SELCELL: SELF-CERTIFYING CELL VELOCITY FROM DESTRUCTIVE SNAPSHOTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatial transcriptomics destroys the tissue it measures, so every cell is observed exactly once: cell velocity—where cells are going—must be inferred from a single snapshot and can never be checked against a later observation. Current practice reads RNA velocity as cell motion, an implicit assumption that real data do not always support: coherent arrows are drawn even on embryonic tissue where the motion signal is indistinguishable from a shuffled control, and existing confidence scores do not separate a method's reliable cells from the rest. We replace the assumption with an explicit observation model that separates a cell's intrinsic expression change from the change caused by its motion across the tissue, and invert it in closed form—training-free, compatible with any RNA-velocity backbone. Three per-cell certificates then follow analytically, each answering a different question: *where* to trust the recovered arrows, *how much* (direction-wise error bars, order-optimal for the noise they certify), and *what* is being measured (motion, or a wave of expression sweeping past still cells—bounded by planting synthetic waves). On real tissue the error bars are calibrated and remain valid on the hardest cells where split-conformal bands collapse; ranking cells by the validity certificate is the only ranking we found that improves the kept cells. On certified tissue SELCELL outperforms the strongest baseline (CBDir 0.232 vs 0.183, non-overlapping intervals) while abstaining where no motion signal exists. Code, raw-read rebuild pipelines, and the preregistered, hash-sealed protocols are in the supplementary material. We believe this recipe—replace an unchecked assumption with an explicit observation model, then certify what its inversion returns—extends to any estimand that can never be observed twice.

1 INTRODUCTION

Spatial transcriptomics measures a tissue by destroying it, so every cell is observed exactly once. Yet the questions such atlases are built for—morphogenesis, immune-cell maturation, cancer invasion—concern *cell velocity*: where cells are going. The standard tool for recovering time from a snapshot is RNA velocity (La Manno et al., 2018; Bergen et al., 2020), an estimate, from unspliced and spliced transcript ratios, of how each cell's expression profile is changing—a change of *expression*, not a movement of the cell.

Recent spatially aware methods take this remaining step with a projection rule: RNA-velocity vectors are matched to spatial displacements, so that a change of expression is read as physical motion ($v \approx J^\top w$) (Gu et al., 2026). But changing expression and moving through the tissue are different causes that can leave the same observation—a gap the newest spatial estimators now acknowledge, disclaiming that their projected vectors imply physical migration (Raghavan et al., 2026; Liu et al., 2026), while no method measures where the two can be told apart. As a consequence, the recovered arrows come with no credentials: the field is reported without a statement of where it can be trusted, by how much, or whether what it shows is motion at all—and the arrows do not stop where the assumption breaks (§4.3). The distinction at stake fits in one sentence: spatial gradients of expression tell us in which direction expression changes (Chitra et al., 2025); the question here is in which direction cells move—and where that can be told from expression at all.

Nor is the gap closed from the representation side: single-cell foundation models and their spatially aware successors (Cui et al., 2024; Theodoris et al., 2023; Tejada-Lapuerta et al., 2025; Birk et al.,

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.

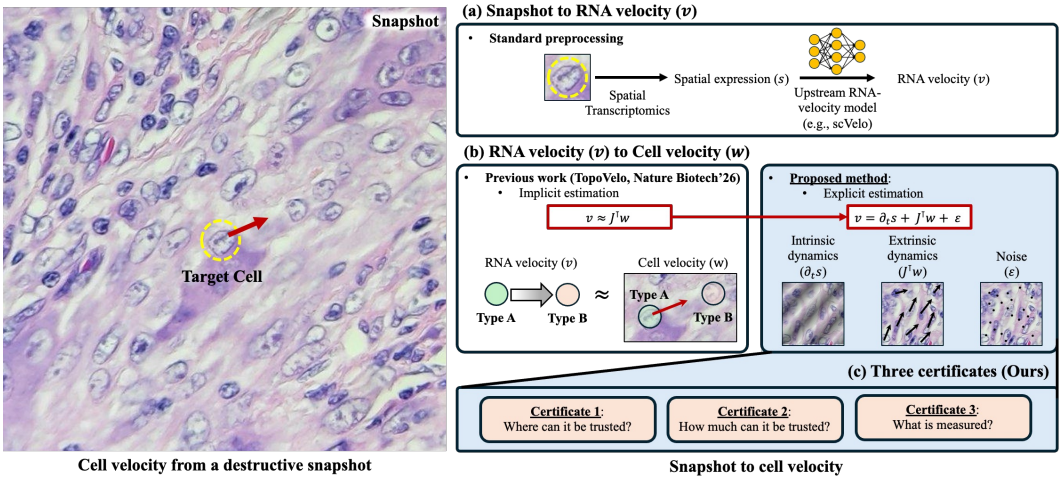


Figure 1: **RNA velocity is not cell velocity: prior work reads v as $J^\top w$; SelCell inverts $v = \partial_t s + J^\top w + \varepsilon$ and certifies what it returns.** Left: the snapshot destroys the tissue, so the target cell's velocity (red arrow) must be recovered from a single time point. (a) Standard preprocessing yields per-cell expression s and RNA velocity v; nothing in this step measures motion. (b) The usual assumption ($v \approx J^\top w$) reads expression change as motion; our observation model names the three causes that produce the same v—intrinsic dynamics, extrinsic (motion-driven) dynamics, noise—only the second carries the cell's motion. (c) Three certificates make the inversion self-auditing—where to trust, how much, whether what was measured is motion at all—abstaining on failure (§3).

2026) keep the *order* of neighbors and discard the *metric* structure—distances and directions—on which any observation model of motion must stand (Appendix D); scaling such models does not narrow the gap.

We replace the implicit assumption with an explicit observation model: instead of reading expression change as motion, we write down what can change expression in the first place (Fig. 1; §3). Once the model is written, the usual assumption stops being an inherited premise and becomes a testable special case. And inverting the model yields more than an estimate: three per-cell certificates follow analytically, answering the three questions above—where the recovered arrows can be trusted (only where motion beats a shuffled control), how much (direction-wise error bars), and what is being measured (cell motion, or a wave of expression sweeping past still cells). The chain that consumes them—gate, error bars, reporting license, abstention—is SELCELL (§3), training-free on top of any RNA-velocity backbone.

Contributions.

- *Where to trust:* a validity certificate derived from the observation model—not a learned confidence—selects reliable cells, and transfers unchanged across tissues.
- *How much to trust:* error bars are derived, not calibrated post hoc, and verified on real tissue—including where distribution-free bands break down.
- *What is measured:* the silence of a detector whose sensitivity is gauged by planting waves of known size turns an untestable assumption into a measured upper bound.
- *And abstention:* when a certificate fails, the field is not reported; performance is claimed only on certified data.

What this paper asks for is not better arrows, but arrows with credentials.

2 Related Work

Spatial RNA velocity. SIRV and KSRV transfer velocities from dissociated references (Abdelaal et al., 2024; He et al., 2025); spVelo, STT, and probabilistic models refine the expression-dynamics side (Long et al., 2025; Zhou et al., 2024; Gayoso et al., 2024); the newest spatially informed estimators disclaim the motion reading of their vectors outright (Raghavan et al., 2026; Liu et al., 2026). TopoVelo (Gu et al., 2026)—an RNA-velocity ODE coupled to a graph neural network, projected onto tissue by a cosine-similarity vote—is the closest antecedent and the baseline we reproduce and benchmark against (§4). All of them report a field; none states an observation model for the step from expression change to motion, so none comes with an error law for that step, or a rule for abstaining when it fails.

Lineages the certificates draw on. The advective part of our observation model, $v = J^\top w$, is brightness constancy (Horn & Schunck, 1981) with G gene channels, and its inversion and error law are classical (structure tensor, scalar velocimetry, ocean currents—full lineage in Appendix D, revisited on radar-measured currents in Appendix C.5). What is new is the regime—$G \gg 2$ channels with a per-cell validity statistic rather than a handful of channels with regularization—and the certificates built on it: certificate 1 is selective prediction (Geifman & El-Yaniv, 2017) with a score derived from the observation model rather than learned; certificate 2 makes a distribution-level promise of the kind conformal prediction makes (Vovk et al., 2005; Angelopoulos & Bates, 2023), and we compare against split conformal directly (§4.3); certificate 3 is a negative-control design (Lipsitch et al., 2010), calibrated by injecting waves of known size so that silence carries information.

Extended related work is in Appendix D.

3 Method: three certificates for cell velocity, and their consumption

We propose three *certificates*, earned by making the relationship between RNA velocity and cell velocity explicit as an observation model. Each answers one question with a statistic and a guarantee: *where* to trust (certificate 1), *how much* (certificate 2), and *what* is actually measured (certificate 3). §3.1 sets up the observation model and the prior-free cell-velocity inversion $\hat{w}$; §3.2–3.4 introduce the three certificates; §3.5 spends them to obtain the final cell-velocity estimator $w^\star$. We call this chain—the three certificates and their consumption—SelCell; Fig. 2 previews the chain and certificates 1–2 as per-cell maps. Numbered statements (Proposition k) are our guarantees for certificate k; classical results are stated in prose with their sources; every guarantee is recalled, measured, and judged in §4.3, and *certificate* always means a statistic carrying a stated guarantee whose assumptions and scope—asymptotic where noted—are part of the statement.

3.1 Setup and the prior-free cell-velocity inversion

What a snapshot provides. For each cell, the snapshot delivers a position $x \in \mathbb{R}^2$ and an expression profile $s \in \mathbb{R}^G$, whose spliced/unspliced split yields the upstream RNA velocity $v \in \mathbb{R}^G$—the rate at which *that cell's own* expression is changing (La Manno et al., 2018; Bergen et al., 2020); we take v as given, its error entering below through ε. What no snapshot provides is w, the cell's physical motion: no time axis, and no cell observed twice.

Notation. At each cell and its local window, the spatial gradients of the z-normalized expression $s \in \mathbb{R}^G$ form the columns of the local Jacobian $J \in \mathbb{R}^{2\times G}$ (column $a = \nabla s_a$ for gene a; 2 is the tissue-coordinate dimension), and $v \in \mathbb{R}^G$ denotes the locally averaged RNA velocity. We write $M := JJ^\top \in \mathbb{R}^{2\times 2}$ and drop the cell index. In deployment J is estimated by kernel-weighted local-linear regression (Appendix C.3); the theory is stated in the true-J convention, and the misalignment part of the estimation error acts conservatively on the certificates, while a scale error rescales magnitudes only (Appendix A.9).

Observation model. What can change a cell's RNA readout between snapshots? Three biological causes and one instrumental one: (i) the cell changes its own expression *in place* (reaction, r); (ii) the cell *moves* through a spatially varying expression landscape (advection, migration velocity $w \in \mathbb{R}^2$);

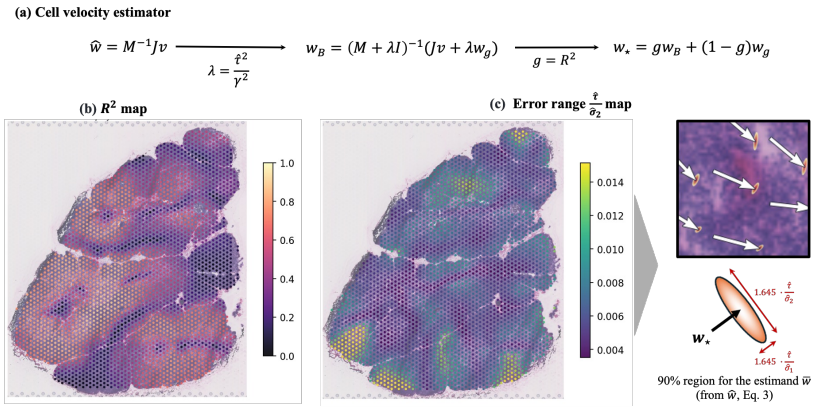


Figure 2: **The SelCell chain, and its certificates as per-cell maps (thymus S1).** (a) The chain: $\hat{w} = M^{-1}Jv$; certificate 2 spent as the shrinkage $\lambda = \hat{\tau}^2/\gamma^2$ toward w_g (Eq. (4)), certificate 1 as the gate $g = R^2$; output $w^\star$. (b) Certificate 1 as a map: per-cell R^2 on the H&E image—read only against a cell-shuffled control (§3.2). (c) Certificate 2 as a map: the error range $\hat{\tau}/\hat{\sigma}_2$ along each cell's least visible direction. Zoom: $w^\star$ arrows, each carrying per-direction 90% bars for the estimand $\bar{w}$, minted at the measurement $\hat{w}$ (Eq. (3)) and carried to $w^\star$ by an observable offset (§3.5). (Certificate 3 is a tissue-level verdict, §3.4.)

