Study design (Part 1)

Clinical problem addressed
N/A. This is a basic-science single-cell study on *Xenopus* tail regeneration, not a clinical prediction task.

Research question clearly stated
Yes. We reproduce and extend ROC-cell analysis and evaluate how two denoising methods and two over-time batch-integration methods affect (1) clustering quality and (2) marker selection.

Training/test cohorts and their characteristics
N/A. Unsupervised analysis; no training–test split is used.

Representativeness of the cohorts for real-world clinical practice
N/A. Non-clinical dataset.

State-of-the-art (SOTA) baselines
Yes. We compare Leiden, Louvain, and K-Means for clustering; "raw" serves as the baseline for denoising (PCA low-rank, kNN smoothing) and over-time batch integration (Harmony, BBKNN).

Data and optimization (Parts 2–3)

Origin of the data and original format
Public single-cell RNA-seq from *Xenopus* tail; stored as AnnData (13199 cells × 31535 genes) with metadata (e.g., DaysPostAmputation, cluster).

Data transformations prior to modeling
Total-count normalization, log1p transform, gene/cell QC (genes detected ≥3; cells with ≥200 genes), HVG selection (2300 genes), scaling, PCA (~50 PCs), kNN graph, UMAP.

Independence between training and test sets
N/A (unsupervised, no split).

Models evaluated and code used to select the best model
Evaluated Leiden/Louvain/K-Means; denoising via PCA low-rank & kNN smoothing; time-batch integration via Harmony & BBKNN. Selection based on internal clustering indices (Silhouette primary; CH & DB secondary) and marker reproducibility. Full code available in a public GitHub repository (xr0706/5243-project-1).

Input data type
Structured: sparse numerical expression matrix with structured cell metadata.

Model performance (Part 4)

Primary metric for algorithm performance (with justification)
Silhouette score (label-agnostic compactness/separation) is primary; Calinski–Harabasz (CH) and Davies–Bouldin (DB) are secondary. Example results: Louvain Silhouette 0.3240, Leiden 0.3048, K-Means 0.2966; CH—K-Means 2916.23, Louvain 2762.10, Leiden 2534.64; DB—Leiden 1.1467, Louvain 1.2187, K-Means 1.3062.

Primary metric for clinical utility (with justification)
N/A (non-clinical task).

Baseline vs proposed comparison and statistical significance
Comparisons presented descriptively: denoising and time-integration variants are contrasted against the raw baseline using Silhouette/CH/DB and marker Jaccard overlap. No hypothesis tests were performed.

Model examination (Part 5)

Examination technique 1
UMAP visualization with multiple cluster labelings (Leiden, Louvain, K-Means) and major lineage overlays to inspect structure and potential batch effects.

Examination technique 2
ROC gene-set scoring and literature cross-validation: ROC-score forms a single high-signal hotspot coinciding with an unsupervised cluster and the biological ROC annotation; three DE methods yield a strong majority-consensus marker panel (84 genes).

Relevance of examination results to performance
Visual cluster compactness/separation aligns with quantitative metrics (e.g., Louvain's higher Silhouette; Leiden's sharper boundaries), confirming behavior seen in the indices.

Feasibility and significance of case-level interpretability
Gene-level markers and gene-set scores support cell-type interpretation at cluster and cell levels; no per-cell saliency method is required for this unsupervised analysis.

Reliability/robustness under data distribution shifts
Addressed via over-time batch integration using DaysPostAmputation: Harmony

improves compactness and separation; BBKNN provides the strongest marker reproducibility versus the raw baseline without over-warping biology.

Reproducibility (Part 6)

Transparency tier
Tier 1: full, executable code and notebooks are publicly available (GitHub: xr0706/5243-project-1), enabling end-to-end reproduction of preprocessing, clustering, denoising/integration, metrics, and figures.