Geometry-Constrained Residual Projection with Optimal Transport for scRNA Cell-State Analysis

Anonymous authors
Paper under double-blind review

Abstract

Observed zeros in single-cell RNA sequencing (scRNA-seq) data couple two problems that are usually treated separately: expression imputation and cell-state progression inference. Imputation seeks to recover expression missed by the assay, whereas optimal transport (OT) infers progression from distributions of cell states. However, imputation changes the same cell-state geometry from which OT infers progression. To limit this change, we introduce *Geometry-Constrained Residual Projection* (GCRP), which treats existing imputers as generators of candidate residuals. GCRP accepts only corrections that are supported by neighboring cells, have a controlled amplitude, and satisfy cell-displacement and population-level OT limits. Per-cell displacement bounds induce Wasserstein bounds on population displacement and on the distortion of pairwise group distances. Candidate selection relies only on agreement with observed positive expression and on OT over clusters of raw expression. GCRP itself uses no biological labels or external measurements, which serve for evaluation. Across controlled-corruption, CITE-seq, disease-progression, and time-course experiments, GCRP and its support gate keep the population geometry used for progression inference. Under controlled corruption, full graph imputation triples the error of the relative state geometry against an uncorrupted reference, whereas GCRP keeps this error at the level of raw data and lowers held-out expression RMSE by 5%. Applied on top of each of five published imputers, GCRP lowers that imputer's distortion of the population geometry 3.7- to 8.2-fold. On CITE-seq cells, scVI and SAVER lower the correlation between RNA and surface-protein levels, and GCRP applied on top of them raises it in every seed, from 0.146 to 0.290 for scVI. At the default radius, the Wasserstein bounds certify that the displacement projection preserves 45% of the raw orderings of group distances. These results show that progression geometry can set how much uncertain expression correction is admitted.

1 Introduction

Single-cell RNA sequencing (scRNA-seq) measures gene expression in individual cells, and many studies use it to infer *cell-state progression*, the movement of cells through developmental, perturbational, or disease states. Optimal transport (OT) methods infer this progression by comparing distributions of cell states. Waddington-OT introduced this approach, and later methods added lineage information, continuous flows, perturbations, multiple modalities, and single snapshots (Schiebinger et al., 2019; Forrow & Schiebinger, 2021; Tong et al., 2020; Bunne et al., 2023; Klein et al., 2025; Yachimura et al., 2024; Tronstad et al., 2025). CellRank provides complementary directed fate mapping (Lange et al., 2022). These methods depend on the distances between cells and between cell populations. However, scRNA-seq data contain many zeros, and some zeros reflect missed molecular capture instead of absent expression. *Imputation* methods estimate which zeros to fill and how much expression to restore, using assumptions about dropout, neighborhoods, low-rank structure, diffusion, uncertainty, graphs, or generative models (Li & Li, 2018; Gong et al., 2018; Huang et al., 2018; van Dijk et al., 2018; Linderman et al., 2022; Eraslan et al., 2019; Arisdakessian et al., 2019; Lopez et al., 2018; Huang et al., 2025; Lee et al., 2024). Foundation models add further representations and reconstructions (Cui et al., 2024; Theodoris et al., 2023). Because imputation changes expression values, it also moves cells and populations in the space where OT measures distances. An imputer can therefore improve expression recovery while changing the population distances that determine

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the inferred progression. Imputation has similar side effects on gene correlations and reconstructed regulatory networks (Ly & Vingron, 2022). Because the two analyses usually run in sequence, this distance change passes directly into progression analysis. We therefore use the progression geometry to limit which imputed corrections are accepted before progression analysis.

We formalize this goal as *geometry-constrained recovery*. Given observed expression X^{obs} and a candidate reconstruction X^{prop}, GCRP treats the residual $R^{\text{prop}} = X^{\text{prop}} - X^{\text{obs}}$ only as a proposal and returns $X^{\text{safe}} = X^{\text{obs}} + \Pi_{\mathcal{M}_{\text{safe}}}(R^{\text{prop}})$, where $\mathcal{M}_{\text{safe}}$ is the set of admissible residuals. The acceptance operator $\Pi_{\mathcal{M}_{\text{safe}}}$ modifies only observed zeros with local support and controlled amplitude, bounds each cell's displacement under fixed reference maps, and rejects candidates that change the OT geometry of raw-expression clusters beyond a declared budget. When every candidate fails, GCRP shrinks one residual uniformly until it passes, which in the extreme case returns X^{obs}. This design separates *proposal generation* from *residual acceptance*: statistical, graph, diffusion, low-rank, generative, or foundation-model imputers can propose corrections, while GCRP limits how far each proposal alters the geometry used for progression inference.

To keep the evaluation independent of this acceptance step, GCRP uses only observed RNA to accept corrections. Agreement with observed nonzero expression weights the candidate proposals, and clusters of raw expression define the OT limit. We hold out cell types, protein abundance, disease states, time points, lineages, and uncorrupted reference data for evaluation. On controlled-corruption, CITE-seq, disease-progression, and time-course datasets, we compare raw expression with full and projected outputs of heterogeneous imputers. Five published imputers, scVI, MAGIC, ALRA, DCA, and SAVER, distort OT geometry, and GCRP applied on top of each lowers this distortion 3.7- to 8.2-fold. Against held-out references, GCRP protects the geometry that full imputers damage. For example, under controlled corruption, full graph imputation triples the error of the relative state geometry, whereas GCRP keeps it at the level of raw data and lowers held-out expression RMSE by 5%. On CITE-seq cells, scVI and SAVER lower the correlation between RNA and surface-protein levels, and GCRP applied on top of them raises it in every seed. Compared with global shrinkage and per-cell caps that set their strength by the same label-free check, GCRP keeps neighborhoods closer to the reference and changes about half as many zeros.

These results also indicate how much recovery to expect on a new dataset. GCRP limits changes relative to the observed geometry, so end-to-end recovery depends on how faithfully that geometry reflects cell states. At the default radius, the Wasserstein bounds certify that the displacement projection preserves 45% of the raw orderings of group distances.

Contributions. We make four contributions: (1) we show that imputation changes the population geometry from which OT infers progression, and define geometry-constrained recovery; (2) we introduce GCRP, a model-agnostic layer that accepts only supported corrections of controlled amplitude, bounds cell displacement in fixed reference maps, and limits OT distortion among raw-expression clusters; (3) we connect per-cell displacement guarantees to Wasserstein bounds on population-geometry distortion and count the observed orderings that these bounds certify; and (4) we show on controlled-corruption, CITE-seq, progression, and time-course data that GCRP recovers expression while keeping the relative state geometry that full imputers distort.

2 Problem Formulation

Observed expression and zero set. Let $X^{\text{obs}} \in \mathbb{R}^{n \times p}_{\geq 0}$ be the observed scRNA-seq expression matrix for n cells and p genes, where row x_i of X^{obs} is the expression vector of cell i. We define an *observed zero* as an entry that equals zero in the measured matrix, whatever its biological or technical cause, and write $Z_0 = \{(i, g) : X^{\text{obs}}_{ig} = 0\}$. We distinguish an observed zero from a *true zero*, which represents a latent absence of expression, and a *missed-capture zero*, which represents positive latent expression that the assay failed to observe.

Candidate residuals. Let C index candidate imputation methods. An imputation method T_m, where $m \in C$, produces:

$$X^{(m)}_{\text{prop}} = T_m(X^{\text{obs}}), \qquad R^{(m)} = X^{(m)}_{\text{prop}} - X^{\text{obs}}.$$

When m is clear, we write X^{prop} and R^{prop}. For consistency, every candidate output X^{prop} is placed on the same scale as X^{obs} before computing the residual. In our experiments, both are counts rescaled per gene to $[0, 1]$ by their nonzero 5th and 95th percentiles (Appendix B). Here R is a *residual proposal*:

R_{ig} represents the proposed numerical change to entry (i, g), and R_i represents the proposed change to cell i. Any imputer can generate a proposal, and GCRP determines which proposed changes are accepted, attenuated, or rejected.

Reference and population geometry. GCRP defines geometric limits at two distinct levels. First, $\mathcal{A} = \{1, \ldots, K\}$ indexes fixed affine reference maps $f_a(x) = A_a x + b_a$ with $A_a \in \mathbb{R}^{d_a \times p}$. A *reference map* is a fixed representation in which we measure imputation-induced cell movement. The radius r_{ia} acts as the *cell-level displacement limit* for cell i under map a. Fixing the reference maps before accepting any residuals prevents the geometry from adapting to the corrections it is supposed to evaluate. Second, $\mathcal{G}$ contains *geometry groups*, which the selector builds as k-means clusters of raw expression. $D_{\mathrm{OT}}^{\mathcal{G}}$ denotes the Sinkhorn or OT discrepancy over these groups, and $\rho_{\mathrm{OT}}^{\mathcal{G}}$ represents the rank correlation between raw and projected group distances. These quantities define a *population-level OT limit* on changes to the relative distances between observed groups.

Geometry-constrained recovery. We determine which components of a candidate residual can be accepted while keeping both cell and population movement within the declared displacement limits. The resulting matrix is $X^{\mathrm{safe}} = X^{\mathrm{obs}} + \Pi_{\mathcal{M}_{\mathrm{safe}}}(R^{\mathrm{prop}})$. Here $\mathcal{M}_{\mathrm{safe}}$ is the admissible residual set defined in Section 4, and $\Pi_{\mathcal{M}_{\mathrm{safe}}}$ maps a proposal into that set. We call this objective *geometry-constrained recovery*. A correction must have local neighborhood support and remain within the declared displacement limits to be admitted.

This constrained target follows from an identifiability limitation. From X^{obs} alone, the latent origin of an observed zero is unidentifiable: the observation $X_{ig}^{\mathrm{obs}} = 0$ is consistent with both a true absence of expression and positive expression lost through incomplete capture. Therefore, we target corrections that preserve downstream geometry, instead of an unconstrained reconstruction of the latent expression matrix.

When every candidate fails the OT limits, GCRP shrinks one residual by a uniform scale found by bisection and returns the result, which equals X^{obs} when only the zero scale passes (Section 4). The GCRP selector uses only observed RNA. Biological annotations, stages, CITE-seq proteins, and other external references serve only to draw an equal number of cells per group and to evaluate outputs, except in the stage-aware zebrafish diagnostic of Appendix J.

3 Related Work

Single-cell imputation and downstream distortion. scRNA-seq recovery methods make different assumptions about observed zeros and how information should be shared across cells and genes. Whether droplet zeros need a dedicated dropout model is itself debated (Svensson, 2020). scImpute and DrImpute target dropout-like entries (Li & Li, 2018; Gong et al., 2018). SAVER models expression uncertainty (Huang et al., 2018). MAGIC relies on cell–cell diffusion (van Dijk et al., 2018). ALRA exploits low-rank structure while preserving a specific subset of zeros (Linderman et al., 2022). Neural approaches include DCA, DeepImpute, and scVI (Eraslan et al., 2019; Arisdakessian et al., 2019; Lopez et al., 2018), while D3Impute and scBFP add dropout discrimination or graph-based feature propagation (Huang et al., 2025; Lee et al., 2024). Foundation models broaden the candidate space: scGPT provides a generative pretrained model for single-cell multi-omics, and Geneformer learns contextual gene representations (Cui et al., 2024; Theodoris et al., 2023).

These approaches primarily aim to reconstruct expression or generate new representations before downstream analysis. However, imputation can alter statistical structures outside its fitting objective, including gene–gene correlations and reconstructed regulatory networks (Ly & Vingron, 2022), and it can create false-positive signals (Andrews & Hemberg, 2018). GCRP targets the analogous risk for progression inference. It evaluates an imputer's output as a residual proposal and controls how much that proposal can alter the underlying cell-state geometry.

Optimal transport and cell-state progression. OT relates destructive single-cell snapshots by coupling empirical cell-state distributions. The field now applies OT across temporal, perturbational, multimodal, and spatial settings (Bunne et al., 2024). Waddington-OT couples developmental populations across time while accounting for proliferation and death (Schiebinger et al., 2019). LineageOT incorporates lineage tracing into ancestor–descendant inference (Forrow & Schiebinger, 2021). TrajectoryNet combines continuous normalizing flows with dynamic OT to learn continuous paths between cross-sectional populations (Tong et al., 2020). CellOT learns neural OT maps for

heterogeneous perturbation responses and developmental transitions (Bunne et al., 2023). moscot provides scalable multimodal OT for temporal, spatial, and spatiotemporal mapping (Klein et al., 2025). scEGOT uses entropic Gaussian-mixture OT to map interpretable population dynamics and intermediate states (Yachimura et al., 2024). MultistageOT infers multistage differentiation from a single snapshot (Tronstad et al., 2025), while GeneTrajectory uses OT distances to order gene programs instead of cells (Qu et al., 2025). CellRank provides a complementary Markov-chain framework for directed fate mapping (Lange et al., 2022).

Recent work also learns Wasserstein geometry itself, through transformer embeddings of Wasserstein distances (Haviv et al., 2024), flow-matching transport (Klein et al., 2024), and flow matching over families of distributions (Haviv et al., 2025). Cluster-regularized OT brings transport into the reconstruction of missing single-cell data (Liu et al., 2026).

Across all these approaches, the central question is how to infer, approximate, or learn transport distances from cellular distributions. GCRP addresses a preceding step: determining which uncertain expression corrections are allowed to redefine the distributions and distances supplied to those downstream methods. It uses OT as an admissibility constraint on residual correction.

4 Geometry-Constrained Residual Projection

GCRP separates two roles that are usually combined during scRNA-seq imputation, predicting missing expression values and deciding how many of those predictions to use in downstream tasks such as cell-state progression. Because the observed expression matrix alone leaves open which zeros represent missing data, GCRP treats every fully imputed matrix as a proposal. Evaluating the residual R^{prop} instead of the full matrix separates two criteria: whether the observed data support filling a specific zero, and whether the new value preserves the population geometry.

We must evaluate these two criteria at different scales because local gene expression can support filling an individual zero, whereas many small changes can collectively distort the population structure. Therefore, GCRP uses three distinct objects. First, fixed affine *reference maps* $f_a(x) = A_a x + b_a$ control how far a single cell can move. The implementation fits PCA maps to standardized raw expression on the linear, log, and square-root scales. Only the linear-scale map is affine, so Proposition 1 applies to it, and the other two maps add heuristic limits (Appendix B). Second, *geometry groups* $\mathcal{G}$, which are k-means clusters of raw expression, define OT checks from RNA alone. Finally, we hold out biological variables for validation, including cell annotations, orthogonal modalities, disease or time labels, and simulation truth. We report separately the analyses in which biological annotations define the OT groups. Table 4 in Appendix B lists these objects for each study.

This hierarchical design matches our goals. Entry-level constraints verify that the observed data support the proposed change. Cell-level constraints limit how far those changes move individual cells. Population-level OT checks prevent these individual movements from rewriting the overall group geometry. We define a practical accepted residual set as:

$$\begin{aligned}\mathcal{M}_{\text{safe}} = \{R : {} & R_{ig} = 0 \text{ for } (i,g) \notin Z_0, \quad R_{ig} \geq 0 \text{ for } (i,g) \in Z_0, \\ & R_{ig} = 0 \text{ when } S_{\text{loc}}(i,g) < \alpha, \quad \|A_a R_i\|_2 \leq r_{ia} \ \forall a \in \mathcal{A}, \\ & D_{\text{OT}}^{\mathcal{G}}(X^{\text{obs}} + R, X^{\text{obs}}) \leq \epsilon_{\mathcal{G}}\}.\end{aligned}$$

Here R is a candidate residual matrix. The set $\mathcal{M}_{\text{safe}}$ describes admissibility; the implemented operator reaches it by gating and radial scaling, which differs from a metric projection onto this set. The first three constraints determine where a correction is allowed. The multi-map constraint limits the effect of the correction on each cell. GCRP enforces these entry- and cell-level constraints constructively by gating and by radial contraction into the declared displacement budgets. Residual scaling bounds each cell, and GCRP enforces the population-level OT limit separately by rejecting candidates during selection over $\mathcal{G}$ (Cuturi, 2013; Peyré & Cuturi, 2019; Schiebinger et al., 2019).