(iii) an expression *wave* sweeps the tissue—spatially coordinated in-place change indistinguishable from movement (phase velocity $c \in \mathbb{R}^2$; e.g., the segmentation clock, Palmeirim et al., 1997); and (iv) measurement noise $\varepsilon \sim (0, \tau^2 I_G)$. The chain rule ties these into $v = \partial_t s + J^\top w + \varepsilon$: the intrinsic term $\partial_t s$—the change of the landscape at a fixed position, never itself observed—carries causes (i) and (iii). Any method that reads motion from a snapshot must take a stance on the intrinsic term—the identity above leaves no third option; the graph-projection lineage takes it implicitly, matching RNA-velocity vectors to spatial displacements by a cosine-similarity *vote* that silently reads whatever intrinsic change v contains as motion (every baseline of §4 inherits this). We state the stance instead, in a form weaker than the usual $v \approx J^\top w$: separating the intrinsic causes needs no extra assumption, because $\partial_t s$ splits uniquely by the orthogonal decomposition along $\mathrm{range}(J^\top)$:

$$\partial_t s \;=\; -J^\top c \;+\; r, \qquad Jr = 0$$

(closed forms and proofs in Appendix A.1). Here c collects *all* intrinsic dynamics that can masquerade as motion—cause (iii), gradient-aligned differentiation included—and $r \in \mathbb{R}^G$ is incoherent with the gradients by construction rather than by assumption. Substitution yields the working model:

$$v \;=\; J^\top w \;-\; J^\top c \;+\; r \;+\; \varepsilon. \tag{1}$$

Throughout the estimation chain we assume only two things: *local linearity*—within the window, s is first-order and the expression landscape is quasi-stationary at observation time—and *isotropic noise*, $\varepsilon \sim (0, \tau^2 I_G)$ independent across genes. No assumption is placed on r: it is cancelled exactly by $Jr = 0$, and its magnitude is measured from the residual. Certificate-specific conditions are stated where they are used (Proposition 1, §3.4).

Remark 1 (Estimand). Shifting (w, c) to $(w+\delta, c+\delta)$ leaves (1) unchanged for any constant $\delta \in \mathbb{R}^2$, so a single snapshot identifies only $\bar{w} := w - c$—an algebraic fact no statistical refinement can lift (Appendix A.1); what can be said about c is certificate 3's subject (§3.4), and an orthogonal-design synthetic exhibits the algebra in isolation (Appendix C.8).

The prior-free cell-velocity inversion. We estimate $\bar{w}$—the identifiable part of the cell's velocity (Remark 1)—by least squares, $\hat{w} := \arg\min_w \|v - J^\top w\|^2$, taken deliberately without any prior so that the certificates built on it grade the data and not the prior:

$$\hat{w} \;=\; M^{-1}Jv, \qquad \hat{r} \;:=\; v - J^\top \hat{w}. \tag{2}$$

The whitening factor M^{-1} is standard least squares; unlike the cosine vote, it removes the metric bias of an anisotropic landscape (steep one way, shallow the other). And the inversion lands

exactly on the estimand: the allocation identity $\hat{w} = \bar{w} + M^{-1}J\varepsilon$ holds *exactly* (Lemma A.1, Appendix A.1)—r is cancelled by $Jr = 0$, so this is not a "no reaction" assumption but an identity of the decomposition (1). The residual $\hat{r}$ exists because the G gene equations of (1) outnumber the two unknowns, and it supplies the certificates' raw material (R^2, $\hat{\tau}^2$): the surplus of equations is what the certificates are made of.

3.2 Certificate 1 — where to trust (R^2)

Statistic. R^2 is the share of the observed RNA velocity that the fit $J^\top \hat{w}$ reproduces,

$$R^2 \;:=\; \frac{\|J^\top \hat{w}\|^2}{\|v\|^2} \;=\; 1 - \frac{\|\hat{r}\|^2}{\|v\|^2},$$

the rest being the residual. It is spent twice: as a discrete verdict—fail the shuffled control, *abstain* (Remark 2)—and, where certified, as the continuous gate $g = R^2$ of §3.5, warranted by Proposition 1.

Remark 2 (No absolute thresholds; ours). Even pure noise explains about $2/G$ by chance (Appendix A.2), and correlated real tissue sits well above that. R^2 is therefore never read as an absolute number: it is compared only to a *cell-permutation control*—re-pairing velocities with locations, which leaves the expression landscape and its spatial correlation intact and breaks only the velocity–gradient pairing—and a window that fails to beat its shuffle *abstains*.

Proposition 1 (Higher R^2, lower risk; ours)**.** *Under local linearity and isotropic noise, R^2 converges, as G grows, to the cell's signal-to-noise ratio, in which the M-weighted relative risk of $\hat{w}$ is strictly decreasing. Hence—absent shared components outside* $\mathrm{range}(J^\top)$ *correlated with the score (*no common-mode dominance*)—ranking by $\hat{R}^2$ asymptotically ranks by risk, and any kept top fraction improves average risk (rates and proof in Appendix A.4).*

3.3 Certificate 2 — how much to trust (τ/σ_k)

Statistic. How wrong can $\hat{w}$ be? Two quantities decide, and both are already in hand. The residual prices the *noise*: $\hat{\tau}^2 := \|\hat{r}\|^2/G$. The gradient geometry prices the *visibility*: the eigendecomposition $M = U\Sigma^2 U^\top$ yields directions u_1, u_2 and scales $\sigma_1 \geq \sigma_2$, where σ_k measures how strongly the expression landscape varies along u_k—how much evidence a unit of motion in that direction leaves in the data.

The error law (classical). Under local linearity and isotropic noise, the standard deviation of the direction-wise error is the noise divided by the visibility:

$$\mathrm{sd}\big[u_k^\top(\hat{w} - \bar{w})\big] \;=\; \frac{\tau}{\sigma_k}, \qquad k = 1, 2, \tag{3}$$

giving the per-cell, per-direction 90% error bar $1.645\,\hat{\tau}/\hat{\sigma}_k$: a direction one can barely see is a direction one can barely measure, and the law says exactly how barely (Gauss–Markov in our observation model, Simoncelli et al., 1991; SVD derivation in Appendix A.3).

Proposition 2 (Conservative validity of the bands; ours, informal)**.** *We do* not *assume $r = 0$: contamination pushes $\hat{\tau}^2$* upward only*, so the asymptotic coverage of the plug-in bands is* at least *nominal, at the price of a quantified width inflation. However much biological signal is mixed in, the error bars keep their promise as long as that signal is incoherent with the gradients. Formal statement and proof in Appendix A.5.*

Scope. The law certifies the noise-propagation (stochastic) component only, and a minimax lower bound (Appendix A.8) shows no estimator can improve it beyond a constant factor; if the residual departs from isotropy (structured residuals), the bands can turn optimistic (§4.4).

3.4 Certificate 3 — what is being measured (wave bound; outside the chain)

Recall Remark 1: the chain measures $\bar{w} = w - c$. The license for a results table to report w is issued here. The general form is a **negative-control test with injection-calibrated power**: given a coordinate subset believed to carry no confound, it either issues an upper bound on the confound or

degrades to an uninformative abstention—*a mislabeled partition attenuates injected and real waves alike (Proposition 3), so lost power self-reports as uninformative rather than as a wrong bound*. The subset is instantiated from sealed, literature-derived gene annotations, fixed before any data were examined: the *niche-anchored* subset $\mathcal{A}$ (resident-structure genes, pinned to the tissue scaffold) and the *cell-intrinsic program* subset $\mathcal{I}$ (genes of the migrating cells, travelling with the program); lists, curation protocol, and sources in Appendix B.1. Only unambiguous genes are admitted—purity outranks coverage (Proposition 3).

Proposition 3 (Partition identification; ours, informal)**.** *Under local linearity, isotropic noise, and partition purity, the two subset inversions cancel the motion and expose the wave:* $\mathbb{E}[\hat{w}_{\mathcal{A}} - \hat{w}_{\mathcal{I}}] = c$*; under symmetric mislabeling* φ *the detection attenuates to* $(1 - 2\varphi)\, c$ *with direction preserved. Formal statement—the asymmetric subset equations and the precise cancellation conditions—and proof in Appendix A.7.*

Statistic, null, and verdict. The per-cell detections $\hat{c}_i := \hat{w}_{\mathcal{A},i} - \hat{w}_{\mathcal{I},i}$ are pooled into the coherence statistic Λ—the norm of their coordinate-wise median relative to the median field norm—so that only a component the tissue shares survives; its zero point is a *shuffled-partition null*, and an observed Λ that fails to exceed the null's upper quantile is declared *silent* (construction, p-values, and diagnostics in Appendix B.2). Silence alone proves nothing—the detector might be deaf—so sensitivity is measured on the same data by planting waves of known relative size κ on the $\mathcal{I}$ genes, and θ is the smallest κ that alarms in every tested direction (Appendix B.3). The verdict is three-way: (*issue*) silence with confirmed sensitivity yields, by contraposition, the bound $\|c\| \le \theta \cdot \mathrm{median}_i \|\hat{w}_i\|$; (*exclude*) an alarm withholds the motion reading and reports the detection as a finding; (*uninformative*) no response to injection means silence carries no information, and we abstain in both directions. Bounds are issued per tissue and never transferred (verdict table and issuance ledger in Appendix B.4).

This certificate sits *outside* the estimation chain: the partition never enters $\hat{w}$ or $w^\star$ and is used only for the audit—domain knowledge is not used to produce the answer but to attempt its refutation (demonstrated in §4.3).

3.5 Assembly — from certificates to the cell-velocity estimator

Spending certificate 2 (direction-wise retreat). A poorly visible direction need not be discarded—certificate 2's number can be spent: only directions with wide error bars retreat toward the backbone prediction w_{g}—an arbitrary reference method's output, direction untouched, magnitude matched by one global scalar; γ is the prior width. The MAP solution is

$$w_B \;=\; (M + \lambda I)^{-1}\,(Jv + \lambda\, w_{\mathrm{g}}), \qquad \lambda \;=\; \frac{\hat{\tau}^2}{\gamma^2}. \tag{4}$$

In the eigenbasis of M, the per-direction data weight is $\sigma_k^2/(\sigma_k^2+\lambda) = \gamma^2/(\gamma^2+(\hat{\tau}/\hat{\sigma}_k)^2)$: shrinkage *is* the act of spending certificate 2's error bar against γ—reporting (3) and acting are the same quantity read twice (Appendix A.6).

Spending certificate 1 (per-cell gate). Use the inverted estimate only to the extent it explains the data, and return the rest to the backbone:

$$w^\star \;=\; g\, w_B + (1 - g)\, w_{\mathrm{g}}, \qquad g = R^2.$$

As $R^2 \to 0$, $w^\star \to w_{\mathrm{g}}$—the continuous semantics of abstention: *we measured nothing, so you get the backbone back, labeled as such*. The discrete gate is Remark 2's permutation contrast: windows that fail it are excluded from certification; and only tissues holding certificate 3's bound report $w^\star$ as "w" in the performance tables (certificate 3 changes no numbers).

The error bars travel with the estimator. Certificate 2's bars are per-direction confidence intervals for the estimand $\bar{w}$, centered at the measurement $\hat{w}$; the reported $w^\star$ inherits valid intervals by enlarging each half-width by the *observable* offset $|u_k^\top (w^\star - \hat{w})|$ (Appendix A.6)—sharp where the gate keeps the data, honestly wider where the prior fills in.

Global constants and provenance. Two global scale constants close the system—a unit match c_s for the backbone and the prior width $\gamma = c_\gamma \gamma_0$, defined with the pseudocode in Appendix C.4; the single constant c_γ was fixed a priori, not tuned (sensitivity in Appendix C.10). Every output ships with its provenance trace (g, g_k)—which cells, and which directions, are data-supported. SELCELL is deterministic and seed-free; the backbone is exchangeable; the prior enters the answer (fusion) but never the grading (certificates).

4 EXPERIMENTS

4.1 SETUP

Data. We experiment on TopoVelo's two synthetic benchmarks (ground-truth fields available) and five real tissues from the TopoVelo study, the thymus contributing three serial sections (S1–S3); full data description in Appendix C.1. Uncertified tissues are excluded from performance tables by §3.4's reporting rule and listed with their verdicts—never silently dropped.

Baselines and inheritance. Five baseline families: scVelo, VeloVAE (and its rate-prior variant), TopoVelo (GCN and GAT), and STT.[1] All share the cosine-vote projection step of §3.1 and inherit TopoVelo's released preprocessing unchanged (same upstream v, same spatial graphs), so the comparison isolates precisely the projection-versus-inversion choice. Before comparing, we reproduce the authors' committed outputs to within ± 0.01 CBDir.

Metrics. CBDir scores the cosine alignment of a cell's arrow with the annotated source-to-target direction; K-CBDir(k) its agreement with the local differentiation order among k spatial neighbors; ICCoh is reference-free neighbor agreement; VelAcc (per-cell cosine to the true vector) exists on the synthetics only. Arrows in tables mark the preferred direction; full definitions and the protocol (TopoVelo's official implementation, unchanged) are in Appendix C.6.

4.2 BENCHMARK PERFORMANCE

The certificates route each cell, and each direction within a cell, to whichever source is quantitatively trustworthy there; winning the benchmark is what that routing looks like from the outside. Ground-truth worlds first; then the same machinery, unchanged, on tissue. On the multiradial synthetic the margin over GAT is positive for every seed ($+0.072$ to $+0.109$; Tab. 1); on the bilinear synthetic the leaders tie.[2] On real tissue the routing pays off: the primary thymus section reaches CBDir 0.232 $[0.215, 0.250]$ against TopoVelo-GAT's 0.183 $[0.163, 0.205]$—disjoint 95% intervals (Tab. 2); a constant-field control scores -0.001. Appendix Fig. 7 contrasts the two leading fields on an H&E-registered window chosen by a declared rule (Appendix C.6; all six fields in Fig. 8); the exhibit zooms where the methods differ, not where both succeed—and the window is not the field's luck: across all 119 sliding windows of the section, SELCELL's field aligns better than the baseline's in 76.5%.

