328$ and the text matches the table counts (exact).
10. ✓ **C10_table2_column_consistency_grand_total** (p.5 Table 2 (Totals) and p.4 §3.1 (cell totals))
- **Claim:** Grand totals are consistent: Table 2 Total=45,691 equals Table 1 Overall Number of Cells=45,691 equals p.4 narrative total 45,691.
 - **Checks:** repeated_constant_match
 - **Verdict:** PASS
 - **Notes:** All three totals match (45,691) exactly.

11. ✓ **C11_table1_overall_median_norm_expr_weighted_range** (p.5 Table 1)
 - **Claim:** Overall median normalized expression per cell is **2058.59**, while lab median is **2093.79** and field median is **1813.51**; overall median should lie between subgroup medians.
 - **Checks:** range_sanity_check
 - **Verdict:** PASS
 - **Notes:** 2058.59 lies within [1813.51, 2093.79].
12. ✓ **C12_table1_overall_median_genes_per_cell_range** (p.5 Table 1)
 - **Claim:** Overall median genes per cell is **937**, while lab median is **954** and field median is **802**; overall median should lie between subgroup medians.
 - **Checks:** range_sanity_check
 - **Verdict:** PASS
 - **Notes:** 937 lies within [802, 954].
13. ✓ **C13_sparsity_percent_bounds** (p.4 Results §3.1 (first paragraph))
 - **Claim:** Matrix sparsity is reported as **80.25%**.
 - **Checks:** percent_bounds_check
 - **Verdict:** PASS
 - **Notes:** 80.25% is within 0% to 100%.
14. ✗ **C14_four_fold_claim_links** (p.11 Results §3.5.2 (Sexual Trajectories paragraph))
 - **Claim:** Sexual Field has **1,917** putative links vs Sexual Lab **404** links; text claims a 'four-fold increase'.
 - **Checks:** ratio_check
 - **Verdict:** FAIL
 - **Notes:** Computed ratio = $1,917/404 = 4.745$, which exceeds the 15% relative tolerance for a “four-fold” claim.

Limitations

- Checks are restricted to explicit numeric statements in the provided PDF text/tables; underlying data files (geneexpression.csv, labels.csv) are not available here.
- No values are extracted from plot images (e.g., scree plot, volcano plots, pseudotime plots); image-based numeric validation is out of scope per instructions.
- Many methodological thresholds (e.g., HVG selection, PCA elbow, GAM peak definitions) cannot be numerically audited without intermediate outputs.