Local support and correction magnitude. For each cell i, let $\mathcal{N}(i)$ be its nearest neighbors in a representation learned from raw observed expression. For each zero entry (i, g), we define the local support as

$$S_{\text{loc}}(i,g) = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \mathbf{1}\{X_{jg}^{\text{obs}} > 0\}.$$

GCRP first forms a residual $R^{\text{prop}} = X^{\text{prop}} - X^{\text{obs}}$. We then enforce non-negative corrections only on observed zeros, require local support, and apply a global correction magnitude $\lambda \in [0, 1]$:

$$R_{ig}^{(0)} = \begin{cases} \lambda \max(R_{ig}^{\text{prop}}, 0), & X_{ig}^{\text{obs}} = 0 \text{ and } S_{\text{loc}}(i, g) \geq \alpha, \\ 0, & \text{otherwise.} \end{cases}$$

This stage evaluates whether observable evidence supports changing a particular zero. We consider a change only when enough raw-expression neighbors express the same gene, and the parameter λ controls the size of this change. Local support measures neighborhood evidence and leaves the dropout status of each zero open. The support threshold removes proposed changes that lack neighborhood evidence. Support and shrinkage control where and how much expression changes, and the multi-map displacement bound controls how far the resulting changes move each cell.

Multi-map displacement bound. We treat cell vectors and residual rows as columns when applying A_a. For the locally supported residual $R_i^{(0)}$, we define $d_{ia} = \|A_a R_i^{(0)}\|_2$. Given cell-specific and map-specific radius limits r_{ia}, GCRP defines $r_{ia}/d_{ia} = +\infty$ when $d_{ia} = 0$ and computes:

$$\gamma_i = \min\left(1, \min_{a \in \mathcal{A}} \frac{r_{ia}}{d_{ia}}\right), \qquad R_i^{\text{safe}} = \gamma_i R_i^{(0)}, \qquad x_i^{\text{safe}} = x_i + R_i^{\text{safe}}.$$

Even if every corrected entry has local support, the accumulated changes can still move a cell substantially within the state space. The displacement step prevents this by bounding the total distance a cell can travel. Each reference map supplies a view of the observed state space, and r_{ia} bounds the movement of cell i in that view. Because we share γ_i across maps, the most restrictive map determines the accepted residual, ensuring feasibility for every map in $\mathcal{A}$. In the displacement experiments, we set r_{ia} to 0.25 times the distance to the 11th non-self neighbor in a fixed PCA-20 map fitted to standardized raw expression. Across radius multipliers of 0.25, 0.50, and 1.0 for scVI, MAGIC, and ALRA, the projected outputs pass the OT limit in 24, 12, and 3 of 45 runs, respectively (Appendix B). A shared scale remains conservative when one map is much more restrictive than the others, and it avoids a more costly map-specific calculation.

Candidate selection. Candidate selection tests the aggregate effect of each candidate on the population. For each candidate $m \in C$, GCRP applies the support step and the displacement step above to $R^{(m)}$ and outputs $x_i^{\text{safe}}(m) = x_i + \gamma_i(m) R_i^{(0)}(m)$.

GCRP only scores candidates that satisfy the population-level geometric constraints. The scoring function $L(m)$ incorporates residual magnitude, the fraction of altered zeros, and OT cost. Before scoring, *reliability weighting* forms a weighted combination of the proposals. A proposal receives more weight when it agrees with the observed nonzero values, which remain unmasked, when its fills fall on supported zeros, and when it moves cells less. Appendix B specifies the weights and the scoring function. We select the candidate:

$$m^\star = \arg\max_{m \in C} L(m) \quad \text{subject to} \quad \rho_{\text{OT}}^{\mathcal{G}} \geq \rho_{\min},\ D_{\text{OT}}^{\mathcal{G}} \leq \epsilon_{\mathcal{G}}.$$

Proposition 1 guarantees $V(m) = \max_{i,a}(\|f_a(x_i^{\text{safe}}(m)) - f_a(x_i)\|_2 - r_{ia})_+ = 0$ for every displacement-scaled candidate, so GCRP filters candidates only on the population-level OT limits. We verify this equality up to numerical precision. GCRP rejects any candidate that changes the group geometry beyond $\epsilon_{\mathcal{G}}$, even if it achieves a high score $L(m)$. The selector enforces the rank condition with a stricter *transport-order check*, which requires pairwise OT costs among at most four raw-expression clusters to keep their raw sorted order, and it sets $\epsilon_{\mathcal{G}}$ from the gaps between these raw costs. When every candidate fails, GCRP shrinks the residual with the most filled zeros by a uniform scale found by bisection and returns X^{obs} when only the zero scale passes (Appendix B). Amplitude-controlled baselines undergo the same evaluation.

Hyperparameter grid and geometry checks. To evaluate the model, we test a grid of hyperparameters. We use raw RNA PCA with 20 principal components. The grid covers the neighborhood evidence thresholds $\alpha \in \{0.10, 0.20, 0.30\}$ and the correction magnitudes $\lambda \in \{0.25, 0.50, 1.0\}$. The released selector fixes $\alpha = 0.10$ and chooses λ by $L(m)$ under the OT limits, using RNA alone. Appendix F reports its sensitivity to α, to the number of clusters, and to the coefficients of $L(m)$. The local neighborhood $\mathcal{N}(i)$ contains the 21 nearest non-self cells. The evaluation requires an OT Spearman correlation of at least 0.98 and a scaled OT error of at most 0.10 per run. We summarize model

stability across runs using a pass fraction of 0.60. Appendix B details how these evaluation thresholds differ from the baseline implementation's acceptance rules. We use raw-cluster OT to evaluate geometry preservation from RNA alone, and biological-group OT to measure how well annotated populations are preserved.

4.1 Constrained Projection Bounds

Next, we use standard Wasserstein stability arguments (Peyré & Cuturi, 2019; Bunne et al., 2024) to obtain three bounds for the GCRP output. These bounds concern the transport geometry that trajectory-inference methods rely on (Schiebinger et al., 2019; Forrow & Schiebinger, 2021; Tong et al., 2020). First, we show that every accepted correction remains within all declared reference-map budgets. Second, these cell-level budgets bound how much population distributions and inter-group distances can change. Third, a progression ordering is preserved whenever its separation exceeds this induced uncertainty. Appendix A gives the proofs and two extensions. All three bounds measure change relative to the observed geometry, and they apply to that geometry only.

Proposition 1 (Multi-Map Feasibility). For each cell, GCRP scales the proposed residual by the largest factor $\gamma_i \in [0, 1]$ that satisfies every declared reference-map constraint. The resulting correction therefore obeys

$$\|f_a(x_i^{\text{safe}}) - f_a(x_i)\|_2 \le r_{ia} \qquad \text{for every } a \in \mathcal{A}.$$

Maximality holds only along the proposed residual direction.

Proposition 2 (Population and Geometry Stability). For a group $u \in \mathcal{G}$ and reference map a, let $P_{u,a}^{\text{raw}}$ and $P_{u,a}^{\text{safe}}$ be the empirical distributions of $f_a(x_i)$ and $f_a(x_i^{\text{safe}})$ over the cells $i \in u$. Define the group displacement budget $B_{u,a}^{(p)} = (|u|^{-1} \sum_{i \in u} r_{ia}^p)^{1/p}$. Then the distribution of group u moves by at most

$$W_p(P_{u,a}^{\text{safe}}, P_{u,a}^{\text{raw}}) \le B_{u,a}^{(p)}.$$

Consequently, the Wasserstein distance between any two groups u and h can change by at most $B_{u,a}^{(p)} + B_{h,a}^{(p)}$. The proof pairs each cell with its corrected version and applies the triangle inequality for W_p (Peyré & Cuturi, 2019). With $B_{\max} = \max_{u \in \mathcal{G}} B_{u,a}^{(2)}$, every entry of the pairwise W_2 distance matrix between groups changes by at most $2B_{\max}$. The same bound holds for the Gromov–Wasserstein distance between the raw and corrected group geometries when both carry a common group measure (Mémoli, 2011).

Proposition 2 links local correction to downstream geometry for any change in the optimal coupling between groups. The bound applies to the unregularized distance W_p in a fixed reference map. The OT metrics in our experiments use entropic Sinkhorn costs in a PCA embedding fitted separately to each output (Appendix B). Therefore, GCRP checks these metrics directly during candidate selection.

Proposition 3 (Preservation of Progression Ordering). Consider three groups $u, h, k \in \mathcal{G}$, and let d_{uv}^{raw} and d_{uv}^{safe} denote the W_p distance between groups u and v before and after correction. Suppose the raw geometry places h closer to u than k is. If the gap between these two distances exceeds the sum of their distortion bounds, then this ordering is preserved after correction.

Equivalently, writing $\eta_{uv} = B_{u,a}^{(p)} + B_{v,a}^{(p)}$, the condition is

$$d_{uk}^{\text{raw}} - d_{uh}^{\text{raw}} > \eta_{uh} + \eta_{uk},$$

under which $d_{uh}^{\text{safe}} < d_{uk}^{\text{safe}}$.

OT-based progression analyses rely on these relative distances. For example, if u, h, and k are healthy, noninflamed, and inflamed cells, a raw ordering whose margin exceeds $\eta_{uh} + \eta_{uk}$ keeps its order after correction. Below this margin, as in our Crohn's disease analysis, the ordering can still survive, but the bound is inconclusive (Appendix F). Section 5 counts the raw orderings on the CITE-seq panels that meet this condition.

5 Evaluation Methodology and Results

We evaluate how imputation candidates change local separation and the population geometry used for cell-state progression. As references for geometry preservation, we use orthogonal biological

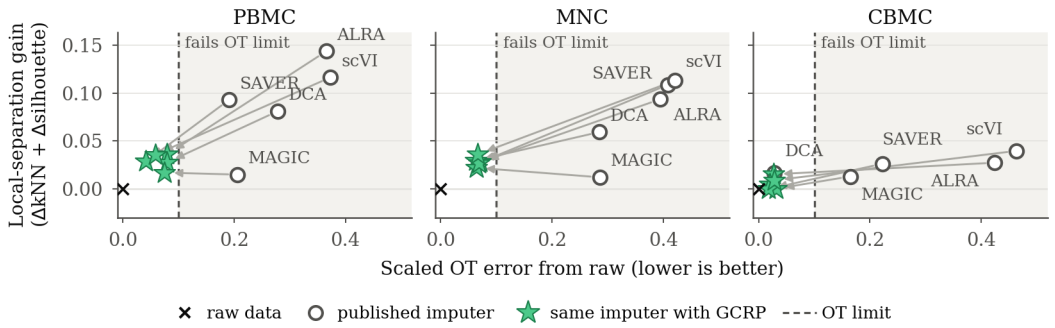

Figure 1: Five published imputers alone and with GCRP applied on top, on CITE-seq panels of PBMC and MNC (300 cells) and CBMC (360 cells) with 600 genes, averaged over five seeds. The horizontal axis is the scaled Sinkhorn error of biological-group distances relative to raw data, and lower is better. The vertical axis is ΔkNN + Δsilhouette relative to raw data and measures how strongly annotated cell types separate. Circles mark imputer outputs, stars mark the same imputer with GCRP, and arrows connect each pair. The dashed line marks the OT error limit of 0.10.

Table 1: Five published imputers alone and with GCRP applied on top, as means over five seeds. OT MAE, OT pass, local separation gain (ΔkNN + Δsilhouette), and zeros filled use the panels of Figure 1, and OT pass counts runs out of 15 that meet OT Spearman ≥ 0.98 and OT MAE ≤ 0.10. RNA–protein correlation uses six surface proteins on 640 CBMC cells, where raw data give 0.247. Parentheses give the change from the imputer alone. Bold marks the better value of each pair for OT MAE, OT pass, and RNA–protein correlation.

	OT MAE		OT pass (of 15)		Local separation gain		Zeros filled		RNA–protein correlation	
Imputer	Alone	With GCRP	Alone	With GCRP	Alone	With GCRP	Alone	With GCRP	Alone	With GCRP
scVI	0.419	**0.051** (−88%)	1	**12**	0.090	0.027	0.93	0.16	0.146	**0.290** (+0.144)
MAGIC	0.219	**0.057** (−74%)	1	**10**	0.013	0.013	0.93	0.17	0.260	**0.288** (+0.028)
ALRA	0.395	**0.059** (−85%)	1	**12**	0.088	0.026	0.70	0.15	**0.373**	0.302 (−0.071)
DCA	0.197	**0.054** (−73%)	5	**11**	0.052	0.021	0.93	0.17	0.262	**0.278** (+0.016)
SAVER	0.274	**0.043** (−84%)	2	**14**	0.076	0.018	0.93	0.17	0.160	**0.285** (+0.125)

measurements, such as CITE-seq antibody-derived tag (ADT) labels and clinical disease states, which are used only for evaluation. Using CITE-seq panels of peripheral blood (PBMC), bone marrow (MNC), and cord blood (CBMC) mononuclear cells, we compare raw expression against standard imputers, the zero-preserving imputer ALRA, shrinkage baselines, and variants gated by neighborhood evidence. We measure local separation by the differences in kNN accuracy (ΔkNN) and silhouette score (Δsilhouette) relative to raw data. To quantify population geometry distortion, we rely on scaled pairwise Sinkhorn error and OT Spearman correlation. We report OT metrics separately for biological groups and raw clusters. In both analyses, the local-separation metrics use biological annotations, which also define the biological OT groups. Tables 2 and 3 isolate the displacement projection and the support gate. Table 5 and Appendices F and J test the full selector.

Figure 1 compares the local-separation gain and the scaled OT error of five published imputers on the CITE-seq panels. Every imputer raises local separation over raw data. On MNC and PBMC, every imputer also exceeds the OT error limit, with mean errors of 0.19 to 0.42, and on CBMC four of the five exceed it. ALRA, which preserves a subset of zeros, passes the OT limit in 1 of its 15 runs. Applied on top of each imputer, GCRP lowers its mean OT error by a factor of 3.7 to 8.2, to between 0.04 and 0.06, and raises its OT pass rate from 1–5 to 10–14 of 15 runs (Table 1). It fills 15% to 17% of zeros, whereas the imputers fill 70% to 93%.

Takeaway (OT Filtering): Published imputers improve local separation but distort the population geometry. GCRP applied on top of any of them restores the geometry while filling about a fifth as many zeros.

To separate the effect of the displacement bound from population-level OT rejection, we apply the displacement projection of Section 4 to scVI, MAGIC, and ALRA (Lopez et al., 2018; van Dijk et al., 2018; Linderman et al., 2022) across five seeds (Table 2). The projection reduces the mean

Table 2: Displacement projection of three published imputers across five seeds. Base is the candidate method alone, and Projected is its projected output. Projected outputs skip the GCRP OT candidate selection, which would reject a failing output (Section 4). ΔkNN and Δsilhouette are changes in kNN accuracy and silhouette score from raw. Lower OT MAE and higher OT Spearman mean less distortion. Pass is the fraction of seeds in which the projected output meets OT Spearman ≥ 0.98 and OT MAE ≤ 0.10, and $\bar{\gamma}$ is the mean projection scale. Bold marks the lower distortion of each pair. Appendix E reports sample standard deviations and paired changes.