4.3 CERTIFICATE VERIFICATION

The numbers above stand on the certificates, so this subsection is the paper's center of gravity. Where ground truth exists—the synthetic benchmarks—we check that each certificate delivers what it states; on real tissue, where none can exist, we manufacture it by planting expression waves of known size.

Certificate 1, discrete verdict (Remark 2). Remark 2 promised that R^2 is read only against its shuffle (whose level reaches 0.61 on real tissue). Measured: in five tissues with no detectable advective signal the excess of R^2 over its cell-permutation control is at most 0.068 (per-tissue values in Appendix C.1), and the method abstains tissue-wide. Abstention has a mundane face (lung:

[1] STT requires its authors' legacy stack and runs on their preprocessing grid ($2{,}165 \times 100$); all other rows share our rebuilt grid ($2{,}843 \times 1{,}000$).

[2] Seed-level margins vs GAT: $+0.017/+0.004/-0.015/-0.007/+0.030$—mixed signs, so we report a tie; that landscape is well-conditioned everywhere, leaving the routing nothing to route.

Table 1: Synthetic benchmarks (ground truth available). Metrics match Tab. 2, plus the ground-truth-only VelAcc. A constant-field control stays at chance on both (CBDir -0.020 / 0.005). Full metric grid: Appendix Tab. 5.

	bilinear				*multiradial*			
Method	VelAcc↑	CBDir↑	ICCoh↑	K-CBDir(5)↑	VelAcc↑	CBDir↑	ICCoh↑	K-CBDir(5)↑
scVelo	0.259	0.348	0.253	0.662	0.511	0.444	0.522	0.860
VeloVAE	0.428	0.430	0.256	0.743	0.015	0.001	0.127	0.726
VeloVAE-RatePrior	0.506	0.534	0.414	0.797	0.300	0.209	0.211	0.808
TopoVelo (GCN)	**0.564**	0.609	**0.552**	0.853	0.599	**0.533**	0.605	0.898
TopoVelo (GAT)	0.526	0.610	0.543	0.848	0.522	0.445	0.527	0.881
SELCELL (ours)	0.540	**0.627**	0.547	**0.854**	**0.621**	0.528	**0.656**	**0.908**

Table 2: Primary thymus section S1, reported as cell velocity under the certificate-3 license (§4.3); bootstrap CIs in the text, per-method selection gains in Fig. 3a. STT runs on its authors' grid (§4.1). Latent-time metrics: appendix grid (SELCELL's on $\hat{v} = J^\top w^\star$; parity with GAT, no superiority claimed).

Method	CBDir↑	ICCoh↑	K-CBDir(5)↑
scVelo	0.107	0.326	0.462
VeloVAE	-0.172	0.407	0.333
STT	0.034	0.330	0.481
TopoVelo (GCN)	-0.230	0.345	0.349
TopoVelo (GAT)	0.183	0.560	0.503
SELCELL (ours)	**0.232**	**0.692**	**0.530**

-0.0005, plainly nothing there) and a deceptive one: on the E9 embryo every baseline draws coherent arrows, yet $R^2 = 0.231$ against a shuffle of 0.228 on all 16 pucks (Appendix Fig. 6)—as *no absolute thresholds* warned, 0.231 means nothing until its shuffle is consulted.

Certificate 1, gate and selection (Proposition 1). Proposition 1 promised that ranking by $\hat{R}^2$ ranks by risk. Measured on the synthetics first: selecting the most-trusted 60% of cells improves the benchmark by $+0.046$ to $+0.179$ (curves in Appendix C.7). On thymus the same ranking, at the preregistered keep-40% point, gains $+0.099$ (full chain; $+0.115$ for the inversion alone) against a random-ranking null of ± 0.020—while the baselines' confidence scores (their own where one is defined, the standard consistency score otherwise; Appendix C.6) are non-selective or anti-selective (GAT -0.095, scVelo -0.012, VeloVAE -0.013, STT -0.034; Fig. 3a). A preregistered transfer test on radar-measured currents is in Appendix C.5.

Certificate 2, error law and conservative validity (Eq. (3), Proposition 2). The error law is an equality: predicted and realized errors must match with unit slope. Measured by gene half-splits: slopes $0.90/1.01$ (bilinear), $1.02/1.14$ (multiradial), $1.03/0.99$ (thymus) across the two principal axes (Fig. 3b). Proposition 2 promised coverage at least nominal under contamination. Measured: 90% bands cover $0.864/0.878/0.899$—the synthetic shortfall is finite-G plug-in slack—with no collapse in reaction-active regions, which inflate $\hat{\tau}$ and populate the widest-band deciles of Fig. 3c. A constant split-conformal baseline matches marginal coverage but collapses conditionally where our bands hold, at greater width (Fig. 3c and caption); an *adaptive* conformal baseline is conditionally valid only when our $\hat{\tau}/\hat{\sigma}_k$ supplies its difficulty score (Appendix C.9); band width tracks realized error (Spearman 0.47–0.62).

Certificate 3, the three-way verdict in action (§3.4). On the primary section the detector is silent—$\Lambda = 0.12$ against a shuffled-partition 95% quantile of 0.17 ($p = 0.22$)—while injection confirms sensitivity: planted $\kappa = 0.5$ waves alarm in every direction ($\cos = 0.994$). By contraposition the bound $\|c\| \le 0.5 \cdot \mathrm{med}\|\hat{w}\|$ is issued (Appendix Fig. 5). The full ledger (Appendix Tab. 3) exercises all three verdicts on serial cuts from one donor: S3 receives only a *conservative* bound of 1.0 (borderline at $\kappa = 0.5$); S2 is *uninformative*—certificate 1 finds signal, but only four $\mathcal{A}$ genes survive its panel. The verdicts track the data, not the tissue. Separately, the cortex partition, declared weak at sealing, fails its gates and issues nothing—a wrong prior degrades to abstention.

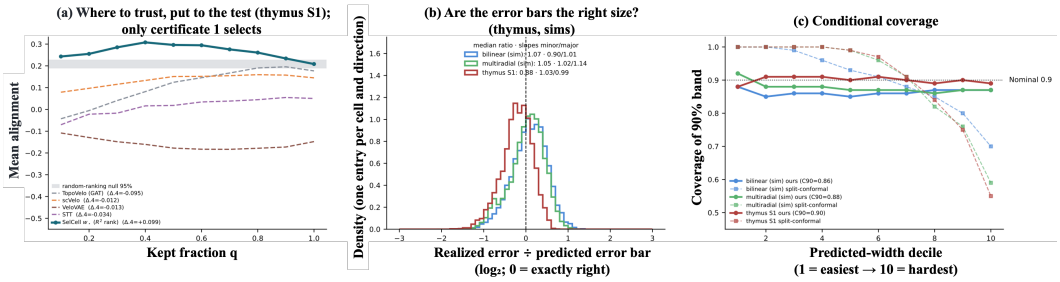


Figure 3: **Certificates 1 and 2 deliver what they state.** (a) Selection on thymus S1: only the R^2 ranking raises the alignment of the kept cells ($\Delta_{.4} = +0.099$, beyond the random-ranking null, grey); baseline scores are flat or anti-selective (GCN: none exposed). (b) The error law is an equality, and it holds: realized over predicted error is centred on 1 in every dataset (medians 0.88–1.07). (c) Our 90% bands hold near 0.9 in every decile; constant split-conformal collapses to 0.55–0.70 on the hardest—at matched marginal coverage, 0.67–0.82× the width (§4.3). Protocols: Appendix C.6.

Certificate ablation. Switching each consumption off isolates its contribution (Appendix Tab. 5): on thymus the raw inversion scores 0.214, the gate of certificate 1 lifts it to 0.232, and certificate 2 alone leaves it nearly unchanged (0.215)—as designed: its contribution is the calibrated error bars (Fig. 3b–c), not a table row. The multiradial column shows the honest flip side: with a weak backbone (GAT 0.445) the gate's fallback costs accuracy (0.611 → 0.528)—abstention's price is the backbone's own quality. Certificate 3 changes no numbers; removing it reports unlicensed arrows everywhere (Appendix Fig. 6).

4.4 Failure modes and the cost of abstention

Two assumption violations occur in the wild; we exhibit both. On cortex, a constant field ties the benchmark (CBDir 0.364): *common-mode dominance*, the regime excluded by Proposition 1. On section S3 the highest-R^2 band inverts (-0.05 to -0.07): *structured residuals*, the optimistic-band regime flagged in §3.3 and the diagnosis behind the benchmark reversal (Appendix Tabs. 4, 3)—*quality gates precede certificate gates*. Beyond these, the five no-signal tissues abstain wholesale (§4.3) while the baselines complete all seven; abstention costs no ranking—where trust is withheld, $w^\star$ returns the backbone (cortex: 0.351 vs GAT 0.350).

5 Conclusion

Reporting cell velocity from a measurement that destroys its subject requires moving the burden from the estimator to the observation model: written down, the usual assumption becomes a testable special case and three per-cell certificates follow analytically; on tissue the machinery that outperforms the strongest baseline abstains wholesale where the motion signal fails its shuffle—instruments, not ornaments—and the recipe applies wherever a quantity must be recovered from observations that can never be repeated. We believe no cell-velocity arrow should ship without a certificate—and the certificates can be free.

Limitations. (i) *Resolution.* The certified results are at Visium resolution, where the estimand is the average motion within a spot; the chain runs unchanged on near-single-cell platforms (Slide-seq, where the certificates abstained); a high-resolution tissue passing the gates would certify at single-cell resolution. (ii) *Ground truth.* Real cell motion is never observed on tissue; we mitigate it by scoring against the annotation TopoVelo itself adopts and by verifying the certificates on synthetics and injections.

REPRODUCIBILITY STATEMENT

Every number in this paper maps to a released script. The supplementary material contains (i) a documented reference implementation of the SELCELL chain and the exact scripts that produced every quantitative result, with a README mapping each number to its command; (ii) the committed small outputs (cached fields, curves, verdicts), so every reported number can be verified without recomputation; (iii) the raw-read rebuild pipeline for the thymus sections (ENA FASTQ $\rightarrow$ STARsolo $\rightarrow$ AnnData $\rightarrow$ backbone training); and (iv) a pinned environment specification. SELCELL itself is closed-form, deterministic, and seed-free, and its single a-priori constant was fixed before evaluation (§3.5); re-running the chain from raw inputs reproduces the committed fields bit-exactly. Evaluation protocols were preregistered and hash-sealed before execution: the sealed documents are released verbatim, and each result file embeds the seal identifier of the protocol it was produced under, so the chain of custody is checkable offline by comparing SHA-256 prefixes. All datasets are public, with accessions in Appendix C.1; baselines use their authors' released implementations, and the primary baseline's committed outputs are reproduced to within ± 0.01 CBDir before any comparison (§4.1). Protocol details for every exhibit are in Appendix C.6.

AI USE STATEMENT

Large language models (LLMs) were used extensively throughout this work, under author direction and with author verification at every step. Specifically, LLMs assisted in (i) developing and stress-testing the conceptual framing and mathematical claims of §3, including drafting derivations that were subsequently checked by the authors; (ii) implementing the analysis, evaluation, and figure code released in the supplementary material; (iii) executing the preregistered evaluation protocols, whose designs were fixed and hash-sealed before execution; (iv) literature search, with every citation verified against the published record; and (v) drafting and revising the manuscript text. LLMs were not used to generate or select experimental results: all quantitative claims derive from committed outputs of the released code, evaluation protocols were sealed before any result was seen, and negative and failed preregistered outcomes are reported as such. All scientific decisions—the problem framing, the estimand, what is claimed, and what is withheld—were made by the authors, who take full responsibility for the entire content of this paper.

REFERENCES

T. Abdelaal et al. SIRV: spatial inference of RNA velocity at the single-cell resolution. *NAR Genomics and Bioinformatics*, 6(3):lqae100, 2024.

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: a gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

O. Mac Aodha, A. Humayun, M. Pollefeys, and G. J. Brostow. Learning a confidence measure for optical flow. *IEEE TPAMI, 35(5):1107–1120*, 2013.

Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12): 1408–1414, 2020. doi: 10.1038/s41587-020-0591-3.

S. Birk et al. Multi-scale modeling of human tissues from spatial transcriptomics with terra. *bioRxiv 2026.07.29.741565*, 2026.

W. Chen, R. P. Mied, and C. Y. Shen. Near-surface ocean velocity from infrared images: Global Optimal Solution to an inverse model. *J. Geophys. Res. Oceans, 113:C10003*, 2008.

U. Chitra et al. Mapping the topography of spatial gene expression with interpretable deep learning. *Nature Methods, 22:298–309*, 2025.

L. Chizat, S. Zhang, M. Heitz, and G. Schiebinger. Trajectory inference via mean-field langevin in path space. In *NeurIPS*, 2022.

H. Cui et al. scGPT: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods, 21:1470–1480*, 2024.

W. J. A. Dahm, L. K. Su, and K. B. Southerland. A scalar imaging velocimetry technique for fully resolved four-dimensional vector velocity field measurements in turbulent flows. *Physics of Fluids A, 4(10):2191–2206*, 1992.

W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *ISPRS Intercommission Conf.*, 1987.

W. A. Fuller. Measurement Error Models. Wiley, 1987.

A. Gayoso et al. Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. *Nature Methods, 21:50–59*, 2024.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems 30*, pp. 4878–4887, 2017.

G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal., 17(6):883–893*, 1980.

