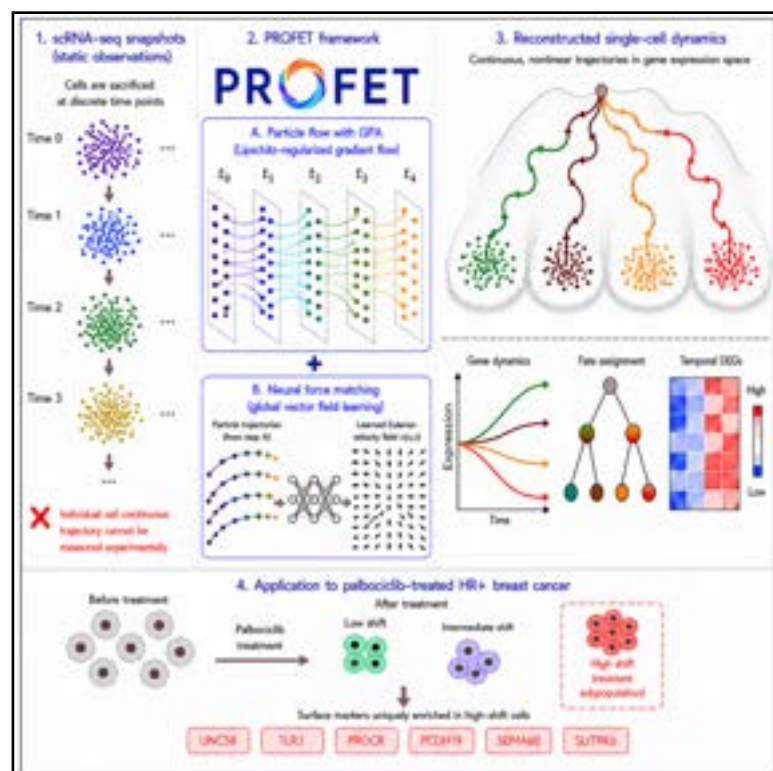


PROFET predicts continuous gene expression dynamics from scRNA-seq data to elucidate heterogeneity of cancer treatment responses

Graphical abstract



Authors

Yu-Chen Cheng, Hyemin Gu, Thomas O. McDonald, ..., Rinath Jeselsohn, Markos A. Katsoulakis, Franziska Michor

Correspondence

markos@umass.edu (M.A.K.), michor@jimmy.harvard.edu (F.M.)

In brief

Cheng and Gu et al. introduce PROFET, a framework that reconstructs continuous single-cell trajectories from sparse scRNA-seq time series. By combining particle-based gradient flows with neural force matching, PROFET accurately captures nonlinear cell-state dynamics and reveals heterogeneous responses and candidate surface markers associated with breast cancer treatment resistance.

Highlights

- PROFET reconstructs continuous single-cell trajectories from sparse scRNA-seq time series
- PROFET outperforms ten trajectory inference methods across diverse datasets
- PROFET recovers nonlinear cell-state divergence and convergence dynamics
- PROFET identifies resistant breast cancer cells and candidate surface markers

Article

PROFET predicts continuous gene expression dynamics from scRNA-seq data to elucidate heterogeneity of cancer treatment responses

Yu-Chen Cheng,^{1,2,3,4,20} Hyemin Gu,^{5,20} Thomas O. McDonald,^{1,2,3,4} Wenbo Wu,⁶ Shubham Tripathi,⁷ Cristina Guarducci,⁸ Douglas Russo,¹ Daniel L. Abravanel,^{8,9,17} Madeline Bailey,^{1,10} Yue Wang,¹¹ Yun Zhang,¹² Yannis Pantazis,¹³ Herbert Levine,^{14,15} Rinath Jeselsohn,^{8,16,17,18} Markos A. Katsoulakis,^{5,*} and Franziska Michor^{1,2,3,4,18,19,21,*}

¹Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA

²Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA

⁴Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

⁵Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA, USA

⁶Harvard-MIT Health Sciences and Technology, Cambridge, MA, USA

⁷Yale Center for Systems and Engineering Immunology and Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA

⁸Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

⁹Harvard Medical School, Boston, MA, USA

¹⁰Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

¹¹Irving Institute for Cancer Dynamics and Department of Statistics, Columbia University, New York, NY, USA

¹²State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

¹³Institute of Applied and Computational Mathematics, Foundation for Research and Technology-Hellas, Heraklion, Greece

¹⁴Center for Theoretical Biological Physics, Northeastern University, Boston, MA, USA

¹⁵Department of Physics, Northeastern University, Boston, MA, USA

¹⁶Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA

¹⁷Breast Oncology Center, Dana-Farber Cancer Institute, Boston, MA, USA

¹⁸The Broad Institute of MIT and Harvard, Cambridge, MA, USA

¹⁹The Ludwig Center at Harvard, Boston, MA, USA

²⁰These authors contributed equally

²¹Lead contact

*Correspondence: markos@umass.edu (M.A.K.), michor@jimmy.harvard.edu (F.M.)