		ΔkNN		Δsilhouette		OT Spearman		OT MAE			
Dataset	Candidate	Base	Projected	Base	Projected	Base	Projected	Base	Projected	Pass	$\bar{\gamma}$
CBMC	scVI	0.025	0.012	0.015	0.004	0.785	**0.991**	0.463	**0.026**	0.80	0.38
MNC	scVI	0.018	0.001	0.095	0.009	0.874	**0.988**	0.421	**0.078**	0.40	0.15
PBMC	scVI	0.025	0.011	0.091	0.008	0.903	**0.954**	0.373	**0.072**	0.40	0.16
CBMC	MAGIC	0.004	0.001	0.009	0.005	0.964	**0.986**	0.165	**0.044**	0.60	0.74
MNC	MAGIC	−0.025	−0.005	0.037	0.013	0.872	**0.990**	0.286	**0.049**	0.40	0.20
PBMC	MAGIC	−0.045	0.003	0.060	0.009	0.913	**0.961**	0.206	**0.086**	0.40	0.17
CBMC	ALRA	0.011	0.011	0.017	0.006	0.872	**0.986**	0.425	**0.027**	0.80	0.37
MNC	ALRA	0.022	−0.001	0.072	0.007	0.910	**0.993**	0.394	**0.068**	0.60	0.14
PBMC	ALRA	0.035	0.012	0.109	0.011	0.910	**0.966**	0.366	**0.073**	0.40	0.13

Table 3: Crohn's disease progression in Enterocytes TMIGD1 MEP1A across five seeds. Columns follow Table 2, and OT metrics use the disease states as groups. Bold marks the preserved rank order and the lower OT MAE. Values are mean ± sample SD.

Method	ΔkNN	Δsilhouette	OT Spearman	OT MAE
support-gated, $\lambda = 0.25$	0.0077 ± 0.0099	0.0105 ± 0.0003	**1.000**	**0.008**
support-gated, $\lambda = 0.50$	0.0070 ± 0.0103	0.0179 ± 0.0006	**1.000**	0.017

OT MAE for scVI, MAGIC, and ALRA on every panel. After projection, the mean Δsilhouette remains positive in every case, and the mean ΔkNN remains positive except for MAGIC on MNC (−0.005) and ALRA on MNC (−0.001). On MNC, projecting scVI lowers its OT MAE from 0.421 to 0.078. The projection also lowers the OT MAE of DCA and SAVER (Eraslan et al., 2019; Huang et al., 2018) on every panel (Table 7). To connect these results to Section 4.1, we compute exact W_2 distances between OT groups in the fixed PCA-20 reference map. In all 135 projected runs over three radius multipliers, every group distance stays within its Proposition 2 bound. At the default radius, Proposition 3 certifies that 804 of 1,800 raw orderings are preserved (Appendix F).

Takeaway (Displacement Bounds): Displacement projection lowers OT MAE for all five imputers on every panel, and its Wasserstein bounds certify 45% of raw orderings. Projected outputs pass the OT limit in two to five of five seeds, so the OT check remains necessary.

To test whether the constraints preserve a clinical ordering, we hold the Enterocytes TMIGD1 MEP1A identity fixed in the Crohn's disease data (Kong et al., 2023) across three known clinical states: healthy, noninflamed, and inflamed (Table 3). We evaluate support-gated candidates that fill each supported zero with λ times the mean expression of its raw-expression neighbors, with a support threshold of $\alpha = 0.10$. Disease states define only the OT groups used for evaluation. Increasing λ from 0.25 to 0.50 increases Δsilhouette from 0.0105 to 0.0179. In contrast, the same increase lowers ΔkNN from 0.0077 to 0.0070 and raises the OT MAE from 0.008 to 0.017. The OT Spearman correlation is 1.000 at both values of λ. All 15 orderings keep their raw order, including those below the certified margin of Proposition 3.

Takeaway (Disease Progression): Support-gated corrections increase silhouette relative to raw data while preserving the rank order of distances among healthy, noninflamed, and inflamed cells.

To identify which components control geometric distortion, we run a leave-one-out ablation of GCRP over five seeds and 105 checks, reported in Table 5 of Appendix B. The full method passes 97 checks and leaves every observed nonzero unchanged. Removing the neighborhood evidence threshold raises the mean filled-zero fraction from 0.103 to 0.254 and lowers the total to 94 checks. Removing reliability weighting also lowers the total to 94 checks. Removing the displacement bound lowers

the total to 93 checks and raises the mean scaled OT error from 0.0102 to 0.0202. Removing the transport-order check keeps 97 checks but raises this error to 0.0135. The full method returns its top-scoring passing candidate in all 15 runs. Halving or doubling any coefficient of $L(m)$ keeps 97 checks, and three to six transport-order clusters give 96 to 98. Per-cell Euclidean caps and global shrinkage pass at most 97 checks, with a higher OT error than GCRP in every setting (Appendix F). On 20,000 cells and 2,000 genes, GCRP needs 706 ± 359 s and 14.0 GB of memory (Appendix G). On panels five times larger, it passes 91 of 105 checks with a mean scaled OT error of 0.016 (Table 14).

Takeaway (Ablation Results): The neighborhood evidence threshold, reliability weighting, and displacement bound each raise the check count. The transport-order check lowers the OT error.

Protection against held-out references. To test the geometry against references withheld from GCRP, we mask observed expression, corrupt it with dropout, and compare each output with the uncorrupted data (Appendix J). Full graph and SVD imputation lower held-out expression RMSE by 37% and 32%, and all 30 of their runs exceed the OT limit. We measure the relative state geometry by the error of median-normalized distances between cell types against the uncorrupted reference. Full graph imputation triples this error (+0.036, 95% CI 0.021 to 0.051). GCRP keeps it at the level of raw data (paired change −0.0001, 95% CI −0.006 to 0.006) while lowering held-out RMSE by 5%. Global shrinkage and per-cell caps that set their strength with the same label-free geometry check keep the reference geometry less closely. GCRP has the highest mean rank agreement with the uncorrupted reference, keeps neighborhoods closer to the reference than SVD shrinkage and caps do in all 15 runs, and changes about half as many zeros. These controls change more entries to recover more masked expression (Table 19). Surface proteins give a second independent reference. On CBMC, GCRP raises the RNA–protein correlation of six markers from 0.247 to 0.288 in every seed, whereas scVI and SAVER lower it to 0.146 and 0.160. ALRA raises it to 0.373 (Table 21). Applied on top of scVI and SAVER, GCRP raises their correlation to 0.290 and 0.285 in every seed (Table 1). On held-out zebrafish stages, GCRP lowers the error of OT interpolation by 0.2% in all five seeds, whereas full SVD raises it by 0.4% in all five (Table 22). On BMMC pseudotime, full SVD lowers the correlation between bin distances and pseudotime separation by 0.109, five times more than conservative support-gated fills, which lower it by 0.021 (Table 23).

Takeaway (Protection): GCRP keeps the relative state geometry at the level of raw data, changes the fewest entries, and still recovers expression. Full imputers recover more expression at the cost of this geometry or of protein concordance.

6 Conclusion

Zero imputation and cell-state progression inference in scRNA-seq are coupled problems: filling missing expression values changes the cellular geometry used to infer state relationships. We introduced GCRP, which treats the outputs of existing imputers as candidate residuals instead of true expression. It admits these corrections only when they have local neighborhood evidence and remain within cell-level displacement and population-level OT limits. Its bounds link cell displacement to the Wasserstein distortion of populations under affine reference maps.

On the CITE-seq panels, GCRP applied on top of each of five published imputers lowers that imputer's OT distortion 3.7- to 8.2-fold. At the default radius, the Wasserstein bounds certify that the displacement projection preserves 45% of the raw orderings of group distances. In the Crohn's disease analysis, support-gated corrections keep the rank order of distances among the three disease states. Against held-out references, GCRP keeps the relative state geometry at the level of raw data while full graph imputation triples its error. GCRP also raises RNA–protein concordance where scVI and SAVER lower it, and it changes about half as many zeros as simple controls tuned by the same label-free check. Because these guarantees hold relative to the observed geometry, end-to-end recovery on a new dataset depends on how faithfully that geometry reflects cell states. Imputation for progression analysis therefore needs evaluation against the population structure used for biological inference, and this structure can set how much correction to accept.

Reproducibility statement. Code is available at `https://anonymous.4open.science/r/GCRP-CA49`. It omits the learned base-imputer combination in Table 12 and the ablation variants in Table 5. Appendices A, B, and I give the proofs, implementation details, and datasets.

AI use statement. In this work, we used generative AI tools to correct grammar and polish text. We have reviewed all AI-assisted work and take responsibility for the final content of this work.

References

Tallulah S. Andrews and Martin Hemberg. False signals induced by single-cell imputation. *F1000Research*, 7:1740, 2018.

Cédric Arisdakessian, Olivier Poirion, Breck Yunits, Xun Zhu, and Lana X. Garmire. Deepimpute: an accurate, fast, and scalable deep neural network method to impute single-cell rna-seq data. *Genome Biology*, 20(1):211, Oct 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1837-6. URL https://doi.org/10.1186/s13059-019-1837-6.

Charlotte Bunne, Stefan G. Stark, Gabriele Gut, Jacobo Sarabia Del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, 20(11):1759–1768, 2023. doi: 10.1038/s41592-023-01969-x.

Charlotte Bunne, Geoffrey Schiebinger, Andreas Krause, Aviv Regev, and Marco Cuturi. Optimal transport for single-cell and spatial omics. *Nature Reviews Methods Primers*, 4(1):58, 2024. doi: 10.1038/s43586-024-00334-2.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, 21(8):1470–1480, 2024. doi: 10.1038/s41592-024-02201-0.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

Gökçen Eraslan, Lukas M. Simon, Maria Mircea, Nikola S. Mueller, and Fabian J. Theis. Single-cell rna-seq denoising using a deep count autoencoder. *Nature Communications*, 10 (1):390, 2019. doi: 10.1038/s41467-018-07931-2. URL https://doi.org/10.1038/s41467-018-07931-2.

Aden Forrow and Geoffrey Schiebinger. Lineageot is a unified framework for lineage tracing and trajectory inference. *Nature Communications*, 12:4940, 2021. doi: 10.1038/s41467-021-25133-1.

Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L. Nazor, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature Methods*, 18 (3):272–282, 2021. doi: 10.1038/s41592-020-01050-x.

Wuming Gong, Il-Youp Kwak, Pruthvi Pota, Naoko Koyano-Nakagawa, and Daniel J. Garry. Drimpute: imputing dropout events in single cell rna sequencing data. *BMC Bioinformatics*, 19(1):220, 2018. doi: 10.1186/s12859-018-2226-y.

Doron Haviv, Russell Zhang Kunes, Thomas Dougherty, Cassandra Burdziak, Tal Nawy, Anna Gilbert, and Dana Pe'er. Wasserstein wormhole: Scalable optimal transport distance with transformer. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17697–17718, 2024.

Doron Haviv, Aram-Alexandre Pooladian, Dana Pe'er, and Brandon Amos. Wasserstein flow matching: Generative modeling over families of distributions. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 22238–22258, 2025.

Mo Huang, Jingshu Wang, Eduardo Torre, Hannah Dueck, Sydney Shaffer, Roberto Bonasio, John I. Murray, Arjun Raj, Mingyao Li, and Nancy R. Zhang. Saver: gene expression recovery for single-cell rna sequencing. *Nature Methods*, 15(7):539–542, 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0033-z. URL https://doi.org/10.1038/s41592-018-0033-z.

Siyi Huang, Linfeng Jiang, Ming Yi, and Yuan Zhu. D3impute: Dropout-aware discrimination, distribution-aware modeling, and density-guide imputation for scrna-seq data. *PLOS Computational Biology*, 21(12):e1013744, 2025. doi: 10.1371/journal.pcbi.1013744.

Dominik Klein, Théo Uscidda, Fabian J. Theis, and Marco Cuturi. Genot: Entropic (gromov) wasserstein flow matching with applications to single-cell genomics. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, Laetitia Meng-Papaxanthos, Michael Sterr, Lama Saber, Changying Jing, Aimée Bastidas-Ponce, Perla Cota, Marta Tarquis-Medina, Shrey Parikh, Ilan Gold, Heiko Lickert, Mostafa Bakhti, Mor Nitzan, Marco Cuturi, and Fabian J. Theis. Mapping cells through time and space with moscot. *Nature*, 638(8052):1065–1075, 2025. doi: 10.1038/s41586-024-08453-2.

Lingjia Kong, Vladislav Pokatayev, Ariel Lefkovith, Grace T. Carter, Elizabeth A. Creasey, Chirag Krishna, Sathish Subramanian, Bharati Kochar, Orr Ashenberg, Helena Lau, Ashwin N. Ananthakrishnan, Daniel B. Graham, Jacques Deguine, and Ramnik J. Xavier. The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. *Immunity*, 56(2):444–458.e5, 2023. doi: 10.1016/j.immuni.2023.01.002.

Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B. Schiller, Dana Pe'er, and Fabian J. Theis. Cellrank for directed single-cell fate mapping. *Nature Methods*, 19(2):159–170, 2022. doi: 10.1038/s41592-021-01346-6.

Junseok Lee, Sukwon Yun, Yeongmin Kim, Tianlong Chen, Manolis Kellis, and Chanyoung Park. Single-cell rna sequencing data imputation using bi-level feature propagation. *Briefings in Bioinformatics*, 25(3):bbae209, 2024. doi: 10.1093/bib/bbae209.

Wei Vivian Li and Jingyi Jessica Li. An accurate and robust imputation method scimpute for single-cell rna-seq data. *Nature Communications*, 9(1):997, Mar 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-03405-7. URL https://doi.org/10.1038/s41467-018-03405-7.

George C. Linderman, Jun Zhao, Manolis Roulis, Piotr Bielecki, Richard A. Flavell, Boaz Nadler, and Yuval Kluger. Zero-preserving imputation of single-cell rna-seq data. *Nature Communications*, 13 (1):192, 2022. doi: 10.1038/s41467-021-27729-z.

Yuyu Liu, Jiannan Yang, Ziyang Yu, Weishen Pan, Fei Wang, and Tengfei Ma. Efficient imputation for patch-based missing single-cell data via cluster-regularized optimal transport. In *Proceedings of the 17th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '26, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400726538. doi: 10.1145/3807503.3819480. URL https://doi.org/10.1145/3807503.3819480.

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0229-2. URL https://doi.org/10.1038/s41592-018-0229-2.

Lam-Ha Ly and Martin Vingron. Effect of imputation on gene network reconstruction from single-cell rna-seq data. *Patterns*, 3(2):100414, 2022. doi: 10.1016/j.patter.2021.100414.

Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Found. Trends Mach. Learn.*, 11(5–6):355–607, February 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL https://doi.org/10.1561/2200000073.

Rihao Qu, Xiuyuan Cheng, Esen Sefik, Jay S. Stanley, III, Boris Landa, Francesco Strino, Sarah Platt, James Garritano, Ian D. Odell, Ronald Coifman, Richard A. Flavell, Peggy Myung, and Yuval Kluger. Gene trajectory inference for single-cell data by optimal transport metrics. *Nature Biotechnology*, 43(2):258–268, 2025. doi: 10.1038/s41587-024-02186-3.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. doi: 10.1016/j.cell.2019.01.006.

Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14(9):865–868, 2017. doi: 10.1038/nmeth.4380.

Valentine Svensson. Droplet scrna-seq is not zero-inflated. *Nature Biotechnology*, 38(2):147–150, Feb 2020. ISSN 1546-1696. doi: 10.1038/s41587-019-0379-5. URL https://doi.org/10.1038/s41587-019-0379-5.