Yichen Gu, Jialin Liu, Kun H. Lee, Chen Li, Lu Lu, Jaimee Moline, Renxiang Guan, and Joshua D. Welch. Topological velocity inference from spatial transcriptomic data. *Nature Biotechnology*, 44:771–782, 2026. doi: 10.1038/s41587-025-02688-8.

T. Hashimoto, D. Gifford, and T. Jaakkola. Learning population-level diffusions with generative RNNs. In *ICML*, 2016.

Y. He et al. KSRV: a kernel PCA-based framework for inferring spatial RNA velocity at single-cell resolution. *Frontiers in Genetics, 16:1695803*, 2025.

D. Heitz, É. Mémin, and C. Schnörr. Variational fluid flow measurements from image sequences: synopsis and perspectives. *Experiments in Fluids, 48(3):369–393*, 2010.

B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence, 17:185–203*, 1981.

E. Ilg et al. Uncertainty estimates and multi-hypotheses networks for optical flow. In *ECCV*, 2018.

J. K. Kearney, W. B. Thompson, and D. L. Boley. Optical flow estimation: an error analysis of gradient-based methods with local optimization. *IEEE TPAMI, 9(2):229–244*, 1987.

K. A. Kelly. An inverse model for near-surface velocity from infrared images. *J. Phys. Oceanogr., 19(12):1845–1864*, 1989.

D. Klein et al. Mapping cells through time and space with moscot. *Nature, 638:1065–1075*, 2025.

Ludger Klein, Bruno Kyewski, Paul M. Allen, and Kristin A. Hogquist. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nature Reviews Immunology*, 14(6):377–391, 2014. doi: 10.1038/nri3667.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018. doi: 10.1038/s41586-018-0414-6.

H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger. Toward a mathematical theory of trajectory inference. *Ann. Appl. Probab., 34(1A):428–500*, 2024.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116.

Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388, 2010. doi: 10.1097/EDE.0b013e3181d61eeb.

T. Liu and L. Shen. Fluid flow and optical flow. *J. Fluid Mech., 614:253–291*, 2008.

Zhiyuan Liu, Yaru Li, Dafei Wu, Weiwei Zhai, and Liang Ma. STEER: decoupling kinetics with Spatial-Temporal Explainable Expert model for RNA velocity inference. *National Science Review*, 13(9):nwag199, 2026. doi: 10.1093/nsr/nwag199.

W. Long, T. Liu, L. Xue, and H. Zhao. spVelo: RNA velocity inference for multi-batch spatial transcriptomics data. *Genome Biology, 26:239*, 2025.

K. Neklyudov, R. Brekelmans, D. Severo, and A. Makhzani. Action matching: Learning stochastic dynamics from samples. In *ICML*, 2023.

Marjorie A. Oettinger, David G. Schatz, Carolyn Gorka, and David Baltimore. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, 248(4962):1517–1523, 1990. doi: 10.1126/science.2360047.

Isabel Palmeirim, Domingos Henrique, David Ish-Horowicz, and Olivier Pourquié. Avian *hairy* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*, 91(5):639–648, 1997. doi: 10.1016/S0092-8674(00)80451-1.

Jong-Eun Park, Rachel A. Botting, Cecilia Domínguez Conde, Dorin-Mirel Popescu, Marieke Lavaert, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science*, 367(6480):eaay3224, 2020. doi: 10.1126/science.aay3224.

Howard T. Petrie and Juan Carlos Zúñiga-Pflücker. Zoned out: functional mapping of stromal signaling microenvironments in the thymus. *Annual Review of Immunology*, 25:649–679, 2007. doi: 10.1146/annurev.immunol.23.021704.115715.

Vishvak Raghavan, Brent Yoon, Gregory J. Fonseca, Yue Li, and Jun Ding. Dissecting and steering cell dynamics using spatially-informed RNA velocity with veloAgent. *Molecular Systems Biology*, 2026. doi: 10.1038/s44320-026-00213-w.

K. Sakalyan, A. Palma, F. Guerranti, F. J. Theis, and S. Günnemann. Modeling microenvironment trajectories on spatial transcriptomics with NicheFlow. In *NeurIPS*, 2025.

David G. Schatz, Marjorie A. Oettinger, and David Baltimore. The V(D)J recombination activating gene, RAG-1. *Cell*, 59(6):1035–1048, 1989. doi: 10.1016/0092-8674(89)90760-5.

J. Shi and C. Tomasi. Good features to track. In *CVPR*, 1994.

E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *CVPR*, 1991.

L. K. Su and W. J. A. Dahm. Scalar imaging velocimetry measurements of the velocity gradient tensor field in turbulent flows. i. assessment of errors. *Physics of Fluids, 8(7):1869–1882*, 1996.

E. D. Sun, R. Ma, P. Navarro Negredo, A. Brunet, and J. Zou. Tissue: uncertainty-calibrated prediction of single-cell spatial transcriptomics improves downstream analyses. *Nature Methods, 21:444–454*, 2024.

A. Tejada-Lapuerta, A. C. Schaar, et al. Nicheformer: a foundation model for single-cell and spatial omics. *Nature Methods*, 22(12):2525–2538, 2025. doi: 10.1038/s41592-025-02814-z.

C. V. Theodoris et al. Transfer learning enables predictions in network biology. *Nature, 618:616–624*, 2023.

Tomoo Ueno, Fumi Saito, Daniel H. D. Gray, Sachiyo Kuse, Kunio Hieshima, Hideki Nakano, Terutaka Kakiuchi, Martin Lipp, Richard L. Boyd, and Yousuke Takahama. CCR7 signals are essential for cortex–medulla migration of developing thymocytes. *Journal of Experimental Medicine*, 200(4):493–505, 2004. doi: 10.1084/jem.20040643.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

Johan Wallin, Hermann Eibel, Annette Neubüser, Jörg Wilting, Haruhiko Koseki, and Rudi Balling. Pax1 is expressed during development of the thymus epithelium and is required for normal T-cell maturation. *Development*, 122(1):23–30, 1996.

A. S. Wannenwetsch, M. Keuper, and S. Roth. Probflow: Joint optical flow and uncertainty estimation. In *ICCV*, 2017.

C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, and A. M. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *PNAS, 115(10):E2467–E2476*, 2018.

Michael L. Whitfield, Gavin Sherlock, Alok J. Saldanha, John I. Murray, Catherine A. Ball, Karen E. Alexander, John C. Matese, Charles M. Perou, Myra M. Hurt, Patrick O. Brown, and David Botstein. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular Biology of the Cell*, 13(6):1977–2000, 2002. doi: 10.1091/mbc.02-02-0030.

Nadav Yayon, Veronika R. Kedlian, Lena Boehme, Chenqu Suo, Brianna T. Wachter, et al. A spatial human thymus cell atlas mapped to a continuous tissue axis. *Nature*, 635(8039):708–718, 2024. doi: 10.1038/s41586-024-07944-6.

P. Zhou, F. Bocci, T. Li, and Q. Nie. Spatial transition tensor of single cells. *Nature Methods*, 21: 1053–1062, 2024. doi: 10.1038/s41592-024-02266-x.

Appendix

Contents

A Theory: full statements and proofs **15**

A.1 The orthogonal decomposition and the allocation identity 15

A.2 The model-null law of R^2 15

A.3 The direction-wise error law: SVD derivation 15

A.4 Proof of Proposition 1 (selection consistency) 15

A.5 Proof of Proposition 2 (conservative validity) 16

A.6 The shrinkage identity and the fusion error 16

A.7 Proof of Proposition 3 (partition identification) 17

A.8 Minimax lower bound (order-optimality of τ/σ_k) 17

A.9 Errors-in-variables account 17

B Certificate-3 protocol: issuing the wave bound **18**

B.1 Gene partition: curation protocol and sealed lists 18

B.2 Statistics and null 18

B.3 Injection calibration 19

B.4 Verdict and issuance ledger 19

C Implementation **21**

C.1 Data description 21

C.2 Qualitative exhibits on the certified section 21

C.3 Local Jacobian estimation 21

C.4 Pseudocode 21

C.5 Transfer test on radar-measured surface currents 22

C.6 Protocols behind the certificate exhibits 23

C.7 Certificate-1 selection on the synthetic benchmarks 24

C.8 Orthogonal-design synthetic: the estimand algebra in isolation 24

C.9 Adaptive-conformal baselines for certificate 2 25

C.10 Robustness and sensitivity 25

D Extended related work **25**

A THEORY: FULL STATEMENTS AND PROOFS

A.1 THE ORTHOGONAL DECOMPOSITION AND THE ALLOCATION IDENTITY

Let $\Pi := J^\top (JJ^\top)^{-1} J$ denote the orthogonal projection onto $\mathrm{range}(J^\top)$. For the intrinsic dynamics $\partial_t s$ define

$$c := -M^{-1} J\, \partial_t s, \qquad r := (I - \Pi)\, \partial_t s,$$

so that $\partial_t s = -J^\top c + r$ and $Jr = 0$ *by construction* (since $J(I - \Pi) = 0$). The decomposition is unique because $\mathbb{R}^G = \mathrm{range}(J^\top) \oplus \ker(J)$.

Lemma A.1 (Allocation identity)**.** *Under the observation model* (1)*, the inversion* $\hat{w} = M^{-1} J v$ *satisfies*

$$\hat{w} \;=\; (w - c) \;+\; M^{-1} J \varepsilon \qquad \textit{exactly, with no further assumption.}$$

Proof. $M^{-1} J \big(J^\top w - J^\top c + r + \varepsilon \big) = w - c + M^{-1} J r + M^{-1} J \varepsilon$, and $Jr = 0$. □

The reaction component is cancelled without any approximation—this is an identity of the decomposition, not a "no reaction" assumption. The same computation gives the residual $\hat{r} = v - J^\top \hat{w} = r + (I - \Pi)\varepsilon$, used in Appendix A.5.

Proof of Remark 1 (indistinguishability). The map $(w, c) \mapsto (w + \delta, c + \delta)$ leaves $J^\top w - J^\top c$ unchanged, hence leaves the distribution of v unchanged, hence leaves the distribution of every function of v (residuals and goodness-of-fit included) unchanged. Only $\bar{w} = w - c$ is identifiable. □

A.2 THE MODEL-NULL LAW OF R^2

Under the model null $v = \varepsilon \sim \mathcal{N}(0, \tau^2 I_G)$ we have $R^2 = \|\Pi\varepsilon\|^2 / \|\varepsilon\|^2$. By orthogonal projection, $\|\Pi\varepsilon\|^2 \sim \tau^2 \chi^2_2$ and $\|(I - \Pi)\varepsilon\|^2 \sim \tau^2 \chi^2_{G-2}$ are independent, so by the chi-square ratio identity

$$R^2 \;=\; \frac{\chi^2_2}{\chi^2_2 + \chi^2_{G-2}} \;\sim\; \mathrm{Beta}\big(1, \tfrac{G-2}{2}\big)\,, \qquad \mathbb{E}[R^2] = \tfrac{2}{G}. \qquad \square$$

(the classical null law of the coefficient of determination, derived above).

A.3 THE DIRECTION-WISE ERROR LAW: SVD DERIVATION

Take the thin SVD $J = U \Sigma Q^\top$ with $U = [u_1\, u_2]$, $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2)$, and $Q = [q_1\, q_2] \in \mathbb{R}^{G \times 2}$, $Q^\top Q = I_2$; then $M = U \Sigma^2 U^\top$. By Lemma A.1 and $J\eta = J\varepsilon$ (since $Jr = 0$),

$$u_k^\top (\hat{w} - \bar{w}) = u_k^\top M^{-1} J \varepsilon = \sigma_k^{-2}\, u_k^\top U \Sigma Q^\top \varepsilon = \frac{\sigma_k}{\sigma_k^2}\, q_k^\top \varepsilon = \frac{q_k^\top \varepsilon}{\sigma_k}.$$

Since $\mathbb{E}[q_k^\top \varepsilon] = 0$ and $\mathrm{Var}[q_k^\top \varepsilon] = q_k^\top (\tau^2 I) q_k = \tau^2 \|q_k\|^2 = \tau^2$ (orthonormality of the SVD), $\mathrm{sd}[u_k^\top (\hat{w} - \bar{w})] = \tau / \sigma_k$. □This is the eigenbasis restatement of the Gauss–Markov covariance $\mathrm{Cov}(\hat{w}) = \tau^2 M^{-1}$ of weighted least squares.

A.4 PROOF OF PROPOSITION 1 (SELECTION CONSISTENCY)

Write $S := \|J^\top \bar{w}\|^2$ and define the per-cell SNR $\rho := S/(S + G\tau^2)$—the quantities the main-text statement refers to; the exact rates below are $R^2 \to_p \rho$ and relative risk $(2/G)(1 - \rho)/\rho$.