<https://doi.org/10.1016/j.cels.2026.101710>

SUMMARY

Single-cell RNA sequencing (scRNA-seq) profiles cellular heterogeneity but captures only static snapshots, limiting inference of gene expression dynamics. We developed PROFET (particle-based reconstruction of generative force-matched expression trajectories), a framework that reconstructs continuous, nonlinear single-cell trajectories from sparsely sampled scRNA-seq time series. PROFET combines a particle-based gradient-flow algorithm with simulation-free force matching to accurately infer cellular dynamics. Across mouse and human *in vitro* datasets and an *in vivo* axolotl regeneration dataset, PROFET achieved 2.6–12.5× lower prediction error than ten state-of-the-art trajectory inference methods. Applying PROFET to newly generated scRNA-seq data from a palbociclib-treated MCF7 cell line and three published breast cancer patient datasets, we reconstructed treatment-response trajectories and identified a resistant cell subpopulation exhibiting large phenotypic shifts and enrichment of the surface markers *UNC5B*, *TLR3*, *PCDH19*, *PROCR*, *SLITRK6*, and *SEMA6B*. PROFET provides a biologically grounded framework for reconstructing cell-state dynamics from static single-cell data across development, regeneration, and therapeutic response. A record of this paper's transparent peer review process is included in the supplemental information.

INTRODUCTION

Modern live single-cell imaging technologies, such as those using fluorescently labeled reporter proteins, offer high temporal resolution for tracking gene expression dynamics at the single-

cell level.¹ Despite this advantage, live-cell experimental approaches are limited by the small number of markers that can be tracked simultaneously, challenges in maintaining system stability, and the complexity of analyzing imaging data over prolonged periods.² In contrast, single-cell RNA sequencing

(scRNA-seq) technologies provide invaluable insights into cellular heterogeneity by capturing genome-wide expression patterns^{3,4} but face challenges because expression patterns can vary over time due to cell-state changes, intrinsic gene expression noise, and experimental variability. Moreover, scRNA-seq data are inherently cross-sectional as individual cells must be lysed for sequencing at each time point. These constraints limit the ability to reconstruct true single-cell trajectories from noisy time-course scRNA-seq datasets.⁵

As a result, neither live-cell imaging nor scRNA-seq alone can provide both the high temporal resolution and the broad gene expression profiling required to capture dynamic changes in cellular states. To bridge this gap, recent data-driven approaches have leveraged mathematical frameworks from optimal transport (OT)^{6–13} and spline interpolation¹⁴ to infer cellular trajectories by linking distributions of cell states across discrete time points. The majority of existing methods are grounded in OT theory, spanning a range of formulations: static OT maps (Waddington-OT [WOT]⁶ and CellOT¹³), Schrödinger bridges (DeepRUOT¹²), Sinkhorn-based schemes (MIOFlow,⁸ PI-SDE,¹¹ and VGFM¹⁰), and Benamou-Brenier dynamic OT (TrajectoryNet⁷ and TIGON⁹). Others construct trajectories via spline interpolation with OT-derived couplings (multi-marginal flow matching [MMFM]¹⁴). Many of these formulations employ a quadratic transport cost, leading to the Wasserstein-2 (W2) distance. In dynamic OT, finding the most cost-efficient path can be viewed as a control-theoretic approach that interpolates between two probability measures by minimizing kinetic energy over the entire path, often augmented with additional constraints to enforce flow consistency or incorporate biological priors.^{7,9} Although these approaches have enabled important advances in trajectory inference, W2-based formulations can remain sensitive to complex biological data geometries, including multimodal distributions, branching trajectories, heavy-tailed distributions, and manifold structure.^{7,8,15}

To address these limitations, we developed PROFET (particle-based reconstruction of generative force-matched expression trajectories), a computational framework that formulates trajectory inference as a Wasserstein gradient flow of the Lipschitz-regularized Kullback-Leibler (KL) divergence—mathematically equivalent to a Wasserstein-1 (W1) proximal regularized KL divergence—which enables robust trajectory inference in three respects. First, this divergence imposes no structural or moment conditions on the target distribution, requiring only a finite first-moment condition on the source—a mild assumption that holds in essentially all practical settings.¹⁵ Second, the Lipschitz constraint on the dual potential ensures that the induced velocity field has a uniformly bounded norm, yielding finite-speed dynamics that guarantee numerical stability under forward Euler discretization. Third, the dissipative structure of the Wasserstein gradient flow admits a greedy sequential implementation—each time step solves a local variational problem independently—rather than requiring global optimization over the entire trajectory, enabling adaptive time horizons and resolution sufficient to resolve transport through complex geometries.