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, Elizabeth M. Brydon, Zexian Zeng, X. Shirley Liu, and Patrick T. Ellinor. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023. doi: 10.1038/s41586-023-06139-9.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9526–9536, 2020.

Magnus Tronstad, Johan Karlsson, and Joakim S. Dahlin. Multistageot: Multistage optimal transport infers trajectories from a snapshot of single-cell data. *Proceedings of the National Academy of Sciences*, 122(50):e2516046122, 2025. doi: 10.1073/pnas.2516046122.

David van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J. Carr, Cassandra Burdziak, Kevin R. Moon, Christine L. Chaffer, Diwakar Pattabiraman, Brian Bierie, Linas Mazutis, Guy Wolf, Smita Krishnaswamy, and Dana Pe'er. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3):716–729.e27, 2018. ISSN 0092-8674. doi: 10.1016/j.cell.2018.05.061. URL https://www.sciencedirect.com/science/article/pii/S0092867418307244.

Toshiaki Yachimura, Hanbo Wang, Yusuke Imoto, Momoko Yoshida, Sohei Tasaki, Yoji Kojima, Yukihiro Yabuta, Mitinori Saitou, and Yasuaki Hiraoka. scegot: single-cell trajectory inference framework based on entropic gaussian mixture optimal transport. *BMC Bioinformatics*, 25:388, 2024. doi: 10.1186/s12859-024-05988-z.

A Detailed Proofs and Auxiliary Stability Results

This appendix gives the full statements and proofs underlying Section 4.1, followed by two auxiliary extensions. We first make explicit the quantities suppressed in the main text. Proposition 1 gives the cell-level certificate; Proposition 2 lifts it to empirical populations and the induced group metric space; Proposition 3 converts those metric error bounds into an ordering guarantee. The final two results extend the same argument to nonlinear reference maps and Lipschitz downstream summaries. As in Section 4.1, all guarantees are relative to the observed raw geometry.

A.1 Proposition 1: Multi-Map Feasibility

For cell i and reference map a, define

$$d_{ia} = \|A_a R_i^{(0)}\|_2, \qquad \gamma_i = \min\left(1, \min_{a\in\mathcal{A}} \frac{r_{ia}}{d_{ia}}\right), \qquad R_i^{\text{safe}} = \gamma_i R_i^{(0)},$$

where $r_{ia}/d_{ia} = +\infty$ if $d_{ia} = 0$.

Restatement of Proposition 1. For every declared affine map $f_a(x) = A_a x + b_a$,

$$\|f_a(x_i^{\text{safe}}) - f_a(x_i)\|_2 \le r_{ia} \qquad \forall a \in \mathcal{A}.$$

Moreover, γ_i is the largest feasible scaling on the ray $\{\gamma R_i^{(0)} : 0 \le \gamma \le 1\}$ satisfying all declared map constraints.

Proof. Because $R_i^{\text{safe}} = \gamma_i R_i^{(0)}$, the affine offsets cancel and

$$\|f_a(x_i^{\text{safe}}) - f_a(x_i)\|_2 = \gamma_i d_{ia}.$$

If $d_{ia} = 0$, this displacement is zero. Otherwise, $\gamma_i \le r_{ia}/d_{ia}$, so $\gamma_i d_{ia} \le r_{ia}$.

For maximality, let $\gamma \in [0, 1]$ be any feasible scaling. For every a with $d_{ia} > 0$, feasibility requires $\gamma d_{ia} \le r_{ia}$ and hence $\gamma \le r_{ia}/d_{ia}$. Therefore

$$\gamma \le \min\left(1, \min_{a\in\mathcal{A}} \frac{r_{ia}}{d_{ia}}\right) = \gamma_i.$$

Thus γ_i is the largest admissible step along the proposed residual direction. This is a radial feasibility result and makes no optimality claim over alternative directions. □

A.2 Proposition 2: Population and Geometry Stability

For group $u \in \mathcal{G}$ and fixed map a, define the raw and accepted empirical distributions

$$P_{u,a}^{\text{raw}} = \frac{1}{|u|}\sum_{i\in u} \delta_{f_a(x_i)}, \qquad P_{u,a}^{\text{safe}} = \frac{1}{|u|}\sum_{i\in u} \delta_{f_a(x_i^{\text{safe}})},$$

and, for $p \ge 1$, the group displacement budget

$$B_{u,a}^{(p)} = \left(\frac{1}{|u|}\sum_{i\in u} r_{ia}^p\right)^{1/p}.$$

Restatement of Proposition 2. The accepted population satisfies

$$W_p(P_{u,a}^{\text{safe}}, P_{u,a}^{\text{raw}}) \le B_{u,a}^{(p)}.$$

For any $u, h \in \mathcal{G}$,

$$\left|W_p(P_{u,a}^{\text{safe}}, P_{h,a}^{\text{safe}}) - W_p(P_{u,a}^{\text{raw}}, P_{h,a}^{\text{raw}})\right| \le B_{u,a}^{(p)} + B_{h,a}^{(p)}.$$

Consequently, if D^{raw} and D^{safe} are the pairwise group W_2 distance matrices, $\|\cdot\|_\infty$ denotes the largest absolute entry, and

$$B_{\max} = \max_{u\in\mathcal{G}} B_{u,a}^{(2)},$$

then

$$\|D^{\text{safe}} - D^{\text{raw}}\|_\infty \le 2B_{\max}.$$

If both group metric spaces carry the same measure μ, then

$$\mathrm{GW}_2\Big((\mathcal{G}, D^{\text{raw}}, \mu), (\mathcal{G}, D^{\text{safe}}, \mu)\Big) \le 2B_{\max}.$$

Proof. The raw and accepted empirical populations contain the same cells. Pair each cell with its corrected version:

$$\pi_{\text{id}} = \frac{1}{|u|} \sum_{i \in u} \delta_{(f_a(x_i^{\text{safe}}), f_a(x_i))}.$$

This is an admissible coupling of $P_{u,a}^{\text{safe}}$ and $P_{u,a}^{\text{raw}}$. Hence, by Proposition 1,

$$\begin{aligned} W_p^p(P_{u,a}^{\text{safe}}, P_{u,a}^{\text{raw}}) &\le \frac{1}{|u|} \sum_{i \in u} \|f_a(x_i^{\text{safe}}) - f_a(x_i)\|_2^p \\ &\le \frac{1}{|u|} \sum_{i \in u} r_{ia}^p = \left(B_{u,a}^{(p)}\right)^p. \end{aligned}$$

Taking p-th roots proves the bound on population displacement.

Now write

$$d_{uh}^{\text{raw}} = W_p(P_{u,a}^{\text{raw}}, P_{h,a}^{\text{raw}}), \qquad d_{uh}^{\text{safe}} = W_p(P_{u,a}^{\text{safe}}, P_{h,a}^{\text{safe}}).$$

The Wasserstein triangle inequality gives

$$d_{uh}^{\text{safe}} \le W_p(P_{u,a}^{\text{safe}}, P_{u,a}^{\text{raw}}) + d_{uh}^{\text{raw}} + W_p(P_{h,a}^{\text{raw}}, P_{h,a}^{\text{safe}}),$$

and therefore

$$d_{uh}^{\text{safe}} - d_{uh}^{\text{raw}} \le B_{u,a}^{(p)} + B_{h,a}^{(p)}.$$

Interchanging "raw" and "safe" yields the reverse inequality, so

$$|d_{uh}^{\text{safe}} - d_{uh}^{\text{raw}}| \le B_{u,a}^{(p)} + B_{h,a}^{(p)}.$$

This argument does not require the optimal coupling between groups to stay fixed.

For $p = 2$, each entry satisfies

$$|D_{uh}^{\text{safe}} - D_{uh}^{\text{raw}}| \le 2B_{\max},$$

which gives

$$\|D^{\text{safe}} - D^{\text{raw}}\|_\infty \le 2B_{\max}.$$

Finally, under a common group measure μ, use the identity coupling between the two group metric spaces in the Gromov–Wasserstein objective:

$$\begin{aligned} \mathrm{GW}_2^2 &\le \int_{\mathcal{G}\times\mathcal{G}} |D_{uh}^{\text{raw}} - D_{uh}^{\text{safe}}|^2 \, d\mu(u) \, d\mu(h) \\ &\le (2B_{\max})^2. \end{aligned}$$

Taking square roots proves the stated GW bound. These are standard coupling and metric-perturbation arguments (Peyré & Cuturi, 2019; Bunne et al., 2024); here they propagate the GCRP cell-level certificate to the full group geometry. □

A.3 Proposition 3: Preservation of Progression Ordering

For $u, v \in \mathcal{G}$, define

$$d_{uv}^{\text{raw}} = W_p(P_{u,a}^{\text{raw}}, P_{v,a}^{\text{raw}}), \qquad d_{uv}^{\text{safe}} = W_p(P_{u,a}^{\text{safe}}, P_{v,a}^{\text{safe}}),$$

and

$$\eta_{uv} = B_{u,a}^{(p)} + B_{v,a}^{(p)}.$$

By Proposition 2,

$$|d_{uv}^{\text{safe}} - d_{uv}^{\text{raw}}| \le \eta_{uv}.$$

Restatement of Proposition 3. For $u, h, k \in \mathcal{G}$, if

$$d_{uk}^{\text{raw}} - d_{uh}^{\text{raw}} > \eta_{uh} + \eta_{uk},$$

then

$$d_{uh}^{\text{safe}} < d_{uk}^{\text{safe}}.$$

Proof. The nearer raw distance can increase by at most η_{uh},

$$d_{uh}^{\text{safe}} \le d_{uh}^{\text{raw}} + \eta_{uh},$$

while the farther raw distance can decrease by at most η_{uk},

$$d_{uk}^{\text{safe}} \ge d_{uk}^{\text{raw}} - \eta_{uk}.$$

The margin condition is equivalent to

$$d_{uh}^{\text{raw}} + \eta_{uh} < d_{uk}^{\text{raw}} - \eta_{uk}.$$

Therefore

$$d_{uh}^{\text{safe}} \le d_{uh}^{\text{raw}} + \eta_{uh} < d_{uk}^{\text{raw}} - \eta_{uk} \le d_{uk}^{\text{safe}}.$$

Whenever the condition holds, the ordering is preserved. An ordering whose margin falls below the condition can still keep its order, and the certificate leaves such an ordering undetermined. By contraposition, if an accepted correction reverses the ordering, the raw margin was at most the certified distortion budget. The argument applies independently to any finite collection of strict distance comparisons. □

A.4 Auxiliary Extension: Nonlinear Reference Maps

The main propositions are stated for the declared affine maps. The same cell-level certificate also bounds cell displacement under a nonlinear preprocessing map whose coordinate-wise derivative is bounded. This extension requires a declared identity map. The implementation in Appendix B does not declare one, so the extension does not bound its log and square-root maps.

Nonlinear reference-map bound. Suppose the declared maps include the identity with radius r_{i0}, so that $\|x_i^{\text{safe}} - x_i\|_2 \le r_{i0}$. Let g act coordinate-wise and satisfy

$$\max_j \sup_{t \ge 0} |g_j'(t)| \le K.$$

For $\widetilde{f}_a(x) = A_a g(x) + b_a$,

$$\|\widetilde{f}_a(x_i^{\text{safe}}) - \widetilde{f}_a(x_i)\|_2 \le K \|A_a\|_{\text{op}} r_{i0}.$$

For $g_j(t) = \log(1+t)$, one may take $K = 1$.

Proof. Let $\Delta_i = x_i^{\text{safe}} - x_i$. GCRP adds nonnegative residuals to nonnegative expression, so every coordinate of x_i and x_i^{safe} lies in $[0, \infty)$. By the coordinate-wise mean value theorem,

$$g(x_i^{\text{safe}}) - g(x_i) = J_i \Delta_i,$$

where J_i is diagonal and $\|J_i\|_{\text{op}} \le K$. Hence

$$\begin{aligned} \|\widetilde{f}_a(x_i^{\text{safe}}) - \widetilde{f}_a(x_i)\|_2 &= \|A_a J_i \Delta_i\|_2 \\ &\le \|A_a\|_{\text{op}} \|J_i\|_{\text{op}} \|\Delta_i\|_2 \\ &\le K \|A_a\|_{\text{op}} r_{i0}. \end{aligned}$$

For $g_j(t) = \log(1+t)$ on $t \ge 0$, $|g_j'(t)| = (1+t)^{-1} \le 1$. □

A.5 Auxiliary Extension: Lipschitz Downstream Summaries

Proposition 2 also controls continuous scalar summaries of the pairwise group geometry.

Bound on Lipschitz downstream summaries. Let

$$\Phi : \mathbb{R}^{\binom{|\mathcal{G}|}{2}} \to \mathbb{R}$$

be L-Lipschitz with respect to the ℓ_1 norm on the vectors $\mathbf{d}^{\text{raw}}$ and $\mathbf{d}^{\text{safe}}$ of upper-triangular pairwise group W_p distances. Then

$$|\Phi(\mathbf{d}^{\text{safe}}) - \Phi(\mathbf{d}^{\text{raw}})| \le L \sum_{u<h} \left(B_{u,a}^{(p)} + B_{h,a}^{(p)} \right).$$

Proof. Let $\mathbf{d}^{\text{raw}}$ and $\mathbf{d}^{\text{safe}}$ collect the upper-triangular entries of the two pairwise group-distance matrices. Lipschitz continuity gives

$$|\Phi(\mathbf{d}^{\text{safe}}) - \Phi(\mathbf{d}^{\text{raw}})| \le L \|\mathbf{d}^{\text{safe}} - \mathbf{d}^{\text{raw}}\|_1.$$

Applying Proposition 2 entrywise,

$$\begin{aligned} |\Phi(\mathbf{d}^{\text{safe}}) - \Phi(\mathbf{d}^{\text{raw}})| &\le L \sum_{u<h} |D_{uh}^{\text{safe}} - D_{uh}^{\text{raw}}| \\ &\le L \sum_{u<h} \left(B_{u,a}^{(p)} + B_{h,a}^{(p)} \right). \end{aligned}$$

This applies to summaries satisfying the stated Lipschitz condition. Discrete orderings or topology-changing procedures require a margin argument such as Proposition 3. □

B Implementation Details

GCRP separates the weighting of candidate proposals from the selection of the final residual. For reliability weighting, it samples at most 20,000 observed nonzero entries and measures each proposal's agreement with the raw values at those entries. The weight of each proposal is the product of this agreement term and heuristic terms that increase with its fill rate and with the fraction of its fills on supported zeros and decrease with its displacement ratio, its residual size, and any change to observed nonzeros. GCRP builds four residual variants from weighted summaries of the proposal residuals: the weighted mean, the median of the highest-weighted proposals, and the weighted mean minus 0.22 times the weighted standard deviation. The direct variant fills locally supported zeros on which the proposals agree. The marker variant also fills zeros whose gene is enriched in the cell's raw-expression cluster. The contrast variant weights the lower value most and fills zeros whose gene separates the cell's cluster from the other clusters. The blended variant takes the entrywise maximum of the other three. Each variant is capped by a per-entry amplitude limit and then passes through the displacement step. The sampled nonzero entries are not masked, and candidate methods are not refitted for this calculation. The resulting quantity is a heuristic based on agreement with observed positive values and does not measure held-out recovery accuracy. GCRP restores all observed nonzero entries to their raw values in the final candidate outputs.