(ii) Relative-risk identity. Using Lemma A.1,

$$\mathbb{E}\|\hat{w} - \bar{w}\|^2_M = \mathbb{E}\big[\varepsilon^\top J^\top M^{-1} J \varepsilon\big] = \tau^2\, \mathrm{tr}(\Pi) = 2\tau^2,$$

while $\|\bar{w}\|^2_M = S$. Hence the relative risk equals $2\tau^2/S$, and substituting $(1 - \rho)/\rho = G\tau^2/S$ gives $(2/G)(1 - \rho)/\rho$, strictly decreasing in ρ. □

(i) The target of R^2. The explained energy $\|J^\top \hat{w}\|^2 = \|\Pi v\|^2 \sim \tau^2 \chi^2_2(S/\tau^2)$ is noncentral chi-square; monotonicity of $\mathbb{E}[R^2]$ in the noncentrality S/τ^2 follows from the stochastic monotonicity

of noncentral chi-square ratios. For large G (effective G_{eff}), the law of large numbers along the gene axis concentrates numerator and denominator at $S+2\tau^2$ and $S+G\tau^2$ respectively, so $R^2 \to_p \rho$, with concentration rate of order $\sqrt{2/G}$ from the Beta tails. With $r \neq 0$ the denominator concentrates at $S + \|r\|^2 + G\tau^2$ instead, so $\hat{R}^2$ under-states ρ—the conservative direction; the ranking is preserved whenever $\|r\|^2$ does not vary against the score, which is the no-common-mode condition. □

(Selection conclusion.) Under the no-common-mode condition, the ranking by $\hat{R}^2$ asymptotically coincides with the ranking by ρ; by (ii) risk is monotone decreasing in ρ, so the average relative risk of the top-q selection is non-decreasing in q and lower than that of the full population. The empirical profile of the violated case is given in §4.4.

A.5 Proof of Proposition 2 (conservative validity)

Formal statement. For any r with $Jr = 0$, $\mathbb{E}[\hat{\tau}^2] = \frac{G-2}{G}\tau^2 + \|r\|^2/G$; for Gaussian ε the asymptotic coverage of the plug-in bands is at least nominal, with equality iff $r = 0$, and the width inflates by $\sqrt{1 + \|r\|^2/(G\tau^2)}$.

(i) By Appendix A.1, $\hat{r} = r + (I - \Pi)\varepsilon$ (note $\Pi r = 0$ because $Jr = 0$). Orthogonality kills the cross term in expectation and $\mathbb{E}\|(I - \Pi)\varepsilon\|^2 = (G - 2)\tau^2$, so

$$\mathbb{E}\|\hat{r}\|^2 = \|r\|^2 + (G-2)\tau^2 \quad \Longrightarrow \quad \mathbb{E}[\hat{\tau}^2] = \frac{G-2}{G}\,\tau^2 + \frac{\|r\|^2}{G}.$$

Dividing by G rather than $G - 2$ is deliberate (an honest plug-in with no additional downward bias); the contamination term moves the expectation *upward only*. □

(ii) By the law of large numbers along the gene axis, $\hat{\tau}^2 \to_p \tau^2 + \|r\|^2/G \geq \tau^2$. By monotonicity of Gaussian quantiles, the band $\hat{w} \pm z_\alpha \hat{\tau}/\hat{\sigma}_k$ then covers at asymptotic level at least $1 - \alpha$, with equality iff $r = 0$. □

(iii) The width ratio converges to $\sqrt{\mathbb{E}\hat{\tau}^2/\tau^2} \to \sqrt{1 + \|r\|^2/(G\tau^2)}$; the estimation error of $\hat{\sigma}_k$ vanishes in the window asymptotics by the continuous mapping theorem. □

A.6 The shrinkage identity and the fusion error

Diagonalizing $w_B = (M + \lambda I)^{-1}(Jv + \lambda w_{\text{g}})$ in the eigenbasis of M gives, per direction,

$$u_k^\top w_B = \frac{\sigma_k^2\, u_k^\top \hat{w} + \lambda\, u_k^\top w_{\text{g}}}{\sigma_k^2 + \lambda}, \qquad \text{data weight } a_k = \frac{\sigma_k^2}{\sigma_k^2 + \lambda}.$$

Substituting $\lambda = \hat{\tau}^2/\gamma^2$,

$$a_k = \frac{\sigma_k^2 \gamma^2}{\sigma_k^2 \gamma^2 + \hat{\tau}^2} = \frac{\gamma^2}{\gamma^2 + (\hat{\tau}/\hat{\sigma}_k)^2}. \qquad \square$$

Fusion error (bias–variance split). From the same diagonalization,

$$u_k^\top (w_B - w) = a_k\, u_k^\top (\hat{w} - w) + (1 - a_k)\, u_k^\top (w_{\text{g}} - w),$$

so, conditionally on w_{g}, the standard deviation of the stochastic term is $a_k\, \tau/\sigma_k \leq \tau/\sigma_k$ (variance never increases), at the price of a bias bounded by $(1 - a_k)\, |u_k^\top (w_{\text{g}} - w)|$—paid only where the data are weak, and proportional to how wrong the backbone is there.

A valid region for the reported estimator. The region of certificate 2 is a statement about the estimand: $|u_k^\top (\hat{w} - \bar{w})| \leq 1.645\, \hat{\tau}/\hat{\sigma}_k$ with probability 0.9 per axis. By the triangle inequality,

$$|u_k^\top (w^\star - \bar{w})| \;\leq\; |u_k^\top (w^\star - \hat{w})| \;+\; |u_k^\top (\hat{w} - \bar{w})|,$$

and the first term is computed from two known vectors. Enlarging each half-axis by this observable offset therefore yields a region for $\bar{w}$ centered at $w^\star$ with coverage at least nominal—no model of the backbone's error is used. Explicitly, $u_k^\top (w^\star - \hat{w}) = (1 - g\, a_k)\, u_k^\top (w_{\text{g}} - \hat{w})$, which vanishes as $g\, a_k \to 1$: the region is sharp exactly where gate and data weight keep the measurement, and honestly wider where the prior fills in.

A.7 Proof of Proposition 3 (partition identification)

Formal statement. The two subsets obey asymmetric equations: $v_{\mathcal{A}} = J_{\mathcal{A}}^{\top} w + \eta_{\mathcal{A}}$ (wave-free) and $v_{\mathcal{I}} = J_{\mathcal{I}}^{\top}(w - c) + \eta_{\mathcal{I}}$ (wave-loaded), with $\eta_\cdot := r_\cdot + \varepsilon_\cdot$. Under local linearity, isotropic noise, and partition purity, the subset inversions $\hat{w}_{\mathcal{A}} := M_{\mathcal{A}}^{-1} J_{\mathcal{A}} v_{\mathcal{A}}$ and $\hat{w}_{\mathcal{I}} := M_{\mathcal{I}}^{-1} J_{\mathcal{I}} v_{\mathcal{I}}$ satisfy $\mathbb{E}[\hat{w}_{\mathcal{A}} - \hat{w}_{\mathcal{I}}] = c$; under symmetric mislabeling φ this becomes $(1 - 2\varphi)\, c$, with power degenerating as $\varphi \to \frac{1}{2}$.

Proof. With $\eta_\cdot := r_\cdot + \varepsilon_\cdot$ and $\mathbb{E}[\eta_\cdot] = 0$,

$$\mathbb{E}[\hat{w}_{\mathcal{A}}] = \mathbb{E}\big[M_{\mathcal{A}}^{-1} J_{\mathcal{A}} (J_{\mathcal{A}}^{\top} w + \eta_{\mathcal{A}})\big] = w, \qquad \mathbb{E}[\hat{w}_{\mathcal{I}}] = w - c, \qquad \mathbb{E}[\hat{c}] = c.$$

Note the two design properties: the unknown motion w cancels in the difference (specificity), and the slope on c is one (the statistic is a meter, not merely a test). □

Precision note (subset cancellation of r). The full-J orthogonality $Jr = 0$ does *not* transfer to subsets ($J_{\mathcal{A}} r_{\mathcal{A}} \neq 0$ in general). The operative ground for the cancellation above is the subset form of the isotropic-noise assumption—the incoherent remainder $\eta_\cdot$ has mean zero—which is exactly what the assumption box of §3.1 declares. Cases where structured residuals break this are disclosed empirically in §4.4.

Attenuation. Under symmetric mislabeling at rate φ (a fraction φ of $\mathcal{A}$ is actually wave-loaded and a fraction φ of $\mathcal{I}$ is not), linearity of the mixtures gives $\mathbb{E}[\hat{w}_{\mathcal{A}}] = w - \varphi c$ and $\mathbb{E}[\hat{w}_{\mathcal{I}}] = w - (1 - \varphi)c$, hence $\mathbb{E}[\hat{c}] = (1 - 2\varphi)\, c$: the magnitude attenuates, the direction is preserved, and power degenerates completely at $\varphi = \frac{1}{2}$—a regime this design cannot resolve, stated explicitly. *Rank condition:* each subset must contain genes whose gradients span two dimensions (≥ 2 non-parallel), so that $M_\cdot$ is invertible; this is enforced by the curation protocol (Appendix B.1). □

A.8 Minimax lower bound (order-optimality of τ/σ_k)

Theorem A.2 (Minimax lower bound)**.** *Under Gaussian isotropic noise, for any estimator $\hat{w}$,*

$$\inf_{\hat{w}} \sup_{\bar{w}} \mathbb{E}\big|u_k^{\top}(\hat{w} - \bar{w})\big| \;\geq\; \frac{1}{4} \cdot \frac{\tau}{\sigma_k}.$$

Proof (two-point argument). Consider the two hypotheses $\bar{w}$ versus $\bar{w} + \delta u_k$ with $\delta = \tau/\sigma_k$. The induced distributions of v have $\mathrm{KL} = \delta^2 \sigma_k^2 / (2\tau^2) = \frac{1}{2}$; Le Cam's two-point bound then gives minimax error at least $(\delta/2)\,(1 - \sqrt{\mathrm{KL}/2}) \geq \delta/4$. □

Under quasi-stationarity violations of size δ_k the bound acquires an additive δ_k term (the price of misspecification). Consequence: the bands of (3) are order-optimal for the stochastic component they certify—no estimator improves them beyond a constant factor; the statement does not cover bias from smoothing, $\hat{J}$ estimation, or model misspecification.

A.9 Errors-in-variables account

In deployment, J is replaced by the estimate $\hat{J}$ (Appendix C.3), whose error splits into a subspace misalignment and a magnitude (scale) component. (i) The residual is taken against $\mathrm{range}(\hat{J}^{\top})$; sampling error of $\hat{J}$ misaligns this subspace with the true one, so part of the true signal leaks into the residual—deflating R^2 and inflating $\hat{\tau}^2$. The misalignment component therefore under-states validity. A pure scale error, by contrast, is invisible to this argument: $\hat{J} = \alpha J$ leaves the residual and R^2 unchanged while rescaling $\hat{w}$ by $1/\alpha$—magnitudes shift, directions (and every direction-based score) do not—so it is checked empirically, by the split-half slopes below, not claimed conservative. (ii) The visibilities $\hat{\sigma}_k$ inherit sampling error, which attenuates the data weight a_k (Appendix A.6) toward the backbone—again the cautious direction. Smoothing at bandwidth ℓ additionally low-passes the landscape (the source of the curl-shrinkage limitation discussed with the diagnostics). A quantitative attenuation account follows standard errors-in-variables arguments and is omitted here; the empirical calibration check of §4.3 (split-half slopes) verifies the net effect end to end.

B CERTIFICATE-3 PROTOCOL: ISSUING THE WAVE BOUND

B.1 GENE PARTITION: CURATION PROTOCOL AND SEALED LISTS

Protocol. (i) The partition is selected from literature knowledge *only* and sealed before any expression or velocity data are examined (anti-forking: lists, gates, and estimator are frozen at sealing; failures are reported as-is, and re-curation is admissible only as a new preregistration). (ii) Only unambiguous genes are admitted: mislabeling at rate φ attenuates sensitivity by $(1 - 2\varphi)$ (Appendix A.7), so *purity outranks coverage*—for the thymus, 69 of $\sim$1,000 genes are assigned. (iii) Each subset must span two gradient dimensions (rank condition of Appendix A.7).

Thymus $\mathcal{A}$ (niche-anchored, 30): TEC / mesenchyme / ECM / endothelium. CCL19, CCL25, PRSS16, CTSV, PAX1, TBATA, KRT16, KRT17, KRT19, DSP, PERP, SPINK5, EGFR, CLU, IGFBP7, FN1, COL23A1, COL27A1, AEBP1, PRRX1, ALDH1A2, TFPI, APOLD1, LIFR, GSN, EMID1, ADAMTS10, WNT5B, PTGDS, C1S.

Thymus $\mathcal{I}$ (cell-intrinsic program, 39): rearrangement / selection / maturation / migration + proliferation block. RAG1, RAG2, CD8A, CD1E, TRBC1, TRBC2, CD27, CD6, CCR7, TCF7, IL7R, SELL, LCP2, FYB1, ACAP1, RIPOR2, ITPKB, BACH2, SCML4, JAML, RASAL3, UBASH3B; MKI67, TOP2A, PCNA, STMN1, CENPE, PRC1, MELK, CHEK1, CCNE2, POLE2, PARPBP, CDCA2, RPA2, MSH2, KIF20B, ARHGAP11A, SPDL1.

Cortex $\mathcal{A}$ (12): VZ-anchored / ECM / non-neuronal. Declared weak *at sealing* (progenitor/ECM mixture; the result carries low standing by predeclaration): Sfrp1, Gli3, Ptprz1, Sox2ot, Pantr1, Igfbpl1, Ccnd2, Tnc, Hexb, Heg1, Zeb1, Wnt7b.

Cortex $\mathcal{I}$ (26): neuronal maturation cascade. Satb2, Sox5, Tle4, Foxp2, Mef2c, Cux2, Rbfox3, Rbfox1, Map2, Mapt, Map1b, Gap43, Stmn2, Syt1, Gria1, Grin2b, Nrxn1, Nrxn3, Nlgn1, Celf4, Myt1l, Basp1, Nrep, Cpe, Dab1, Unc5d.

