Organoid Fidelity vs. Reference

Project: Organoid-Data-Analysis
Notebook: nb3_reference
Dataset: GSE75140 (Camp et al. 2015, *Cell*)

Purpose

Organoids are designed to recapitulate aspects of human development, but how faithfully they reproduce in vivo programs is a key question. This notebook demonstrates how to compare cerebral organoid cells to a fetal brain reference.

- **Within organoids:** Do clusters of cells map to expected fetal lineages (radial glia, progenitors, neurons)?
- **Across organoids:** Is the mapping consistent, or do replicates differ in fidelity?
- **Against reference:** Which fetal populations are missing or underrepresented?

All instructions, annotations, interpretations, and saved artifacts are documented directly in the notebook.

Challenges in Fidelity Assessment

1. **Reference quality** — biological interpretation requires curated fetal annotations.
   In this notebook, provisional labels were derived by clustering since curated labels were not available.
2. **Limited gene overlap** — organoid and fetal data often share only a subset of genes, reducing resolution.
3. **Ambiguous mappings** — some organoid states align confidently, others spread across multiple fetal groups.
4. **Replicate variability** — organoids differ in which fetal programs are best recapitulated.

Framework for Analysis

This notebook structures fidelity evaluation into four layers:

1. **Label transfer** — project organoid cells into fetal space, assign predicted reference labels.

Example: organoid clusters aligning with fetal progenitors or neurons.

2. **Correlation analysis** — compute pseudobulk correlations between organoid units and fetal types.
Example: progenitor-like organoid units scoring highest with fetal progenitors.

3. **Agreement checks** — compare transfer vs. correlation outcomes to flag stable vs. ambiguous alignments.

4. **Guardrails** — apply baselines (random shuffles) and low-N exclusions to avoid overinterpretation.

Steps in This Notebook

1. **Setup** — initialize environment, seed, paths, helper functions.

2. **Load Data** — import organoid query and fetal reference, harmonize shared genes.

3. **Reference embedding** — build fetal PCA/UMAP space as the projection target.

4. **Label transfer** — map organoid cells into fetal manifold and predict reference labels.

5. **QC metrics** — assess confidence, per-organoid alignment rates, and coverage.

6. **Pseudobulk correlations** — quantify similarity between organoid clusters and fetal types.

7. **Agreement** — evaluate convergence/divergence of transfer vs correlation.

8. **Baselines & exclusions** — compute chance-level expectations and filter unstable units.

9. **Summary dashboards** — compile scorecards, global figures, and summary tables.

Addressing Organoid Fidelity — Broader Perspective

- **Major lineages captured:** Organoids reproducibly generate radial glia, progenitors, and neurons.

- **Underrepresentation:** Some fetal populations (e.g., basal progenitors) appear reduced.

- **Variability:** Different organoids vary in fidelity; some align robustly, others weakly.

- **Interpretation caution:** Conclusions here rely on computational clusters as proxies for fetal types; curated references (e.g., fetal cortex atlases) are needed for biological precision.

Guardrails

- **Random baselines** show observed alignments are stronger than chance.
- **Low-N thresholds** prevent small organoid units from skewing results.
- **Confidence scores** provide a tunable filter for stable assignments.

Biological Context

Organoid fidelity is not all-or-nothing. Organoids can recapitulate major neurodevelopmental trajectories but often incompletely. Differences may reflect biological stochasticity (e.g., progenitor-to-neuron balance) and technical constraints (sampling depth, gene coverage).

Camp et al. (2015) showed cerebral organoids recover early cortical lineages but underrepresent basal progenitors compared to fetal cortex. This notebook's framework reproduces such findings and offers a template for evaluating other organoid systems.

How This Notebook Helps

- Provides a **stepwise, reproducible framework** to benchmark organoid fidelity.
- Embeds **guardrails** (baselines, exclusions) to contextualize alignment strength.
- Offers a **tutorial template** for applying the same logic to other references (fetal, adult, disease).

Next Steps Beyond This Notebook

- **Curated fetal references**: Use published fetal atlases with annotated cell types (e.g., Nowakowski et al. 2017) for lineage-specific fidelity.
- **Cross-reference multiple atlases**: Benchmark organoids against several fetal datasets to ensure robustness.
- **Temporal fidelity**: Assess whether organoids recapitulate stage-specific developmental programs (early vs. late progenitors, neuronal maturation).
- **Functional benchmarking**: Extend beyond transcriptomics, adding scATAC-seq or

spatial transcriptomics to test whether organoids capture both molecular identity and tissue architecture.

- **Perturbation studies**: Compare fidelity across protocols, genetic backgrounds, or treatments to test how external factors shape lineage fidelity.
- **Expanded guardrails**: Add more sophisticated baselines (e.g., gene module permutations, bootstrap resampling) to strengthen statistical rigor.

Key Takeaways

- Organoids recapitulate broad fetal lineages, but fidelity is uneven.
- Agreement between label transfer and correlation strengthens confidence.
- Guardrails are essential to contextualize findings and avoid overinterpretation.
- Curated references remain the gold standard for biological interpretation.