For each residual variant m, the final score is:

$$L(m) = e_m + 0.018 \log(1 + 120 f_m) - 0.012 D_m + 10^{-4} \rho_m + b_m.$$

Here, e_m is the mean scaled residual norm across cells, f_m is the fraction of observed zeros increased by more than 10^{-5}, and D_m and ρ_m are $D_{\text{OT}}^{\mathcal{G}}$ and $\rho_{\text{OT}}^{\mathcal{G}}$ for candidate m. The additive offset b_m is 0.005 for the blended residual, 0.003 for the contrast residual, and zero for the other variants. GCRP tests geometry admissibility before attempting to maximize this score.

Let c and $c^{(m)}$ contain the pairwise Sinkhorn costs for the raw and candidate matrices under identical groups. The scaled OT error and rank statistic are:

$$D_m = \frac{1}{|c|} \sum_j \left| \frac{c_j}{\text{med}_+(c)} - \frac{c_j^{(m)}}{\text{med}_+(c^{(m)})} \right|, \qquad \rho_m = \text{Spearman}(c, c^{(m)}),$$

where med_+ is the median over positive costs. We apply a unit denominator if no positive costs exist. The scaled OT error measures relative group-distance distortion following separate normalization, so

it ignores a uniform rescaling of all group distances. The internal selector samples up to 35 cells from each of at most four raw-expression clusters and applies a Sinkhorn regularization of 0.08 across 80 iterations. This transport-order check requires the same sorted pair order when at most six pairwise costs are available, as with the default four clusters. With more costs, the check requires a rank correlation of at least 0.995. The limit $\epsilon_{\mathcal{G}}$ on the scaled OT error is $\mathrm{clip}(0.68\,\mathrm{median}(g), 0.020, 0.080)$, where g contains the positive gaps between sorted, median-normalized raw costs. The fallback limit is 0.055 when no such gaps exist.

Output when no candidate passes. GCRP returns the passing candidate with the highest $L(m)$. If no candidate passes, GCRP runs eight bisection steps on a uniform scale in $[0, 1]$ for the residual with the most filled zeros and keeps the last scale on this path that passes the transport-order check. This output equals X^{obs} when no positive scale passes. If the output fills fewer than a fraction $\max(1/n, 1/p)$ of observed zeros, GCRP adds at most 2×10^{-4} to supported zeros. It keeps this addition only if the transport order still holds with scaled OT error at most $\max(\epsilon_{\mathcal{G}}, 0.03)$. This step does not recheck the displacement budgets. Finally, GCRP restores observed nonzero entries and clips the output to $[0, 1]$. Across the 300 selector runs of Appendix F, GCRP returns its top-scoring passing candidate every time and never adds the small fill. All 25 scaling runs of Appendix G also return a passing candidate.

The hyperparameter grid uses 600 genes. PBMC and MNC evaluations run on 300 cells, while CBMC runs on 360 cells. The optimal grid runs use PCA-20, 25 cells per OT group, and 80 Sinkhorn iterations with a regularization of 0.08. The evaluation embedding for the OT metrics applies $\log(1+x)$, standardization, and PCA fitted separately to each output matrix. The displacement reference map stays fixed to the raw data. We select genes based on the variance of $\log(1 + x)$ and scale each gene to $[0, 1]$ using the nonzero 5th and 95th percentiles. The same scaling rule applies to both the raw and candidate matrices, with quantiles fitted separately to each matrix. The real and shuffled neighborhood evidence controls use 1,000 genes, evaluating 400 cells for PBMC and MNC, and 640 for CBMC. The five evaluation seeds are identical across support conditions. The shuffling protocol preserves the total supported-zero count for each gene while breaking cell alignment.

Biological-group OT and raw-cluster OT represent separate analyses. The local-separation scores displayed in the main text use biological labels and serve only as retrospective evaluation metrics. They are not the implemented label-free selection score $L(m)$. Table 4 lists the OT groups and reporting variables for each analysis.

Table 4: Geometry evaluation protocols. Biological grouping variables enter the indicated OT comparisons. Raw-cluster OT defines its groups from RNA, while its reported local-separation metrics still use biological labels. These comparisons do not establish masked-entry recovery.

Setting	RNA input	OT groups	Reporting variables
CITE-seq biological OT	observed RNA	cell type or ADT clusters	cell type, ADT
CITE-seq raw-cluster OT	observed RNA	raw RNA clusters	cell type, ADT
Crohn progression	observed RNA	disease states	epithelial cell type
Evidence controls	observed RNA	cell type or ADT clusters	real and shuffled evidence

Ablation criteria and scope. The component ablation uses seven checks per panel: positive kNN and silhouette differences from the raw data, OT Spearman of at least 0.98, scaled OT MAE of at most 0.10, a filled-zero fraction of at least 0.001, an observed-nonzero change fraction of at most 0.001, and a mean displacement divided by raw neighbor radius of at most 1. Each seed contributes 21 checks across the three panels, and Table 5 reports totals over five seeds. This count is a diagnostic. The full method is the four-component method in the released code. The seven-component rows use an earlier version with three auxiliary components. The removed identity guard tests for neighborhood overlap, the recovery guard evaluates predictions of raw-cluster labels, and the shadow component supplies an auxiliary neighbor-derived residual. Removing the identity guard leaves every check score unchanged. These component comparisons isolate mechanics but do not establish necessity across all datasets or seeds.

Displacement limit calculation. The displacement bound uses PCA-20 fitted to standardized raw expression and sets r_{ia} to 0.25 times the distance to the 11th non-self neighbor. It maintains a scale of one when the cell displacement is at most 10^{-12}. For larger displacements, it applies $\min(1, r/d)$. The largest recorded radius excess in the reported five-seed comparison is below 1.3×10^{-5} in these PCA-20 coordinates. We verify this numerical residual separately from the OT limit. Across radius

Table 5: Leave-one-out component ablation on the PBMC, MNC, and CBMC panels over five seeds. Checks count the passed geometry and local-separation criteria out of 105, seven per panel and seed. Higher is better for kNN and silhouette improvement; lower is better for Sinkhorn scaled error. kNN, silhouette, and Sinkhorn error are means over panels and seeds. Zero fill is the mean filled-zero fraction, and displacement is the largest panel displacement ratio averaged over seeds. This diagnostic is distinct from the affine displacement bound.

Configuration	Checks	kNN	Silhouette	Sinkhorn err.	Zero fill	Displacement
Full method	97	0.0043	0.0153	0.0102	0.103	0.29
Removals from the full method						
neighborhood evidence threshold off	94	0.0018	0.0138	0.0093	0.254	0.29
reliability weighting off	94	0.0031	0.0161	0.0118	0.106	0.30
displacement bound off	93	0.0012	0.0328	0.0202	0.226	1.55
transport-order check off	97	0.0045	0.0148	0.0135	0.108	0.30
Seven-component method	97	0.0037	0.0117	0.0099	0.121	0.28
Removals from the seven-component method						
neighbor-identity guard off	97	0.0037	0.0117	0.0099	0.121	0.28
recovery guard off	97	0.0042	0.0153	0.0102	0.103	0.29
neighbor-shadow lane off	97	0.0037	0.0117	0.0099	0.121	0.28

multipliers, we apply the projection to scVI, MAGIC, and ALRA on the three CITE-seq panels with biological OT groups and five seeds. As the multiplier increases from 0.25 to 0.50 and 1.0, the mean scaled OT MAE rises from 0.058 to 0.108 and 0.191, and the mean sum of ΔkNN and Δsilhouette rises from 0.013 to 0.025 and 0.039. The full GCRP selector applies the same ratio test in three PCA maps of standardized raw expression. These maps use 20 components on the linear scale, 25 components after $\log(1 + x)$, and 18 components after the square root, with radii of 0.30, 0.34, and 0.33 times the distance to the 11th non-self neighbor. GCRP scales each residual by the smallest of the three ratios. Proposition 1 holds for the linear-scale map because that map is affine. The logarithm and square root are concave, so scaling a residual by the computed ratio can leave its displacement in those two maps above the radius. These two maps therefore act as additional limits without a guarantee.

C Reference Map Sensitivity

GCRP protects geometry only with respect to the reference maps supplied to it. Therefore, we test robustness to changes in those maps. We evaluate sensitivity by setting the dimensionality of the PCA map to ten, twenty, and forty components to test whether the projection depends on a particular truncation of the raw expression space. Across every setting, we evaluate the same scVI candidate and hold the radius rule constant; only the reference map varies. We report OT Spearman to measure how well the ordering of group distances is preserved, and we report scaled OT MAE to measure the relative distance distortion.

Table 6 reports the minimum and maximum ranges across the PCA dimensionalities. As Proposition 1 guarantees, every tested map has zero radius violations up to numerical precision. The downstream geometry remains stable on CBMC and MNC, producing OT Spearman ranges of 0.989–0.991 and 0.981–0.988, respectively. In contrast, PBMC is more sensitive, with the correlation ranging from 0.927 to 0.968, although its scaled OT MAE remains between 0.065 and 0.090. These experiments separate the formal guarantee from empirical robustness. The radius bound holds for every chosen map by construction, but preserving a specific population geometry can still depend on which representation we declare as the reference map. PBMC shows this dependence, which motivates using multiple defensible reference maps when the raw representation remains uncertain.

D Projection Summary

Figure 2 summarizes the mechanism isolated in Table 2. We apply the same projection to three qualitatively different published imputers and compare the results against each unprojected candidate across CBMC, MNC, and PBMC. This comparison separates the effect of the GCRP displacement

Table 6: Reference-map sensitivity of the projected scVI candidate, reported as the minimum and maximum over three PCA dimensionalities. The affine reference map uses ten, twenty, or forty principal components. Across every reference map, the displacement limit holds up to numerical precision. PBMC exhibits the lowest OT Spearman values.

	OT Spearman		OT MAE	
Dataset	Min	Max	Min	Max
CBMC	0.989	0.991	0.026	0.038
MNC	0.981	0.988	0.077	0.088
PBMC	0.927	0.968	0.065	0.090

bound from the choice of imputation model. Dense scVI, MAGIC, and ALRA residuals move cells far enough to distort the population geometry. We find that the GCRP projection enforces the displacement limit and reduces the average OT distortion for all three imputers. This comparison isolates the projection step and does not apply the final OT rejection stage. The radius limit alone does not ensure that a candidate passes the final OT limit: several scVI, MAGIC, and ALRA settings pass the final limit on only two of five seeds. The mean projection scale ranges from 0.13 for ALRA on PBMC to 0.74 for MAGIC on CBMC. Each cell's scale depends on its displacement relative to its radius, so the projection does not shrink all candidates uniformly.

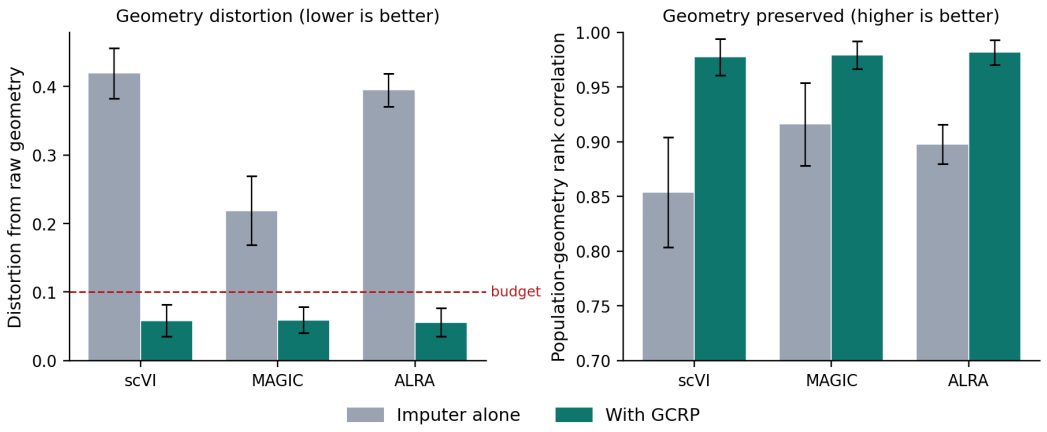

Figure 2: The GCRP projection reduces average OT distortion for scVI, MAGIC, and ALRA. Bars compare each candidate with its projected output, averaged over CBMC, MNC, and PBMC. Error bars are the standard deviation across panels. Left: scaled pairwise OT error relative to raw. Right: OT rank correlation. These averages do not imply that every seed passes the final OT limit.

DCA and SAVER. We apply the same displacement projection to two further published imputers, DCA and SAVER (Eraslan et al., 2019; Huang et al., 2018) (Table 7). We run DCA 0.3.4 in denoise mode and SAVER 1.1.2 from their official packages with default settings, on the same count matrices as the other imputers. The projection lowers the mean OT MAE of both imputers on every panel. DCA on CBMC passes the OT limit in all five seeds even without projection, with a mean OT MAE of 0.028. Apart from this case, the base outputs pass the OT limit in at most two of five seeds, and the projected outputs pass in two to five.

GCRP on top of each imputer. To test GCRP as a layer on a single imputer, we give GCRP only that imputer's output and its blends at 0.25 and 0.50 as proposals, on the panels of Table 2 and the CBMC panel of Table 21 (Table 1). For every imputer, GCRP lowers the mean OT MAE by a factor of 3.7 to 8.2 and lowers it in 13 to 15 of 15 runs. The OT pass rate rises from 1–5 to 10–14 of 15 runs. GCRP keeps the full local gain of MAGIC and 24% to 40% of the gain of the other imputers while filling 15% to 17% of zeros. GCRP raises the RNA–protein correlation of scVI from 0.146 to 0.290 ($p = 8.6 \times 10^{-5}$), of SAVER from 0.160 to 0.285 ($p = 1.2 \times 10^{-4}$), and of MAGIC from 0.260 to 0.288 ($p = 0.02$), each in all five seeds. It changes the DCA correlation little (0.262 to 0.278) and lowers the ALRA correlation from 0.373 to 0.302. The ALRA difference comes almost entirely from CD4, whose correlation ALRA raises from 0.107 to 0.494 by filling 58% of zeros on

Table 7: Displacement projection of DCA and SAVER across five seeds. Columns follow Table 2, with the default radius multiplier 0.25. Values are means over seeds.

		ΔkNN		Δsilhouette		OT Spearman		OT MAE			
Dataset	Candidate	Base	Projected	Base	Projected	Base	Projected	Base	Projected	Pass	$\bar{\gamma}$
CBMC	DCA	0.010	0.005	0.007	0.006	0.994	0.996	0.028	0.020	1.00	0.83
MNC	DCA	0.004	0.003	0.056	0.013	0.947	0.993	0.285	0.058	0.60	0.19
PBMC	DCA	0.013	0.014	0.068	0.013	0.888	0.966	0.278	0.071	0.40	0.18
CBMC	SAVER	0.019	0.005	0.007	0.002	0.889	0.991	0.222	0.021	0.80	0.42
MNC	SAVER	0.014	−0.003	0.095	0.007	0.886	0.993	0.409	0.070	0.40	0.16
PBMC	SAVER	0.014	0.009	0.079	0.010	0.935	0.947	0.191	0.054	0.40	0.31

this panel. GCRP applied on top of ALRA fills 4% of zeros, keeps the CD14 and CD16 gains, raises the CD3 correlation above that of ALRA (0.541 against 0.436), and keeps the correlation across the six markers above that of raw data in all five seeds ($p = 4.2 \times 10^{-4}$).

E Projection Results Across Seeds

Tables 8 and 9 separate the two local-separation metrics and report detailed uncertainty for all three tested candidates. We report all values as means and sample standard deviations evaluated over the same five seeds. We compute the paired differences within each seed before aggregation.

Table 8: Base and linear-projected results across five seeds. ΔkNN and Δsilhouette are differences from raw within each seed. OT Spearman and scaled OT MAE compare biological-group distances against the raw data. All entries are mean ± sample SD.