Sources of the assignments. Tissue atlases: (Yayon et al., 2024; Park et al., 2020). Compartment biology of resident stroma versus migrating thymocytes: (Petrie & Zúñiga-Pflücker, 2007; Klein et al., 2014; Ueno et al., 2004) (the CCL19/CCR7 ligand–receptor separation is the archetype of an $\mathcal{A}/\mathcal{I}$ assignment). Proliferation signature: (Whitfield et al., 2002). Classical single-gene assignments: (Schatz et al., 1989; Oettinger et al., 1990; Wallin et al., 1996).

B.2 STATISTICS AND NULL

Per cell, $\hat{c}_i = \hat{w}_{\mathcal{A},i} - \hat{w}_{\mathcal{I},i}$ (valid cells: both subset inversions finite; subset inversions use the convention of (2) with a stabilizer $M_\cdot + 10^{-9}\,\mathrm{tr}(M_\cdot)\,I$).

- $\Lambda_1 = \|\mathrm{median}_i\,\hat{c}_i\|/\mathrm{median}_i\|\hat{w}_i\|$: global coherence magnitude (dimensionless). The median is coordinate-wise; it commutes exactly with constant shifts, so $\mathrm{median}_i(c + e_i) = c + \mathrm{median}_i(e_i)$ and cell-idiosyncratic components cancel.
- Λ_2 (split-half stability): $\mathcal{A}$ and $\mathcal{I}$ are each randomly halved; the direction agreement $\cos\big(\mathrm{median}\,\hat{c}^{(1)},\,\mathrm{median}\,\hat{c}^{(2)}\big)$ is recorded as the median over $B = 20$ resplits. Only genuinely shared signal reproduces across halves.
- *Shuffled-partition null*: random size-matched $(|\mathcal{A}|, |\mathcal{I}|)$ repartitions of $\mathcal{A} \cup \mathcal{I}$, $K = 200$ draws, give the null distributions of Λ_1, Λ_2; $p = (1 + \#\{\mathrm{null} \geq \mathrm{obs}\})/(K + 1)$. The shuffle destroys only the biological meaning of the labels; geometry, noise, and gene counts are preserved.
- *Reading gates* (fixed in advance): alarm $= \{p(\Lambda_1) \leq 0.05\} \wedge \{p(\Lambda_2) \leq 0.05\}$. The level 0.05 is a sealed convention controlling the false-alarm rate only; control of false *silence* is the role of Appendix B.3.
- *Diagnostics (not gates; honest-reporting duty)*: D1, decomposition of $\hat{c}$ into components parallel/perpendicular to the local $\hat{w}$ (the parallel part can be forged by subset-specific

attenuation; the perpendicular part is forgery-resistant); D2, within-class contrast (split-half differences $\mathcal{A} \leftrightarrow \mathcal{A}$ and $\mathcal{I} \leftrightarrow \mathcal{I}$ at the same B); D3, per-compartment median $\hat{c}$ (descriptive only).

B.3 Injection calibration

A wave of known size is planted in the same measured data, on the program subset only:

$$v'_a = v_a - \hat{J}_a^\top c_{\text{inj}}, \quad a \in \mathcal{I}, \qquad c_{\text{inj}} = \kappa \cdot \text{median}_i \|\hat{w}_i\| \cdot d,$$

with $\kappa \in \{0.25, 0.5, 1.0\}$ and d ranging over a battery of unit directions (the true wave direction is unknown, so all directions are tested). The entire pipeline of Appendix B.2 is re-run for each (κ, d); the computation is deterministic—the axis of repetition is the scenario (direction, resplit), not a random seed. The sensitivity threshold is

$$\theta \;:=\; \min\{\kappa \text{ in the grid : the alarm fires decisively for } \textit{every} \text{ tested direction } d\},$$

selected on the grid with no interpolation or extrapolation—to tighten the bound, a smaller κ must actually be injected and shown to fire. The direction recovery $\cos(\hat{c}, d)$ is reported alongside (the detector should act as a compass, not merely a bell). Calibration is performed *per tissue*, and θ is never transferred across tissues.

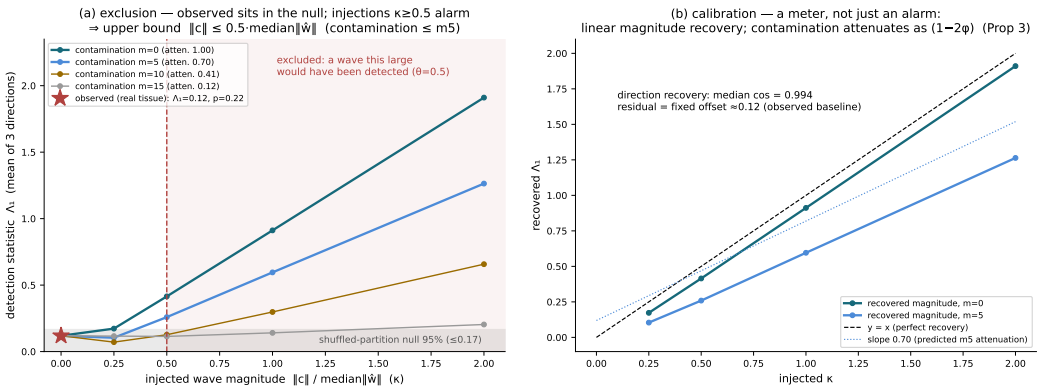


Figure 4: Injection calibration in full (companion to Fig. 5). (a) Mean coherence over the three planted directions against the planted size κ for partition contamination $m \in \{0, 5, 10, 15\}$ mislabelled genes; the attenuation follows $(1 - 2\varphi)$ (Proposition 3). Under $m = 10$ the first size that alarms in every direction is $\kappa = 2.0$; under $m = 15$ no tested size alarms, and the certificate would be uninformative. (b) Recovered magnitude against planted magnitude: linear, with the predicted slope under $m = 5$; median direction recovery $\cos(\hat{c}, d) = 0.994$, so the detector is a meter, not only an alarm.

B.4 Verdict and issuance ledger

Table 3: Certificate ledger for the three serial thymus sections (one donor; N = number of cells; accession mapping in Appendix C.1). Per-section benchmark scores are in the appendix replication table.

Section	N	Cert. 1	Cert. 3 verdict	Grounds
S1 (primary)	2,843	pass	**bound** ≤ 0.5	silent (p=0.22); injection fires (κ=0.5, all dir.)
S2	3,125	pass	**uninformative**	4 $\mathcal{A}$ genes survive intersection—no power
S3 (poor)	793	pass	**bound** ≤ 1.0 (cons.)	κ=0.5: 2/3 dir., p=.0498; next grid point quoted

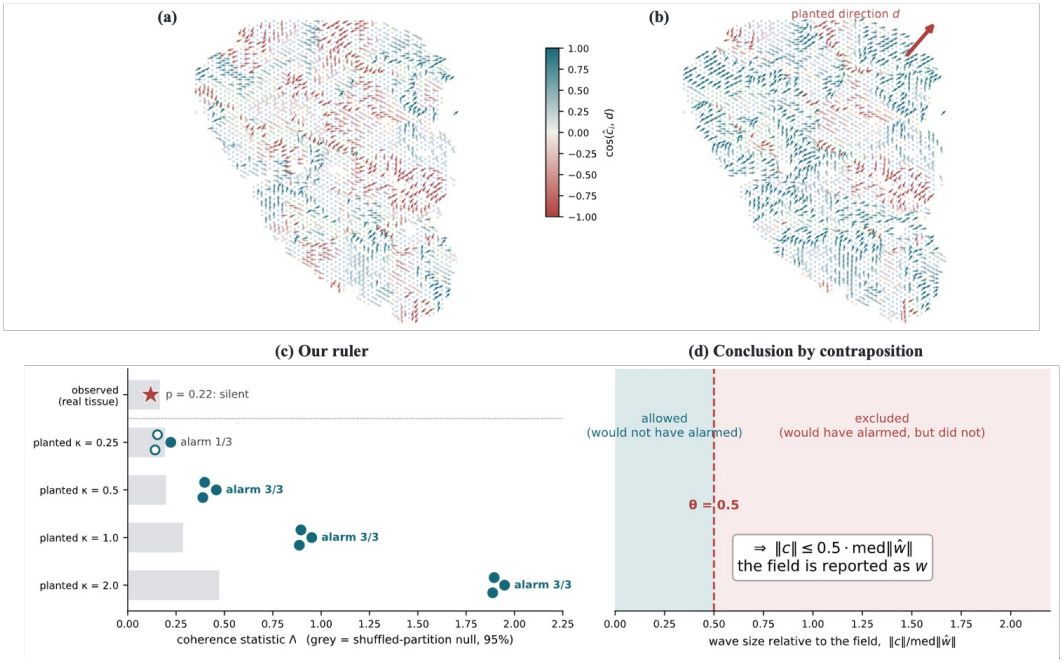


Figure 5: **Certificate 3: silence of a sensitivity-calibrated detector converts an untestable assumption into a measured upper bound, $\|c\| \le 0.5 \cdot \mathrm{med}\|\hat{w}\|$ on thymus S1.** The argument reads left to right. (a) What the detector sees: per cell, the inversion on niche-anchored genes minus the inversion on program genes, $\hat{c} = \hat{w}_{\mathcal{A}} - \hat{w}_{\mathcal{I}}$ (arrows coloured by their cosine with the direction d used in (b)). On real tissue the field is incoherent: its coherence $\Lambda = 0.12$ lies inside the shuffled-partition null ($p = 0.22$), so the detector is silent. (b) Silence alone proves nothing, so a wave of known relative size $\kappa = 0.5$ is planted on the program genes along d: the field becomes coherent and the detector fires ($\Lambda = 0.39$, $p = 0.01$). (c) Both facts on one ruler: grey bars are the null's 95% range, the star is the observed value, dots are the planted values in three directions (filled = alarm). From $\kappa = 0.5$ every direction alarms, so $\theta = 0.5$. (d) Contraposition: any real wave of relative size $\ge \theta$ would have alarmed; none did, so the wave is bounded by θ and the field is reported as w. Partition contamination $m = 0$ (the sealed protocol); attenuation under contamination and the magnitude calibration are in Appendix B.3 (Fig. 4).

By contraposition, for the tissue-coherent waves of the tested family—the component the median statistic detects (§3.4)—

$$\big(\|c\| > \theta \cdot \mathrm{median}_i\|\hat{w}_i\| \Rightarrow \text{alarm}\big) \ \wedge \ \neg\text{alarm} \implies \|c\| \le \theta \cdot \mathrm{median}_i\|\hat{w}_i\|.$$

Observation	Verdict	Report
silence ∧ injection detected	**issue bound**	"any tissue-coherent wave of the tested family is $\le \theta \cdot \mathrm{med}\|\hat{w}\|$"; tables report w
alarm (both gates)	**exclude**	arrows not read as motion; the detection itself ($\hat{c}$) reported as a finding
silence ∧ no injection response	**uninformative**	no statement in either direction; excluded from tables

The issuance ledger (tissue × {issued θ / excluded / uninformative}, with diagnostics D1–D3 attached) is Tab. 3; its verdicts are narrated in §4.3. Synthetic pre-validation ledger line: in a self-constructed random-feature world with a known ground-truth partition, the detector recovers the planted wave with small relative error—a sanity check only; the injection calibration on real tissue (Appendix B.3) is the higher-grade validation.

C IMPLEMENTATION

C.1 DATA DESCRIPTION

Synthetic benchmarks (TopoVelo simulation suite). *bilinear* and *multiradial*, generated by the released simulator with ground-truth velocity fields; used for all ground-truth-level verification (VelAcc, error-law slopes, selection).

Thymus (certified real tissue). Three serial sections from donor Z4 of Yayon et al. (2024), labelled by role in the main text; accession mapping: S1 (primary) = TA11556493 ($N = 2{,}843$ cells), S2 = TA11486162 ($N = 3{,}125$), S3 (poor) = TA11556494 ($N = 793$). All three are rebuilt from raw reads (ENA PRJEB77091) with one pipeline ($G \approx 1{,}000$ genes); S1 carries the primary benchmark, S2 and S3 the replication and the certificate ledger (Tabs. 4 and 3).

Table 4: Replication across the three serial thymus sections (companion to the ledger): CBDir of the strongest baseline and of SELCELL, their gap, the selection gain Δ_{40}, median R^2, and the constant-field control. The S3 reversal is diagnosed in §4.4 (structured residuals).

Section	GAT	SELCELL	Δ	Δ_{40}	median R^2	const. field
S1 (primary)	0.183	**0.232**	$+0.049$	$+0.099$	0.562	-0.001
S2	0.180	**0.200**	$+0.020$	$+0.034$	0.298	-0.005
S3 (poor)	**0.371**	0.345	-0.026	-0.025	0.340	-0.016

The thymus is among the tissues of the TopoVelo study; we rebuild it independently and extend it with the two replicate sections.

Further TopoVelo-study tissues (diagnostic). Cortex E15 ($N = 2{,}525$; rebuilt from raw data), embryoid body D21 ($N = 5{,}559$; authors' published output), Stereo-seq gut ($N = 2{,}381$) and lung ($N = 1{,}030$; authors' published outputs; TopoVelo deposits: outputs figshare 28516184, preprocessed inputs figshare 28516139). The Slide-tags E14 brain ($N = 4{,}623$; authors' output) is held out of the present analysis.

Abstention exhibits (our additions). E9 embryo, 16 Slide-seq pucks, rebuilt from raw reads—the counterexample tissue of Fig. 6; chicken heart D14 and an E9.5 embryo section, cited only for their abstention verdicts (R^2 excess over the cell-permutation control: chicken 0.068, E9.5 0.005, gut -0.001, lung -0.0005). Public identifiers of the abstention exhibits: E9 mouse embryo 3D Slide-seq, figshare 21695879 (16 pucks); E9.5 mouse embryo Slide-seq, SRA SRR18128768; chicken heart D14 Visium, SRA SRR11616466 (GEO series GSE149457).

C.2 QUALITATIVE EXHIBITS ON THE CERTIFIED SECTION

Fig. 7 contrasts the two leading fields on the declared window; Fig. 8 shows all six fields.

C.3 LOCAL JACOBIAN ESTIMATION

At each cell, $\hat{J}$ is obtained by kernel-weighted local-linear regression over spatial neighbors at bandwidth ℓ—in closed form $\hat{J}^\top = C_{xx}^{-1} C_{xs}$, the ratio of position–expression covariation to positional spread—and v is smoothed at the same scale, so that residuals reflect model misfit rather than resolution mismatch. Two costs are stated where they arise: smoothing at scale ℓ low-passes the landscape (curl-shrinkage; discussed with the diagnostics), and the sampling error of $\hat{J}$ propagates as quantified in Appendix A.9.

C.4 PSEUDOCODE

Algorithm 1 lists the full chain of §3.5. When the data are perfect ($\lambda \to 0$, $g \to 1$), $w^\star \to \hat{w}$; when the data are uninformative ($g \to 0$), $w^\star \to w_{\mathrm{g}}$—the chain only ever moves between these two limits.

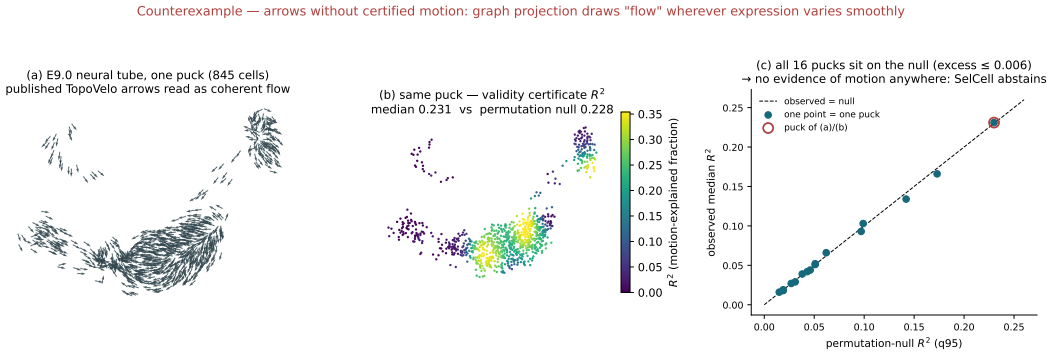


Figure 6: The deceptive face of abstention (E9 embryo, §4.3). (a) On one puck of the E9.0 neural tube (845 cells), the published TopoVelo arrows read as a smooth flow. (b) The same puck under certificate 1: median $R^2 = 0.231$ against a cell-permutation null of 0.228—the apparent flow explains no more of the RNA velocity than a shuffled control. (c) All 16 pucks sit on the null diagonal (excess ≤ 0.006), so SELCELL abstains tissue-wide while every baseline completes the field.

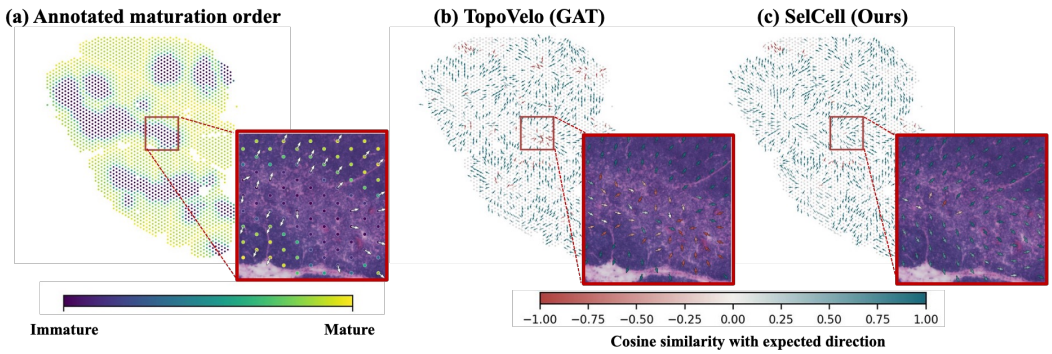


Figure 7: **SelCell recovers an anatomically consistent field.** (a) The annotation-based maturation order that TopoVelo adopts for this tissue (immature $\rightarrow$ mature)—a reference, not ground truth. Inside the zoom window (red box) the expected direction reverses across the central band: above it points up, below it down (median $R^2 = 0.58$). (b) The baseline (CBDir 0.183, Tab. 2), arrows colored by cosine with the local expected direction (Appendix C.6), follows the reference on the upper side only ($+0.49$ vs -0.04 below, though the motion-explained signal there is above the tissue median). (c) SELCELL (CBDir 0.232) tracks the reference's reversal on both sides ($+0.79$/$+0.75$). The window is chosen by a declared max-disagreement rule (Appendix C.6).

C.5 TRANSFER TEST ON RADAR-MEASURED SURFACE CURRENTS

The advective part of the observation model is shared with scalar velocimetry (§2), so the selection rule can be tested where the recovered motion is *measured*: sea-surface temperature maps play the role of expression, and high-frequency radar measures the true surface current. On 16,200 queries the same estimator equations and the same R^2 ranking were applied, preregistered under the certificate-1 selection protocol. Ranking by R^2 raises the directional alignment of the kept queries by $+0.044$, three times the random-ranking null of ± 0.015 (Fig. 9). The honest caveat: the selected queries' *absolute* speed error grows by $+0.127$ m/s, because R^2 prefers strong-flow regions. Both numbers are consistent with what certificate 1 certifies—a direction-level validity, not a speed claim—and are reported as measured.

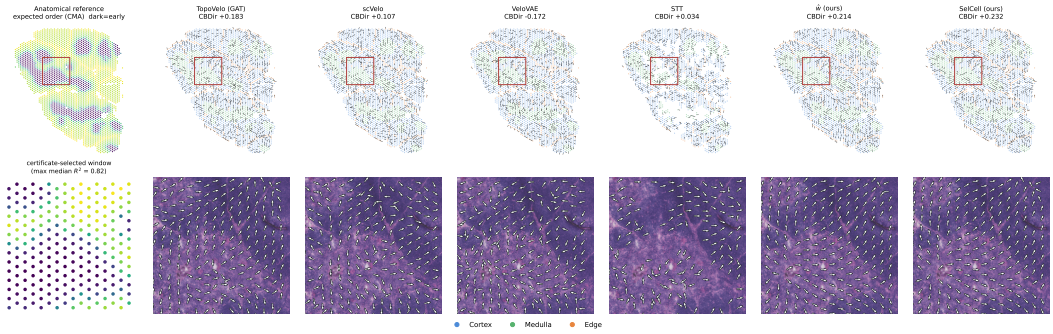


Figure 8: Six estimated fields on thymus S1 (companion to Fig. 7): the four baselines exposing a spatial field on our grid (GAT, scVelo, VeloVAE, STT) alongside the prior-free inversion $\hat{w}$ and SELCELL. Leftmost: expected maturation order (CMA) and the certificate-selected window. Top: unit-normalized fields with CBDir against the expected order per panel. Bottom: the window on the H&E image with every cell drawn. STT ships on the authors' rescaled grid; its window is registered by bounding-box affine.

Algorithm 1 SELCELL

global: $c_s \leftarrow \mathrm{med}\|\hat{w}\|/\mathrm{med}\|w_{\mathrm{g}}^{\mathrm{raw}}\|$; $\quad w_{\mathrm{g}} \leftarrow c_s\, w_{\mathrm{g}}^{\mathrm{raw}}$; $\quad \gamma \leftarrow c_\gamma \cdot \mathrm{med}\|\hat{w} - w_{\mathrm{g}}\|$
for each cell **do**
  $\hat{J}, v \leftarrow$ local WLS and kernel smoothing at bandwidth ℓ (App. C.3)
  $\hat{w} \leftarrow M^{-1}Jv$; $\quad \hat{r} \leftarrow v - J^\top \hat{w}$ (Eq. 2)
  $R^2 \leftarrow 1 - \|\hat{r}\|^2/\|v\|^2$; $\quad \hat{\tau}^2 \leftarrow \|\hat{r}\|^2/G$ (certificates 1–2: raw material)
  $\lambda \leftarrow \hat{\tau}^2/\gamma^2$; $\quad w_B \leftarrow$ Eq. (4) (spend certificate 2)
  $w^\star \leftarrow R^2\, w_B + (1 - R^2)\, w_{\mathrm{g}}$ (spend certificate 1)
  output $w^\star$ with provenance (g, g_k)
end for
report $w^\star$ as "w" only in tissues holding the wave bound (check certificate 3; App. B.4)

C.6 PROTOCOLS BEHIND THE CERTIFICATE EXHIBITS

Metrics (Tabs. 1–2). CBDir is the mean cosine alignment of a source cell's arrow with the displacements to its target-cluster neighbours, per annotated boundary; K-CBDir(k) scores agreement with the local differentiation order among k spatial neighbours under a shared time key (on STT's grid, which carries no common time key, the annotation order serves as the key); ICCoh is the mean cosine agreement of the field among spatial neighbours. Where a ground-truth field exists (synthetics only), VelAcc is the per-cell cosine to the true vector, and the grid of Tab. 5 adds endpoint and median angular error—all undefined on tissue. All community metrics use TopoVelo's official implementation and protocol unchanged, reported at face value.

Selection curves (Fig. 3a, Fig. 10). Each method ranks its own field by its own confidence, fixed by the preregistered protocol before evaluation: ours is R^2; scVelo's is its official velocity confidence (the mean cosine of a cell's velocity with its neighbours'); TopoVelo-GAT and VeloVAE define no confidence of their own, so the same standard consistency score is computed on their velocity layers; STT's is the negative entropy of its own membership probabilities. All neighbourhoods are spatial kNN ($k = 50$, the chain's graph). The per-cell score being averaged is CBDir's per-source-cell ingredient—the mean cosine between a cell's arrow and the displacements to its target-cluster neighbours—because a risk–coverage curve needs a per-instance loss; the official CBDir of Tab. 2 is left untouched. $\Delta_{.4}$ is $\mathrm{curve}(0.4) - \mathrm{curve}(1.0)$, the preregistered reporting point, and the grey band is the 95% range of $\Delta_{.4}$ under $B = 100$ random rankings. The curve's keep-all level (0.208) differs from the official CBDir (0.232) only by aggregation weighting: cells pooled equally versus per-boundary means. The score also transfers across fields: ranking the baseline's own arrows by our R^2 yields $\Delta_{.4} = +0.102$ on the same protocol—the selection power lives in the score, not the

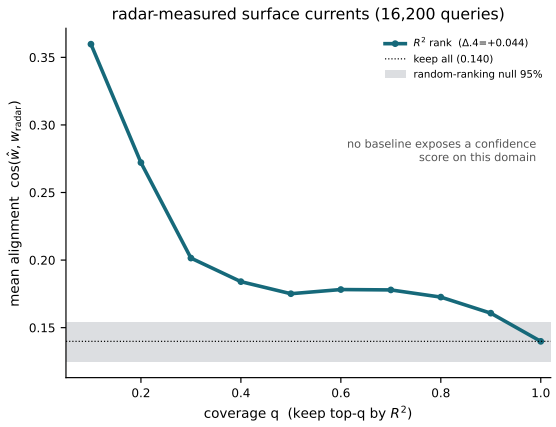


Figure 9: Certificate-1 selection on radar-measured surface currents: mean alignment $\cos(\hat{w}, w_{\mathrm{radar}})$ of the kept queries against coverage q; grey band = random-ranking null (95%).

field—and the kept cells are not concentrated at cluster boundaries (median distance to the nearest other-cluster cell: 110.3 for the top-40% versus 110.4 for all scored cells).

Half-split error (Fig. 3b–c). The error law is verified without ground truth by gene half-split self-consistency: for each of $K = 10$ random splits of the G genes into disjoint halves, the inversion of Eq. (2) is run independently on each half, giving two measurements of the same estimand whose per-direction discrepancy $d_k = u_k^\top(\hat{w}^{(A)} - \hat{w}^{(B)})$ is observable. When Eq. (3) holds, d_k has standard deviation $\sqrt{(\tau_A/\sigma_k^A)^2 + (\tau_B/\sigma_k^B)^2}$ —the two halves' error bars in quadrature. Panel (b) plots, per cell and axis, the observed sd of d_k across the K splits (small-sample c_4-corrected) over this prediction (log-ratio histogram; the quoted slopes are per-axis log–log regressions of observed on predicted), and panel (c) scores the coverage of the corresponding 90% band by predicted-width decile. Because any error *shared* by the two halves cancels in d_k, this verifies the noise-propagation component only—the declared scope of §3.3.