Panel	Candidate	Output	ΔkNN	Δsilhouette	OT Spearman	OT MAE
CBMC	scVI	Base	0.025 ± 0.015	0.015 ± 0.006	0.785 ± 0.155	0.463 ± 0.557
CBMC	scVI	Projected	0.012 ± 0.011	0.004 ± 0.001	0.991 ± 0.015	0.026 ± 0.017
CBMC	MAGIC	Base	0.004 ± 0.018	0.009 ± 0.001	0.964 ± 0.034	0.165 ± 0.273
CBMC	MAGIC	Projected	0.001 ± 0.007	0.005 ± 0.002	0.986 ± 0.013	0.044 ± 0.036
CBMC	ALRA	Base	0.011 ± 0.017	0.017 ± 0.005	0.872 ± 0.110	0.425 ± 0.526
CBMC	ALRA	Projected	0.011 ± 0.014	0.006 ± 0.000	0.986 ± 0.016	0.027 ± 0.015
MNC	scVI	Base	0.018 ± 0.018	0.095 ± 0.018	0.874 ± 0.066	0.421 ± 0.234
MNC	scVI	Projected	0.001 ± 0.012	0.009 ± 0.001	0.988 ± 0.015	0.078 ± 0.047
MNC	MAGIC	Base	−0.025 ± 0.013	0.037 ± 0.011	0.872 ± 0.070	0.286 ± 0.186
MNC	MAGIC	Projected	−0.005 ± 0.019	0.013 ± 0.002	0.990 ± 0.013	0.049 ± 0.033
MNC	ALRA	Base	0.022 ± 0.029	0.072 ± 0.014	0.910 ± 0.033	0.394 ± 0.221
MNC	ALRA	Projected	−0.001 ± 0.011	0.007 ± 0.001	0.993 ± 0.011	0.068 ± 0.042
PBMC	scVI	Base	0.025 ± 0.010	0.091 ± 0.020	0.903 ± 0.041	0.373 ± 0.146
PBMC	scVI	Projected	0.011 ± 0.007	0.008 ± 0.002	0.954 ± 0.052	0.072 ± 0.021
PBMC	MAGIC	Base	−0.045 ± 0.028	0.060 ± 0.013	0.913 ± 0.076	0.206 ± 0.101
PBMC	MAGIC	Projected	0.003 ± 0.020	0.009 ± 0.004	0.961 ± 0.054	0.086 ± 0.022
PBMC	ALRA	Base	0.035 ± 0.017	0.109 ± 0.019	0.910 ± 0.033	0.366 ± 0.139
PBMC	ALRA	Projected	0.012 ± 0.013	0.011 ± 0.003	0.966 ± 0.041	0.073 ± 0.019

F Selector Robustness, Simple Baselines, and Bound Checks

Selector settings. We rerun the released selector on the panels of Table 5 over the same five seeds and change one setting at a time (Table 10). Scaling any coefficient of $L(m)$ or the offsets b_m by 0.5 or 2, or removing the offsets, keeps 97 checks and 10 OT passes. Each of these variants returns the default output in at least 14 of 15 runs. The support threshold matters more. Lowering α to 0.05 raises the filled-zero fraction from 0.103 to 0.173, and GCRP passes 96 checks. Raising α to 0.20 or 0.30 lowers the filled-zero fraction to 0.082 or 0.057, and GCRP passes 92 or 93 checks. With three to six transport-order clusters, the method passes 96 to 98 checks, and the OT error decreases as the number of clusters grows. We chose these defaults on the same panels, so this analysis measures stability around the chosen setting and does not replace tuning on held-out data.

Table 9: Paired projection changes, projected minus base, as mean ± sample SD across five seeds. A negative OT MAE difference indicates reduced geometric distortion.

Panel	Candidate	ΔkNN change	Δsilhouette change	OT Spearman change	OT MAE change
CBMC	scVI	−0.013 ± 0.019	−0.011 ± 0.005	0.206 ± 0.160	−0.438 ± 0.550
CBMC	MAGIC	−0.003 ± 0.017	−0.004 ± 0.001	0.021 ± 0.025	−0.121 ± 0.243
CBMC	ALRA	0.000 ± 0.027	−0.011 ± 0.005	0.114 ± 0.116	−0.398 ± 0.517
MNC	scVI	−0.016 ± 0.007	−0.086 ± 0.019	0.114 ± 0.077	−0.343 ± 0.196
MNC	MAGIC	0.020 ± 0.021	−0.025 ± 0.011	0.119 ± 0.075	−0.237 ± 0.167
MNC	ALRA	−0.023 ± 0.021	−0.065 ± 0.015	0.082 ± 0.029	−0.326 ± 0.187
PBMC	scVI	−0.015 ± 0.010	−0.084 ± 0.019	0.051 ± 0.081	−0.301 ± 0.142
PBMC	MAGIC	0.048 ± 0.031	−0.051 ± 0.010	0.048 ± 0.048	−0.120 ± 0.094
PBMC	ALRA	−0.023 ± 0.024	−0.099 ± 0.017	0.056 ± 0.060	−0.293 ± 0.139

Table 10: Sensitivity of the released selector on the PBMC, MNC, and CBMC panels over five seeds. Checks count passed criteria out of 105 as in Table 5. OT pass counts runs out of 15 that meet OT Spearman ≥ 0.98 and scaled OT MAE ≤ 0.10 with raw-expression clusters as OT groups. OT MAE is mean ± sample SD over the 15 runs, and zero fill is the mean filled-zero fraction. Same output counts runs that return the default output.

Setting	Checks	OT pass	OT MAE	Zero fill	Same output
Default: $\alpha = 0.10$, four clusters	97	10	0.0102 ± 0.0075	0.103	15
Coefficients of $L(m)$					
weight of e_m ×0.5	97	10	0.0099 ± 0.0066	0.105	14
weight of e_m ×2	97	10	0.0101 ± 0.0075	0.102	14
fill weight 0.018 ×0.5	97	10	0.0101 ± 0.0075	0.102	14
fill weight 0.018 ×2	97	10	0.0102 ± 0.0075	0.103	15
fill rate 120 ×0.5	97	10	0.0101 ± 0.0075	0.102	14
fill rate 120 ×2	97	10	0.0102 ± 0.0075	0.103	15
OT error weight 0.012 ×0.5	97	10	0.0102 ± 0.0075	0.103	15
OT error weight 0.012 ×2	97	10	0.0102 ± 0.0075	0.103	15
rank weight 10^{-4} ×0.5	97	10	0.0102 ± 0.0075	0.103	15
rank weight 10^{-4} ×2	97	10	0.0102 ± 0.0075	0.103	15
offsets b_m ×0.5	97	10	0.0101 ± 0.0075	0.102	14
offsets b_m ×2	97	10	0.0102 ± 0.0075	0.103	15
offsets b_m removed	97	10	0.0101 ± 0.0075	0.102	14
Transport-order clusters					
three	98	10	0.0123 ± 0.0093	0.108	11
five	97	11	0.0101 ± 0.0075	0.106	14
six	96	11	0.0098 ± 0.0078	0.107	12
Support threshold					
$\alpha = 0.05$	96	11	0.0102 ± 0.0077	0.173	0
$\alpha = 0.20$	92	9	0.0104 ± 0.0081	0.082	0
$\alpha = 0.30$	93	10	0.0107 ± 0.0090	0.057	0

Simple geometry baselines. We compare GCRP with two simpler controls applied to the scVI, MAGIC, and ALRA residuals and to the unweighted mean of all proposal residuals. Each control keeps only non-negative corrections on observed zeros. The per-cell Euclidean cap scales the residual R_i of cell i by $\min(1, c\, r_i/\|R_i\|_2)$, where r_i is the distance from cell i to its 11th non-self neighbor in raw expression and $c \in \{0.25, 0.50, 1.0\}$. The cap uses no OT. Global shrinkage multiplies the whole residual by the largest factor in $\{1, 0.75, 0.5, 0.35, 0.25, 0.18, 0.12, 0.08, 0.05, 0.03, 0.02, 0.01, 0.005\}$ that passes the evaluation OT limit, and it returns X^{obs} if no factor passes. Global shrinkage is therefore tuned on the evaluation criterion and passes it in every run by construction. In one run per residual, no factor passes, and shrinkage returns X^{obs}.

Table 11 reports the results. GCRP passes 97 checks with a scaled OT error of 0.0102 and a filled-zero fraction of 0.103. The Euclidean caps pass 72 to 97 checks with OT errors of 0.0152 to 0.0544. The best cap, scVI with $c = 0.25$, matches GCRP at 97 checks and gives larger local-separation gains. However, it doubles the OT error, fills 94% of zeros, and passes the OT limit in 9 of 15 runs. Global shrinkage passes 85 to 95 checks with 1.8 to 3.0 times the OT error of GCRP and fills 65% to 93%

of zeros. Unconstrained residuals pass 70 to 81 checks. GCRP therefore matches the highest check count with the lowest OT error and the fewest filled zeros. It passes the evaluation OT limit in 10 of 15 runs without access to that limit.

Table 11: Simple geometry baselines on the PBMC, MNC, and CBMC panels over five seeds. Mean residual is the unweighted mean of all proposal residuals. Cap applies the per-cell Euclidean cap with multiplier c. Shrink applies the largest global factor that passes the evaluation OT limit. Columns follow Table 10. ΔkNN and Δsilhouette are mean differences from raw.

Residual	Control	Checks	OT pass	OT MAE	Zero fill	ΔkNN	Δsilhouette
GCRP	full method	97	10	0.0102 ± 0.0075	0.103	0.0043	0.0153
scVI	none	70	1	0.0988 ± 0.0795	0.938	0.0302	0.0855
scVI	cap, $c = 0.25$	97	9	0.0208 ± 0.0142	0.937	0.0150	0.0257
scVI	cap, $c = 0.50$	80	5	0.0354 ± 0.0242	0.937	0.0192	0.0438
scVI	cap, $c = 1.0$	77	3	0.0544 ± 0.0358	0.938	0.0257	0.0636
scVI	shrink	95	15	0.0310 ± 0.0350	0.874	0.0107	0.0387
MAGIC	none	71	5	0.0366 ± 0.0302	0.932	−0.0245	0.0493
MAGIC	cap, $c = 0.25$	87	9	0.0171 ± 0.0142	0.930	−0.0085	0.0325
MAGIC	cap, $c = 0.50$	72	5	0.0253 ± 0.0191	0.932	−0.0167	0.0459
MAGIC	cap, $c = 1.0$	72	5	0.0335 ± 0.0223	0.932	−0.0261	0.0517
MAGIC	shrink	85	15	0.0180 ± 0.0206	0.867	−0.0159	0.0303
ALRA	none	71	1	0.0756 ± 0.0510	0.702	0.0317	0.0839
ALRA	cap, $c = 0.25$	95	7	0.0157 ± 0.0115	0.701	0.0144	0.0280
ALRA	cap, $c = 0.50$	85	6	0.0271 ± 0.0211	0.701	0.0213	0.0456
ALRA	cap, $c = 1.0$	80	6	0.0403 ± 0.0264	0.701	0.0215	0.0639
ALRA	shrink	95	15	0.0288 ± 0.0319	0.650	0.0165	0.0415
Mean residual	none	81	5	0.0443 ± 0.0367	0.996	0.0147	0.0718
Mean residual	cap, $c = 0.25$	94	8	0.0152 ± 0.0106	0.996	0.0086	0.0310
Mean residual	cap, $c = 0.50$	87	10	0.0252 ± 0.0207	0.996	0.0123	0.0527
Mean residual	cap, $c = 1.0$	83	6	0.0433 ± 0.0369	0.996	0.0154	0.0715
Mean residual	shrink	93	15	0.0245 ± 0.0278	0.930	0.0111	0.0478

Real and shuffled neighborhood evidence. Removing the neighborhood evidence threshold changes both where zeros are filled and how many are filled. To separate these effects, a five-seed control keeps the number of supported zeros of each gene fixed and shuffles which cells receive them (Table 12). The fills in this control come from a learned combination of the base imputers. Real neighborhood evidence gives a higher mean OT MAE on all three panels. It passes the OT limit in four of five PBMC seeds and three of five MNC seeds, and shuffled evidence passes in all five seeds on both panels. The evidence threshold therefore controls OT error mainly through the number of zeros it admits.

Table 12: Real and shuffled neighborhood evidence over five matched seeds. Shuffling preserves each gene's supported-zero count. ΔkNN and Δsilhouette are differences from raw. OT metrics use biological groups. Values are mean ± sample SD. Pass is defined as in Table 2.

Panel	Evidence	ΔkNN	Δsilhouette	OT Spearman	OT MAE	Pass
PBMC	Real	−0.0077 ± 0.0109	0.0027 ± 0.0006	0.9927 ± 0.0108	0.0572 ± 0.0340	0.80
PBMC	Shuffled	−0.0050 ± 0.0104	0.0009 ± 0.0004	0.9952 ± 0.0066	0.0152 ± 0.0074	1.00
MNC	Real	0.0033 ± 0.0079	−0.0016 ± 0.0003	0.9830 ± 0.0138	0.0408 ± 0.0167	0.60
MNC	Shuffled	−0.0003 ± 0.0036	−0.0003 ± 0.0001	0.9952 ± 0.0066	0.0066 ± 0.0034	1.00
CBMC	Real	−0.0004 ± 0.0046	−0.0001 ± 0.0003	0.9984 ± 0.0019	0.0097 ± 0.0066	1.00
CBMC	Shuffled	−0.0002 ± 0.0020	−0.0004 ± 0.0002	0.9986 ± 0.0011	0.0075 ± 0.0084	1.00

Checks of Propositions 2 and 3. We rebuild the runs of Table 2 and the radius multipliers of Section 4 and embed each raw and output matrix in the fixed PCA-20 reference map. We sample 25 cells from each OT group and compute exact W_2 distances between groups. We evaluate two versions of the group budget $B^{(2)}_{u,a}$. The declared budget uses the radii r_{ia}. The realized budget replaces r_{ia} with the displacement of cell i, which is also a valid radius. An ordering is a triple (u, h, k) in which

h is closer to u than k is in raw expression. The six-group CBMC panel contributes 60 orderings per run, and the five-group MNC and PBMC panels contribute 30.

Table 13 reports the results. No group distance changes by more than either bound in any run, and the largest change is 35% of the realized bound. Without projection, the realized budgets are 69% of the median raw group distance and certify 21% of orderings. At the default radius, the declared budgets are 15% of the median raw group distance and certify 45% of orderings. The declared budgets depend only on raw expression and the radius, so the declared counts are identical for the three imputers. Every certified ordering keeps its raw order at every radius, and every reversed ordering lies below the certification margin.

The Crohn's disease states in Table 3 give a contrasting case, in which the bound is inconclusive but every ordering survives. We apply the same procedure with 80 cells per state and realized budgets, because the support-gated candidates have no declared radius. For $\lambda = 0.25$ and 0.50, the mean state budgets are 13% and 25% of the median raw state distance. The summed budgets exceed every raw margin, so no ordering is certified. However, all 15 orderings over five seeds keep their raw order at both values of λ, and each state distance changes by at most 7.3% of its Proposition 2 bound.

Table 13: Proposition checks for scVI, MAGIC, and ALRA on the CBMC, MNC, and PBMC panels over five seeds, with 45 runs and 1,800 raw orderings per row. Reversed counts orderings that flip in the output. Certified counts orderings that satisfy the condition of Proposition 3 under declared or realized budgets, and none of the certified orderings reverses. Max ratio is the largest pair-distance change divided by its realized Proposition 2 bound. Budget and pair change are means relative to the median raw group distance.

		Certified			Budget		
Radius multiplier	Reversed	Declared	Realized	Max ratio	Declared	Realized	Pair change
none	152	–	382	0.35	–	0.69	0.128
0.25	60	804	867	0.34	0.15	0.11	0.018
0.50	87	471	635	0.30	0.30	0.20	0.033
1.0	115	306	463	0.34	0.60	0.33	0.059

Larger panels. To test whether these results depend on panel size, we rerun the GCRP checks and the displacement projection with 300 cells per group and 2,000 genes over the same five seeds (Tables 14 and 15). The check panels contain 1,450 to 2,400 cells, about five times as many as in Table 5. GCRP passes 91 of 105 checks, compared with 97 on the smaller panels. Its mean scaled OT error is at most 0.029 on every panel, and it fills 5% to 11% of zeros. The failed checks are the OT Spearman limit in 8 runs and the kNN gain in 6 runs. The internal transport-order check uses at most four clusters, whereas the evaluation uses up to eight groups. The base imputers distort the geometry more on the larger panels, with mean scaled OT errors of 0.30 to 1.21. The projection lowers this error for every imputer on every panel, to 0.09 to 0.28, but it passes the OT limit in at most three of five seeds.

Table 14: GCRP checks on larger panels with 300 cells per group and 2,000 genes over five seeds. Checks count passed criteria out of 35 per panel as in Table 5. OT pass counts seeds that meet OT Spearman ≥ 0.98 and scaled OT MAE ≤ 0.10. Values are mean ± sample SD.

Panel	Cells	Checks	OT pass	OT Spearman	OT MAE	Zero fill
PBMC	1,450	32	2	0.964 ± 0.024	0.029 ± 0.009	0.066
MNC	1,500	28	2	0.966 ± 0.034	0.015 ± 0.011	0.113
CBMC	2,400	31	3	0.987 ± 0.014	0.005 ± 0.001	0.049

G Computational Cost Details

Because GCRP runs as a post-processing step, we evaluate its computational cost separately from the cost of fitting the candidate imputers. Once the candidate matrices are available, support gating and residual construction evaluate only observed zero entries, so their cost scales linearly with the number of these entries. We compute local support by looking up gene presence on the raw kNN graph, and we reuse this calculation across candidates. For K affine reference maps of dimensions d_a,

Table 15: Displacement projection on larger panels with 300 cells per group and 2,000 genes over five seeds (1,450 PBMC, 1,500 MNC, and 1,800 CBMC cells). CBMC has 1,800 cells here because this analysis keeps six groups, and the checks of Table 14 keep eight. Columns follow Table 2, with the default radius multiplier 0.25. Values are means over seeds.

		OT Spearman		OT MAE			
Panel	Candidate	Base	Projected	Base	Projected	Pass	$\bar{\gamma}$
PBMC	scVI	0.227	0.990	1.213	0.283	0.00	0.04
PBMC	MAGIC	0.476	0.990	1.119	0.168	0.00	0.03
PBMC	ALRA	0.350	0.993	1.208	0.166	0.20	0.03
MNC	scVI	−0.149	0.966	0.488	0.118	0.20	0.04
MNC	MAGIC	0.585	0.949	0.377	0.144	0.00	0.04
MNC	ALRA	0.348	0.971	0.449	0.131	0.00	0.04
CBMC	scVI	0.810	0.993	0.493	0.086	0.60	0.15
CBMC	MAGIC	0.859	0.998	0.351	0.110	0.60	0.17
CBMC	ALRA	0.919	0.990	0.296	0.123	0.40	0.16

the displacement checks apply the map loadings to the residuals. This process scales as $\sum_a O(npd_a)$ in the dense representation; sparse storage can reduce this cost when corrections remain sparse.

The primary computational expense occurs during population-level candidate selection because the algorithm must evaluate Sinkhorn distances across groups or subsamples. This cost scales with the number of candidates and the resolution of the groups. It does not scale with the full gene dimension. To measure scaling, we run the released selector on BMMC subsets of 1,000 to 20,000 cells with 2,000 genes and 72% observed zeros over five seeds (Table 16). The proposals are a 50-component SVD fill and a 20-neighbor graph mean, each at full strength and blended at 0.25 and 0.50, which gives 136 to 156 candidates. Each run uses 8 CPU cores and 64 GB of memory on a shared compute node. The GCRP runtime grows roughly linearly with the number of cells, from 61 s at 1,000 cells to 706 s at 20,000 cells, and the candidate loop accounts for 84% to 91% of this time. The large SD at 20,000 cells comes from one run with low CPU utilization on the shared node. Peak resident memory also grows linearly and reaches 14.0 GB at 20,000 cells. On the 400- and 640-cell panels of Table 5, one GCRP run takes 45 ± 16 s.

Table 16: Scaling of the released GCRP selector on BMMC subsets with 2,000 genes over five seeds. GCRP time and peak memory are mean ± sample SD. Candidate loop and proposal fitting are mean times. Proposal fitting is reported separately and excluded from GCRP time. Peak memory is the peak resident memory of the process.

Cells	GCRP time (s)	Candidate loop (s)	Proposal fitting (s)	Peak memory (GB)
1,000	61 ± 29	55	2.1	0.90 ± 0.00
2,500	99 ± 6	88	2.1	1.96 ± 0.01
5,000	182 ± 36	162	3.8	3.68 ± 0.01
10,000	351 ± 90	310	7.3	7.10 ± 0.01
20,000	706 ± 359	593	17.9	13.97 ± 0.04

H Scope and Boundary Cases

The bounds proven in this paper hold relative to a declared geometry; they do not guarantee that the protected geometry matches biological ground truth. This distinction limits the scope of the method in five ways.

Boundary 1: CITE-seq Validation Limits.
CITE-seq supplies an orthogonal validation signal, but it does not supply RNA-level ground truth. The available surface-protein panel remains small relative to the entire transcriptome, and the correspondence between RNA and protein is imperfect. We therefore treat ADT structure only as evidence that a selected correction preserves an independently measured aspect of cell state.

Boundary 2: Representation Dependence.
The geometry depends on the chosen representation and ground metric. Proposition 1 bounds

movement only within the reference maps supplied to GCRP. If a reference map omits biologically important variation, the projection cannot protect that variation. The sensitivity analysis over PCA dimensions varies only the dimension of an affine map, and it does not prove representation invariance. The reference maps on the log and square-root scales are nonlinear in expression, so Proposition 1 does not bound displacement in these two maps. Finally, the Wasserstein bound does not bound the entropic Sinkhorn costs or their rank order.

Boundary 3: The Anchor is Not an Oracle.
Raw RNA geometry acts as a stabilizing anchor, and the anchor can itself contain errors. For example, if dropout has already distorted the raw geometry, a conservative imputation method can preserve some of that distortion. GCRP prevents the unsupported rewriting of observed structure, but it does not claim that the observed structure is correct. The geometry experiments do not resolve whether preserving the raw anchor also preserves the distortions originally introduced by missing capture.

Boundary 4: Stage Validation.
The zebrafish stage-aware analysis uses observed stage labels. Any improvement achieved by using stage data during construction could therefore result from label leakage, so we exclude these results from the label-free evidence. A stronger validation would select candidates without stage information and evaluate them only on held-out stages, pseudotime bins, or lineage fate measurements.

Boundary 5: Non-negative Corrections Only.
The current formulation targets only non-negative corrections applied to observed zeros. Modifying observed nonzero entries is simple to implement, but doing so changes the error model. Correcting ambient RNA or false-positive counts requires independent evidence to support downward correction, just as dropout requires evidence for upward correction. We therefore leave the modification of nonzeros outside the scope of this study.

The evaluated panels show bounded geometric change; they do not establish the recovery of unobserved RNA values, lineages, or fate outcomes. Establishing recovery requires independent reference measurements and a separate evaluation framework.

I Existing Assets and Use Terms

External assets provide the candidate inputs, evaluation variables, and software dependencies. Table 4 identifies the specific analyses that use biological annotations to define OT groups. The CITE-seq and multimodal panels are public datasets credited through the CITE-seq and totalVI literature (Stoeckius et al., 2017; Gayoso et al., 2021). Progression, time-course, BMMC, and trajectory resources are credited through the corresponding single-cell trajectory literature (Lange et al., 2022; Schiebinger et al., 2019). Candidate generation and evaluation rely on publicly released or standard implementations of scVI, MAGIC, ALRA, SVD-based low-rank recovery, analyses following CellRank, and OT and Sinkhorn routines. We credit these tools to their original method papers (Lopez et al., 2018; van Dijk et al., 2018; Linderman et al., 2022; Lange et al., 2022; Cuturi, 2013; Peyré & Cuturi, 2019) and use them under their respective public release terms. This submission redistributes no third-party dataset, trained model, or software package. The PBMC, MNC, and CBMC panels use GEO accessions GSE100501 (sample GSM2685238), GSE128639 (sample GSM3681518), and GSE100866. The BMMC multiome data are GEO GSE194122, and the pseudotime is the `GEX_pseudotime_order` annotation distributed with them. The Crohn's disease data are the colon epithelial atlas of Kong et al. (2023), CELLxGENE dataset 63ff2c52-cb63-44f0-bac3-d0b33373e312. Its Enterocytes TMIGD1 MEP1A cells come from 16 healthy, 17 noninflamed, and 5 inflamed donors. Each seed samples cells across all donors, so the five seeds do not measure donor-level variation. The zebrafish time course is the CellRank zebrafish dataset (Lange et al., 2022).

J Additional Baselines and Boundary Cases

This appendix provides the quantitative data supporting the controlled-dropout, zebrafish stage-aware construction, held-out stage interpolation, BMMC pseudotime, and SERGIO boundary cases discussed in the main text.

Controlled dropout evaluation. We evaluate graph and SVD imputers under controlled corruption to test whether full imputation reduces expression error while failing the OT limit (Table 17). We

mask 10% of the observed positive entries and split the masked entries into two halves at random. A masked-recovery selector uses k-means clusters of raw expression and a fixed PCA embedding of the observed data, and it chooses the accepted projection with the lowest RMSE on the first half. Acceptance requires an OT Spearman correlation ≥ 0.98 and a relative OT MAE ≤ 0.10. Relative OT MAE is the mean absolute error of pair distances divided by the mean raw pair distance. In contrast, the released GCRP selector uses neither half and no labels. We report masked RMSE on the second half, which no method uses.

The selected projection passes the OT limit in all 15 runs. In 2 of these runs, no projection passes and the selector returns the raw matrix. Relative to raw, the projection lowers held-out masked RMSE by 10–17% across panels and by 13% pooled over the 15 runs, with a 95% CI of 9–17% and a paired t-test $p = 9.6 \times 10^{-6}$. The released selector lowers it by 4–7% across panels and by 5% pooled (95% CI 4–6%, $p = 6.9 \times 10^{-9}$). Against the uncorrupted reference, the projection lowers the reference MAE by 22% (95% CI 17–28%, $p = 9.2 \times 10^{-7}$). The released selector lowers it by 13–18% across panels and by 16% pooled (95% CI 14–18%, $p = 7.2 \times 10^{-11}$). Wilcoxon signed-rank tests give $p \leq 2.4 \times 10^{-4}$ for all four reductions. Most of these reductions come from restored expression scale, because the reference MAE compares absolute distances in a shared embedding. To remove this scale effect, the scale-free reference MAE divides each distance vector by its median before comparison. On this measure, GCRP matches raw data (paired change −0.0001, 95% CI −0.0059 to 0.0057), and the projection stays close to raw (+0.0006). In contrast, full graph imputation roughly triples the scale-free error (+0.036, 95% CI 0.021 to 0.051, $p = 1.4 \times 10^{-4}$, higher in 14 of 15 runs), and SVD raises it by a quarter on average ($p = 0.24$). The pooled change in reference Spearman correlation is −0.0037 (95% CI −0.0076 to 0.0002) for the projection and −0.0024 (95% CI −0.0070 to 0.0022) for the released selector. The kNN overlap of the projection with the reference exceeds raw on PBMC and CBMC and falls by 0.010 on MNC. The released selector keeps kNN overlap above raw on every panel and passes the evaluation OT limit in 10 of 15 runs. Its transport-order check can disagree with the evaluation because the check uses four clusters and the evaluation uses six.

Full graph and SVD imputation reach a lower masked RMSE and reduce the reference MAE by 48–66%, but all 30 of their runs fail the OT limit. Their reference Spearman correlation is lower than that of the projection on every panel and falls to 0.920 for graph imputation on MNC. They also reduce kNN overlap with the reference in five of six combinations of panel and method. The OT limit is defined relative to the observed data, so failing it does not by itself show poorer recovery of the uncorrupted geometry.

Table 17: Controlled-dropout evaluation across PBMC, MNC, and CBMC. Values are mean ± sample SD across five seeds. Masked RMSE is evaluated on the half of the masked entries that no method uses. Projection is chosen by masked recovery on the other half. GCRP is the released selector, which uses neither the masked entries nor labels. kNN overlap is the neighbor overlap with the uncorrupted reference embedding. Scale-free ref. MAE divides each distance vector by its median before comparison. Ref. Spearman and Ref. MAE are the rank correlation and relative pair-distance error against the uncorrupted reference. Pass counts the seeds out of five that meet OT Spearman ≥ 0.98 and relative OT MAE ≤ 0.10.

Panel	Method	Masked RMSE	kNN Overlap	OT Spearman	Relative OT MAE	Ref. Spearman	Ref. MAE	Scale-free ref. MAE	Pass
PBMC	raw	0.749 ± 0.004	0.636 ± 0.006	1.000 ± 0.000	0.000 ± 0.000	0.993 ± 0.011	0.220 ± 0.004	0.010 ± 0.002	5/5
PBMC	graph	0.500 ± 0.006	0.586 ± 0.008	0.907 ± 0.064	0.179 ± 0.008	0.983 ± 0.011	0.076 ± 0.005	0.034 ± 0.005	0/5
PBMC	svd	0.532 ± 0.006	0.549 ± 0.009	0.939 ± 0.027	0.217 ± 0.004	0.985 ± 0.013	0.074 ± 0.003	0.014 ± 0.004	0/5
PBMC	projection	0.673 ± 0.035	0.636 ± 0.009	0.988 ± 0.005	0.065 ± 0.019	0.990 ± 0.013	0.172 ± 0.019	0.012 ± 0.004	5/5
PBMC	GCRP	0.719 ± 0.003	0.641 ± 0.006	0.979 ± 0.009	0.047 ± 0.004	0.983 ± 0.011	0.182 ± 0.005	0.013 ± 0.004	2/5
MNC	raw	0.761 ± 0.004	0.639 ± 0.008	1.000 ± 0.000	0.000 ± 0.000	0.978 ± 0.018	0.223 ± 0.004	0.011 ± 0.002	5/5
MNC	graph	0.514 ± 0.004	0.588 ± 0.002	0.903 ± 0.077	0.158 ± 0.016	0.920 ± 0.063	0.106 ± 0.013	0.037 ± 0.005	0/5
MNC	svd	0.547 ± 0.005	0.549 ± 0.007	0.931 ± 0.057	0.193 ± 0.021	0.930 ± 0.054	0.085 ± 0.010	0.026 ± 0.005	0/5
MNC	projection	0.662 ± 0.059	0.628 ± 0.008	0.991 ± 0.009	0.070 ± 0.039	0.971 ± 0.018	0.174 ± 0.029	0.017 ± 0.007	5/5
MNC	GCRP	0.728 ± 0.004	0.643 ± 0.006	0.989 ± 0.009	0.051 ± 0.002	0.981 ± 0.018	0.184 ± 0.005	0.012 ± 0.004	4/5
CBMC	raw	0.765 ± 0.003	0.435 ± 0.014	1.000 ± 0.000	0.000 ± 0.000	0.996 ± 0.003	0.231 ± 0.006	0.030 ± 0.021	5/5
CBMC	graph	0.415 ± 0.006	0.461 ± 0.008	0.963 ± 0.025	0.247 ± 0.025	0.977 ± 0.021	0.082 ± 0.009	0.087 ± 0.039	0/5
CBMC	svd	0.478 ± 0.008	0.418 ± 0.010	0.979 ± 0.024	0.120 ± 0.011	0.986 ± 0.007	0.120 ± 0.020	0.023 ± 0.006	0/5
CBMC	projection	0.638 ± 0.072	0.449 ± 0.005	0.995 ± 0.007	0.059 ± 0.033	0.995 ± 0.003	0.179 ± 0.030	0.024 ± 0.013	5/5
CBMC	GCRP	0.711 ± 0.008	0.444 ± 0.014	0.986 ± 0.025	0.034 ± 0.013	0.996 ± 0.004	0.201 ± 0.016	0.025 ± 0.013	4/5

Matched distortion. To compare GCRP with simpler controls at equal distortion, we apply global shrinkage with factors from 0.05 to 1 and per-cell Euclidean caps with multipliers from 0.05 to 2 to the SVD and graph residuals in the same runs (Table 18). For each run and control, we keep the members whose relative OT MAE does not exceed that of GCRP. We then take the best such member

separately for each metric, which favors the control. The controls reach a lower masked RMSE than GCRP in 12 to 15 of 15 runs, by 0.018 to 0.036 on average. However, GCRP has a lower reference MAE than the best matched member of each shrinkage and cap control in 12 to 15 of 15 runs, with paired t-test $p \leq 1.4 \times 10^{-3}$. It also has a lower reference MAE than the best matched projection in all 11 runs that have one, and it fills fewer zeros than every control. On the scale-free reference MAE, the best matched members of the controls reach 0.012 to 0.015, and GCRP reaches 0.017, the level of raw data. The absolute advantage of GCRP therefore reflects restored expression scale. At matched distortion, shrinkage preserves the relative state geometry slightly better than GCRP ($p \leq 0.014$), and the caps match GCRP.