Split-conformal baseline (Fig. 3c). Cells are split $50/50$ (seed 0) into calibration and evaluation halves; the constant band width is the 90% quantile of $|d_k|$ on the calibration half (both directions pooled), and coverage and width are measured on the evaluation half. Adaptive-conformal variants are in Appendix C.9.

Window rule (Fig. 7). Among fully tissue-covered windows straddling the reversal band and passing certificate 1 in aggregate (median $R^2 \geq 0.5$), the exhibit shows the one where the two fields disagree most.

Expected direction (Fig. 7, Fig. 8). The per-cell expected direction is the unit spatial gradient of the annotation order $t = -\mathrm{CMA}$ (early $\rightarrow$ late), estimated by the same kernel-weighted local-linear regression used for $\hat{J}$ (Appendix C.3) and undefined where the annotation is locally flat. Arrows in the exhibits are colored by the cosine between the field and this direction; it is a reference derived from anatomy, not ground truth.

C.7 Certificate-1 selection on the synthetic benchmarks

C.8 Orthogonal-design synthetic: the estimand algebra in isolation

Remark 1 says the chain measures $\bar{w} = w - c$; this exhibit isolates that algebra with migration and wave placed on *orthogonal* axes, so that—unlike a tissue whose anatomy aligns maturation with migration—the two cannot be conflated by construction. On an analytic random-feature landscape ($N = 3{,}000$ cells, $G = 500$ genes) we set $w = w_0\, e_x$ and plant an intrinsic wave

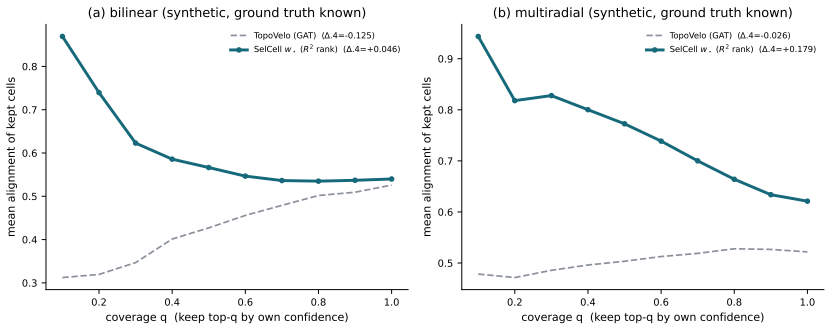


Figure 10: Certificate-1 selection where ground truth is known: mean alignment $\cos(F, v_{\text{true}})$ of the kept cells against coverage q on (a) bilinear and (b) multiradial. Per the preregistered protocol, the synthetic arm compares against the strongest baseline's confidence only (TopoVelo-GAT); the full own-confidence comparison is the real-tissue claim (Fig. 3a). $\Delta_{.4} = +0.046$ (bilinear) and $+0.179$ (multiradial) against GAT's -0.125 and -0.026.

$c = c_0\, e_y$ of relative size $c_0/w_0 \in \{0, 0.5, 1, 2\}$. Measured: the migration component of the inversion is unchanged across all four settings ($\hat{w}_x/w_0 = 1.000$–1.002), the wave enters exactly as $-c$ ($\hat{w}_y/w_0 = -0.501/-1.002/-2.004$ against expected $-0.5/-1/-2$), and certificate 1 is blind to the interchange (median $R^2 = 0.985$–0.988 throughout)—as it must be, since both causes produce the same observation. The exhibit confirms both halves of the estimand statement: the chain recovers $w - c$ exactly, no more and no less, which is why the license to report "w" is issued by certificate 3 and can never be issued by fit quality. (Script: `o2_orthogonal.py`.)

C.9 Adaptive-conformal baselines for certificate 2

The main-text conformal baseline uses a constant width, which targets marginal coverage only; adaptive conformal constructions exist, so we run the comparison they call for (difficulty-normalized split conformal, $50/50$ calibration split, same half-split protocol as §4.3). Two difficulty scores are compared. (i) With *our* $\hat{\tau}/\hat{\sigma}_k$ as the difficulty score, normalized conformal is conditionally valid (worst decile $0.881/0.889/0.893$ on bilinear/multiradial/thymus) at 1.02–$1.09\times$ our median width—conformalizing certificate 2's own score reproduces certificate 2 with exact marginal coverage, at the price of a calibration split. (ii) With generic difficulty scores (local density, neighbor smoothness), the same construction loses conditional validity (worst decile $0.862/0.745/0.826$) at 2.2–$2.7\times$ the width. The comparison locates the value in the score, not the wrapper: adaptivity comes from the model-derived $\hat{\tau}/\hat{\sigma}_k$, and any conformal wrapper is welcome to consume it. (Script: `diag_conformal.py`.)

C.10 Robustness and sensitivity

The full certificate-ablation grid follows: every certified dataset (bilinear, multiradial, thymus S1 and S3; S2 is uninformative under certificate 3 and excluded), all four consumption variants of §3.5, and every metric the dataset admits (ground-truth metrics exist on the synthetics only).

Sensitivity of the results to the single a-priori constant c_γ (grid $\{0.5, 1, 2\}$) and a masked variant of the global medians (restricted to the top-σ_2 quartile) are reported here with the ablation; all numbers live with the results (§4.3).

D Extended related work

Table 6 situates the chain against the closest methods along the five elements of our contribution; Table 7 pins each element to its evidence and to what it does *not* claim. The individual lineages are discussed below.

Table 5: Full certificate-ablation grid. Variants as in the main ablation table: raw inversion $\hat{w}$; + certificate 2 only (λ); + certificate 1 only (g); SELCELL (both). Arrows mark metrics where lower is better.

	Variant	VelAcc↑	CBDir↑	ICCoh↑	K-CBDir(5)↑	EPE↓	MedAng↓
bilinear	$\hat{w}$	0.255	0.318	0.422	0.678	0.905	58.2
	+ cert. 2	0.300	0.350	0.443	0.699	0.901	54.1
	+ cert. 1	**0.541**	**0.628**	0.546	**0.855**	**0.822**	**31.1**
	SELCELL	0.540	0.627	**0.547**	0.854	**0.822**	31.0
multiradial	$\hat{w}$	0.623	0.611	0.656	0.914	0.875	**24.1**
	+ cert. 2	**0.670**	**0.619**	**0.696**	**0.929**	**0.860**	24.6
	+ cert. 1	0.622	0.531	0.659	0.909	0.862	27.9
	SELCELL	0.621	0.528	0.656	0.908	0.854	27.8

thymus S1	CBDir↑	ICCoh↑	K-CBDir(3)↑	K-CBDir(5)↑	Δ_{40}
$\hat{w}$	0.214	**0.744**	0.399	**0.533**	+0.115
+ cert. 2	0.215	0.746	0.400	**0.533**	+0.115
+ cert. 1	**0.232**	0.692	**0.406**	0.530	+0.100
SELCELL	**0.232**	0.692	**0.406**	0.530	+0.099

S3, CBDir↑	$\hat{w}$	+ cert. 2	+ cert. 1	SELCELL
	0.270	0.276	0.343	**0.345**

Table 6: **Scope of inference in related methods**, against the five elements of the SELCELL chain. ✓: provided by the method as published; ×: not provided (possibly outside the method's stated goals; criteria definitions in the *Comparison criteria* paragraph).

Method	Motion vs. intrinsic obs. model	Training-free, backbone-agnostic	Per-cell trust vs. shuffled null (abstains)	Direction-wise error law (inversion step)	Motion-vs-wave audit (injection-calibrated)
(a) RNA-velocity backbones					
scVelo (Bergen et al., 2020)	×	×	×	×	×
veloVI (Gayoso et al., 2024)	×	×	×	×	×
(b) Spatial RNA-velocity methods					
TopoVelo (Gu et al., 2026)	×	×	×	×	×
STT (Zhou et al., 2024)	×	×	×	×	×
spVelo (Long et al., 2025)	×	×	×	×	×
veloAgent (Raghavan et al., 2026)	×	✓	×	×	×
STEER (Liu et al., 2026)	×	×	×	×	×
(c) Post-hoc uncertainty for spatial transcriptomics					
TISSUE (Sun et al., 2024)	×	✓	×	×	×
(d) Flow inversion with uncertainty					
Simoncelli et al. (Simoncelli et al., 1991)	×	×	×	✓	×
SELCELL (ours)	✓	✓	✓	✓	✓

Comparison criteria. Column 1 asks whether the method writes an observation model separating motion-driven from intrinsic expression change, with a named estimand for what its output recovers (§3). Column 2: applies training-free on top of any RNA-velocity backbone. Column 3: a per-cell trust decision read against a negative (cell-shuffled) control, with abstention on failure (§3.2). Column 4: a closed-form direction-wise error law for the expression-gradient inversion step—not for the upstream RNA-velocity posterior (§3.3). Column 5: an audit distinguishing motion from a travelling expression wave, with injection-calibrated sensitivity (§3.4). A × records that the criterion is not provided by the method as published and may lie outside its stated goals: veloAgent and STEER *disclaim* the motion reading of projected vectors rather than certifying it, and veloVI's posterior quantifies the upstream RNA-velocity fit, not the spatial inversion. Simoncelli et al. give the direction-wise law for optical flow—the analogue our certificate 2 derives for expression gradients.

Table 7: **Element-wise contribution, evidence, and claimed scope.** Each row names one element, the claim the paper makes for it, where that claim is measured, and the neighbouring claim it does *not* make.

Element	Claim	Evidence	Not claimed
Certificate 1 (*where*)	the only ranking found that improves the kept cells; transfers across fields	$\Delta_{.4} = +0.099$ vs. the baselines' own confidences -0.095 to -0.012 (Fig. 3a); applied to the GAT field, $+0.102$ (App. C.6)	certification of true motion
Certificate 2 (*how much*)	conservative, order-optimal error bars	decile-wise coverage near 0.9 where constant split conformal drops to 0.55–0.70 at 0.67–0.82$\times$ the width (Fig. 3c); minimax bound (App. A.8); adaptive conformal conditionally valid only with $\hat{\tau}/\hat{\sigma}_k$ as its difficulty score (App. C.9)	coverage of shared (common-mode) bias
Certificate 3 (*what*)	injection-calibrated bound on tissue-coherent waves	S1: silence + planted κ=0.5 waves alarm in every tested direction $\Rightarrow$ $\|c\| \leq 0.5 \cdot \mathrm{med}\|\hat{w}\|$; three-way ledger (App. B.4)	a bound on arbitrary per-cell confounds
Inversion + fusion	outperforms the strongest baseline on certified tissue	CBDir 0.232 vs 0.183, non-overlapping CIs (Tab. 2); full ablation including the multiradial fallback cost $0.611 \to 0.528$ (Tab. 5)	universal fusion superiority

Trajectory inference from temporal marginals. A mature line of work infers population dynamics from snapshots taken at successive times: partial-identifiability results for dynamics fit to marginals (Hashimoto et al., 2016), the observation that near-equilibrium marginals cannot see divergence-free components of the drift (Weinreb et al., 2018), consistency of entropic-OT trajectory inference under gradient dynamics (Lavenant et al., 2024), and scalable extensions (Chizat et al., 2022; Neklyudov et al., 2023; Klein et al., 2025). Our setting differs in its data: a single snapshot with no second time point, whose usable signal is the spatial gradient of expression—an object that marginal formulations do not use even when spatially informed transport costs use the coordinates themselves (Klein et al., 2025).

Foundation and world models for tissue. Single-cell foundation models (Cui et al., 2024; Theodoris et al., 2023) and their spatially aware successors (Tejada-Lapuerta et al., 2025; Birk et al., 2026) encode space, when at all, through neighbor graphs or proximity ranks; generative models of sequential slides model coordinates directly but require multiple time points (Sakalyan et al., 2025). In none of these are the metric gradients of expression that our observation model stands on exposed at the representation level (§1). Their goals are orthogonal to ours, and their outputs can serve as upstream inputs to our chain wherever an RNA-velocity layer is exposed.

Optical flow, scalar velocimetry, and errors-in-variables (full lineage). Beyond the correspondences in §2: the map of M's smallest singular value is the Shi–Tomasi criterion (Shi & Tomasi, 1994; Förstner & Gülch, 1987); scalar imaging velocimetry was developed in Dahm et al. (1992); Su & Dahm (1996), and the same inverse problem underlies ocean-current estimation from sea-surface temperature (Kelly, 1989; Chen et al., 2008) and fluid optical flow (Liu & Shen, 2008; Heitz et al., 2010). Gradient-noise bias in such inversions is classical (Kearney et al., 1987; Fuller, 1987; Golub & Loan, 1980); Appendix A.9 gives the direction-resolved account for locally regressed Jacobians.

Selective prediction and learned confidences. Unlike learned confidences (Aodha et al., 2013) or flow-specific uncertainty heads (Wannenwetsch et al., 2017; Ilg et al., 2018), certificate 1's score needs no training data and transfers across domains unchanged; certificate 2's distribution-level promise is compared against split conformal (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023) in §4.3; certificate 3 follows the negative-control design of Lipsitch et al. (2010).

Spatial gradients and uncertainty in spatial transcriptomics. GASTON (Chitra et al., 2025) models spatial expression gradients through a learned scalar topography, answering the complementary question drawn in §1. TISSUE (Sun et al., 2024) quantifies uncertainty on imputed expression; our uncertainty is on the recovered cell velocity and is derived from the observation model rather than calibrated post hoc.