Table 18: Controls at matched distortion under controlled corruption over PBMC, MNC, and CBMC and five seeds. For each run, a control member is eligible when its relative OT MAE is at most that of GCRP. Masked RMSE and reference MAE take the best eligible member of the control for each metric, and zero fill belongs to the member with the lowest masked RMSE. Scale-free ref. MAE divides each distance vector by its median. Runs counts the runs with an eligible member. Lower ref. MAE counts the runs in which GCRP has the lower reference MAE. Values are means over runs.

		Masked RMSE		Ref. MAE		Scale-free ref. MAE			Zero fill	
Control	Runs	Control	GCRP	Control	GCRP	Control	GCRP	Lower ref. MAE	Control	GCRP
Shrinkage, SVD	15	0.693	0.720	0.197	0.189	0.012	0.017	12	0.81	0.37
Shrinkage, graph	15	0.686	0.720	0.197	0.189	0.013	0.017	15	0.76	0.37
Cap, SVD	15	0.702	0.720	0.202	0.189	0.015	0.017	15	0.81	0.37
Cap, graph	15	0.693	0.720	0.203	0.189	0.014	0.017	15	0.76	0.37
Projection grid	11	0.685	0.721	0.190	0.184	0.013	0.014	11	0.68	0.42

Deployable controls. A user without masked entries needs a label-free rule to set the strength of a control. We therefore let global shrinkage and per-cell caps set their strength by the same label-free geometry check that accepts projections. In each run, a control uses the largest strength whose output meets OT Spearman ≥ 0.98 and relative OT MAE ≤ 0.10 on raw-expression clusters, and it returns raw data when no strength passes (Table 19). GCRP has the highest mean rank agreement with the uncorrupted reference, 0.987 against 0.977 to 0.984. Its kNN overlap with the reference exceeds that of SVD shrinkage and SVD caps in all 15 runs ($p \leq 1.4 \times 10^{-4}$). It keeps the relative state geometry closer to the reference than graph shrinkage ($p = 0.016$) and matches the other three controls. GCRP fills 37% of zeros, and the controls fill 64% to 81%. The controls recover more masked expression, with held-out RMSE of 0.643 to 0.667 against 0.720. GCRP fills 73% of the held-out masked entries, and its fills average 6% of the true value. The per-cell displacement radius is the main limit on this fill size.

Table 19: Deployable controls under controlled corruption over PBMC, MNC, and CBMC and five seeds. In each run, each control uses the largest strength that passes the label-free geometry check. Strength is the mean shrinkage factor or cap multiplier. Scale-free ref. MAE divides each distance vector by its median. kNN overlap is the neighbor overlap with the uncorrupted reference embedding. Values are means over 15 runs.

Method	Strength	Held-out RMSE	Ref. MAE	Scale-free ref. MAE	Ref. Spearman	kNN overlap	Zero fill
GCRP	–	0.720	0.189	0.017	0.987	0.576	0.37
Shrinkage, SVD	0.43	0.643	0.175	0.015	0.981	0.559	0.81
Shrinkage, graph	0.34	0.652	0.182	0.024	0.977	0.576	0.76
Cap, SVD	0.25	0.665	0.191	0.017	0.984	0.565	0.81
Cap, graph	0.13	0.667	0.193	0.021	0.980	0.577	0.64

Zebrafish stage-aware construction. We report the stage-aware CellRank zebrafish analysis separately because its construction uses stage labels. The unrestricted variant uses 21 non-self neighbors, and the same-stage variant uses 20. Because the two variants differ in both label use and neighbor count, we treat the comparison as a diagnostic. We find that supplying stage information increases kNN accuracy relative to raw data (Table 20). Higher amplitudes also remain feasible and pass the OT limit when local support is restricted to same-stage neighbors.

RNA–protein concordance. The CBMC data measure 13 surface proteins, and the genes encoding 6 of them are among the 5,000 RNA genes: CD3 (CD3E), CD4, CD45RA (PTPRC), CD16 (FCGR3A), CD14, and CCR7. We add these six genes to the CBMC panel of Table 5. For each protein, we

Table 20: CellRank zebrafish stage-aware diagnostic (1,157 cells, 1,200 genes, 11 stages). Values are paired within-seed changes relative to raw, reported as mean ± sample SD across five seeds. All six variants pass the OT limit (≥ 0.98 Spearman, ≤ 0.10 MAE) in all five seeds.

Neighbor Source	Amplitude	ΔkNN	Δsilhouette	OT Spearman	OT MAE
Unrestricted	0.25	−0.002 ± 0.003	0.002 ± 0.000	0.999 ± 0.000	0.006 ± 0.003
Unrestricted	0.50	−0.004 ± 0.003	0.003 ± 0.000	0.999 ± 0.000	0.010 ± 0.004
Unrestricted	1.00	−0.012 ± 0.004	0.006 ± 0.001	0.998 ± 0.000	0.017 ± 0.005
Same-stage	0.25	0.020 ± 0.004	0.003 ± 0.000	1.000 ± 0.000	0.005 ± 0.001
Same-stage	0.50	0.038 ± 0.005	0.007 ± 0.000	0.999 ± 0.001	0.009 ± 0.002
Same-stage	1.00	0.075 ± 0.007	0.014 ± 0.000	0.998 ± 0.000	0.017 ± 0.002

compute the Spearman correlation across cells between its protein level after centered log-ratio transformation and the RNA expression of its gene (Table 21). No imputer and no selector uses the proteins. On raw data, the mean correlation is 0.247. GCRP raises it to 0.288 in all five seeds (paired change +0.041, 95% CI 0.032 to 0.050) while filling 3% of zeros. ALRA raises it most, to 0.373, and its projection keeps most of this gain at 0.359. MAGIC and DCA change it little. In contrast, scVI and SAVER lower it to 0.146 and 0.160 in all five seeds. The displacement projection recovers part of the scVI loss, reaching 0.182, and leaves the SAVER value unchanged.

Table 21: RNA–protein concordance on the CBMC panel (640 cells) over five seeds. Each protein column is the Spearman correlation across cells between the protein and its gene, averaged over seeds. Mean is the average over the six proteins with its sample SD over seeds. Change is the paired change of the mean from raw with its 95% t interval. Projected outputs use the default radius multiplier 0.25.

Method	Mean	Change	CD3	CD4	CD45RA	CD16	CD14	CCR7	Zero fill
raw	0.247 ± 0.011	–	0.442	0.107	0.020	0.523	0.351	0.042	0.00
GCRP	0.288 ± 0.008	+0.041 [0.032, 0.050]	0.515	0.155	−0.004	0.552	0.470	0.042	0.03
scVI	0.146 ± 0.014	−0.102 [−0.129, −0.075]	−0.070	0.302	0.156	0.260	0.334	−0.107	0.94
scVI projected	0.182 ± 0.011	−0.066 [−0.090, −0.041]	0.050	0.304	0.054	0.285	0.386	0.012	0.94
MAGIC	0.260 ± 0.016	+0.013 [−0.015, 0.041]	0.173	0.398	0.050	0.368	0.379	0.195	0.94
MAGIC projected	0.257 ± 0.014	+0.010 [−0.015, 0.035]	0.188	0.368	0.037	0.369	0.378	0.204	0.94
ALRA	0.373 ± 0.006	+0.126 [0.107, 0.145]	0.436	0.494	0.142	0.596	0.489	0.084	0.58
ALRA projected	0.359 ± 0.007	+0.111 [0.094, 0.128]	0.430	0.481	0.086	0.600	0.485	0.070	0.58
DCA	0.262 ± 0.017	+0.014 [−0.015, 0.044]	0.131	0.456	0.052	0.365	0.369	0.199	0.94
DCA projected	0.260 ± 0.018	+0.013 [−0.016, 0.041]	0.164	0.422	0.017	0.375	0.364	0.217	0.94
SAVER	0.160 ± 0.012	−0.087 [−0.110, −0.065]	0.054	0.350	0.175	0.249	0.171	−0.039	0.94
SAVER projected	0.159 ± 0.011	−0.088 [−0.108, −0.069]	0.123	0.294	0.032	0.234	0.283	−0.013	0.94

Held-out stage interpolation. To test progression inference against held-out truth, we hold out each interior stage of the zebrafish time course and predict it from the two neighboring stages (Table 22). Each seed samples the 1,157 cells of Table 20. For each held-out stage, we subsample 50 cells from it and from each neighboring stage. We compute the exact OT coupling between the two neighboring stages in a PCA-30 embedding fitted to each output and interpolate each coupled pair at the held-out stage time. The error ratio is the W_2 distance between the interpolated and held-out cells divided by the W_2 distance between the nearer neighboring stage and the held-out cells, averaged over five subsamples and nine held-out stages. No compared method uses stage labels. On raw data, the error ratio is 0.887, and OT coupling lowers the error by 6% relative to random coupling. Every label-free fill changes the ratio by less than 1%. Neighbor-mean and support-gated fills lower it by 0.2% to 0.7%, and GCRP lowers it by 0.2% in all five seeds. In contrast, full SVD imputation raises it by 0.4% in all five seeds. At matched OT distortion, GCRP has a lower error than SVD fills after shrinkage in all five seeds and a higher error than neighbor-mean fills after shrinkage in four seeds. On this time course, imputation changes the held-out interpolation error only slightly, and GCRP does not degrade it.

BMMC pseudotime and SERGIO limits. GCRP bounds distortion relative to the observed geometry, so it protects cell-state progression to the extent that the observed geometry reflects that progression. In the BMMC pseudotime evaluation over five seeds, the support-gated variant raises the mean silhouette score from 0.026 for raw expression to 0.032 (Table 23). The Spearman correlation between bin distances and pseudotime separation varies between seeds for every candidate because each seed samples different cells, so we compare candidates within each seed. Within each seed, the

Table 22: Held-out stage interpolation on the CellRank zebrafish time course over five seeds (1,157 cells, 1,200 genes, 11 stages, nine held-out stages). The error ratio is the W_2 distance between OT-interpolated and held-out cells divided by that from the nearer neighboring stage, and lower is better. Change is the paired change from raw with its 95% t interval. OT MAE is the scaled Sinkhorn error of stage distances relative to raw. Support-gated fills use $\alpha = 0.10$. Values are mean ± sample SD.

Method	Error ratio	Change ($\times 10^{-3}$)	OT MAE	Zero fill
raw	0.887 ± 0.009	–	0.000	0.00
neighbor mean	0.882 ± 0.010	−5.4 [−6.9, −3.8]	0.016	0.67
SVD	0.891 ± 0.009	3.6 [2.3, 4.9]	0.018	0.71
support-gated, $\lambda = 0.25$	0.886 ± 0.009	−1.4 [−1.9, −0.9]	0.006	0.42
support-gated, $\lambda = 0.50$	0.884 ± 0.009	−3.3 [−4.0, −2.6]	0.010	0.42
support-gated, $\lambda = 1.0$	0.881 ± 0.009	−5.9 [−7.5, −4.3]	0.017	0.42
GCRP	0.886 ± 0.009	−1.8 [−2.2, −1.4]	0.009	0.35

support-gated variant changes this correlation by -0.021 ± 0.042 relative to raw. SVD imputation lowers it more, with a within-seed change of -0.109 ± 0.123. The SERGIO simulation tests recovery under extreme dropout, reaching 78% to 92% mean observed zeros (Table 24). SVD reduces mean dropout RMSE by 2.2% to 4.4% relative to raw. However, geometry recovery remains unstable between seeds. We observe a boundary where extreme dropout leaves insufficient structure for consistent recovery.

Table 23: BMMC pseudotime evaluation over five seeds (2,700 cells, 1,000 genes, six pseudotime bins). Each seed samples 450 cells per bin. Pseudotime-distance ρ measures correlation between pairwise bin distances and bin-index separation. Values are mean ± sample SD.

Candidate	Bin kNN Acc.	Silhouette	Pseudotime ρ	Far/Adjacent Ratio
raw	0.754 ± 0.007	0.026 ± 0.003	0.561 ± 0.243	1.201 ± 0.126
SVD	0.729 ± 0.008	0.023 ± 0.003	0.452 ± 0.197	1.178 ± 0.114
neighbor mean	0.737 ± 0.010	0.040 ± 0.003	0.429 ± 0.225	1.185 ± 0.124
support-gated, $\alpha = 0.30$, $\lambda = 0.25$	0.752 ± 0.008	0.032 ± 0.003	0.540 ± 0.233	1.192 ± 0.130

Table 24: SERGIO simulation boundary case (360 cells, 100 genes, three simulated states). Dropout percentiles of 45, 65, and 82 give 78%, 85%, and 92% observed zeros. Values are mean ± sample SD across five seeds. Sinkhorn metrics use the simulated expression before dropout as the reference.

Dropout percentile	Candidate	Dropout RMSE	State kNN	OT Spearman	Scaled OT MAE
45.0	raw	0.493 ± 0.016	0.650 ± 0.155	0.200 ± 0.671	0.223 ± 0.189
45.0	SVD	0.471 ± 0.020	0.615 ± 0.146	0.300 ± 0.758	0.116 ± 0.096
65.0	raw	0.500 ± 0.010	0.624 ± 0.119	−0.100 ± 0.548	0.269 ± 0.208
65.0	SVD	0.484 ± 0.011	0.589 ± 0.122	−0.300 ± 0.758	0.159 ± 0.080
82.0	raw	0.506 ± 0.006	0.560 ± 0.136	−0.200 ± 0.671	0.213 ± 0.089
82.0	SVD	0.495 ± 0.005	0.563 ± 0.119	0.100 ± 0.548	0.193 ± 0.137