In this study, we systematically benchmark PROFET against ten representative methods—most rooted in OT but differing in training methodology: discretize-then-optimize (DTO) (MIOFlow,⁸ PRESCIENT,¹⁶ PI-SDE,¹¹ and DeepROUT¹²),

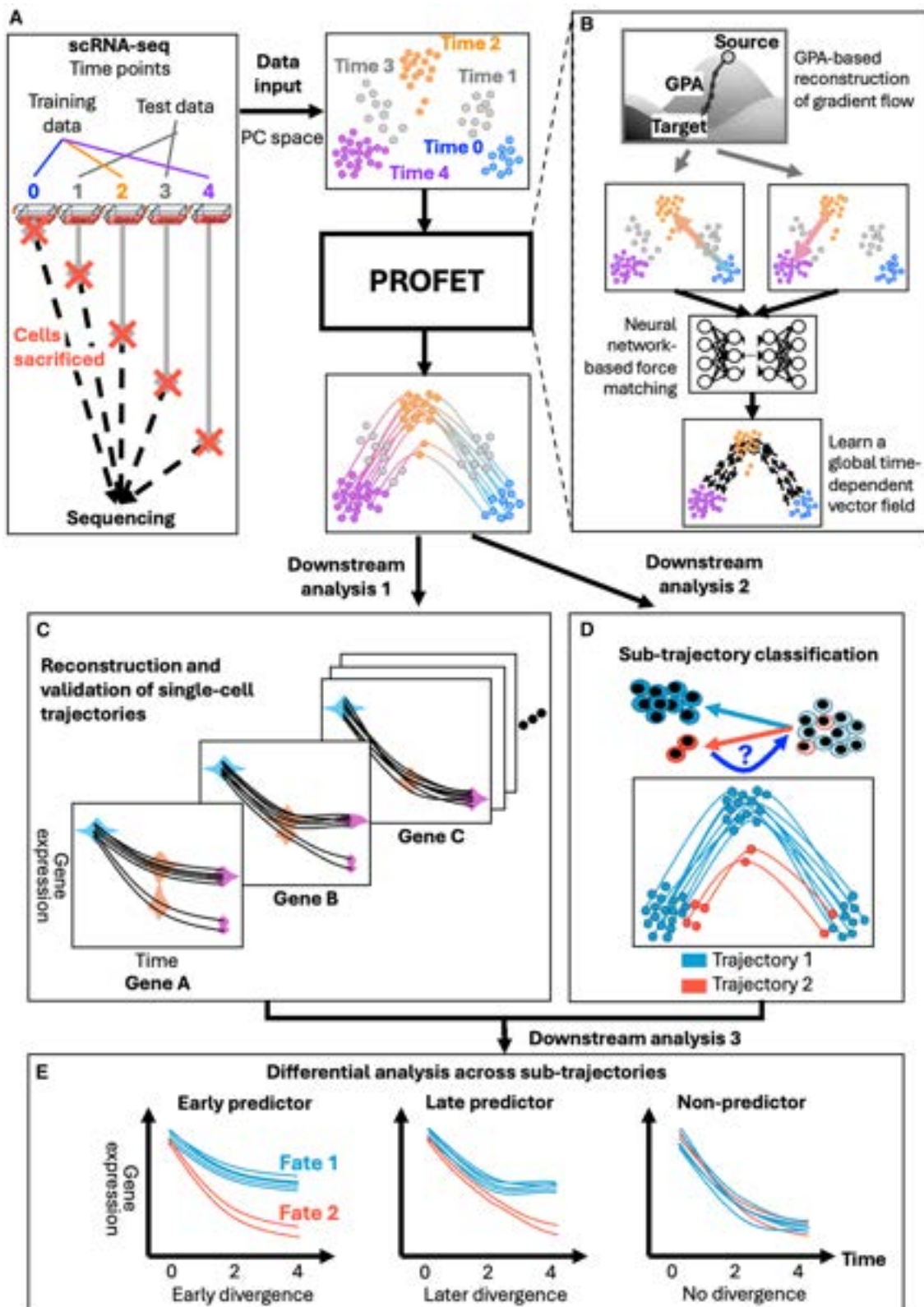
adjoint-based methods (TrajectoryNet⁷ and TIGON⁹), simulation-free methods (MMFM¹⁴ and VGFM¹⁰), and static transport maps (WOT⁶ and CellOT¹³). Using *in vitro* and *in vivo* mouse, axolotl, and human datasets, we demonstrate that PROFET consistently outperforms competing methods across varying levels of biological complexity while remaining data-agnostic and scalable. Across all test datasets, PROFET achieved 2.6- to 12.5-fold lower W2 error than competing methods, based on the geometric mean of W2 error ratios, while maintaining comparable runtime and memory requirements.

For downstream analyses, we then applied PROFET to a scRNA-seq dataset from the MCF7 hormone receptor-positive (HR+) breast cancer cell line,¹⁷ as well as three datasets from HR+ metastatic breast cancer patients,^{18,19} to reconstruct single-cell trajectories spanning pre- and post-palbociclib time points. Palbociclib, a CDK4/6 inhibitor, has extended progression-free survival in HR+ breast cancer; however, acquired resistance to CDK4/6 inhibition has been widely reported in both clinical and preclinical studies.^{20,21} In this study of the MCF7 cell line, 16 weeks of palbociclib treatment led to experimentally confirmed drug resistance, accompanied by a marked downregulation of cell-cycle-related genes. To compare this response with that obtained from patient-derived data, we identified a subpopulation of cells in the patient samples that exhibited similarly pronounced phenotypic shifts, characterized by the same downregulation signature of cell cycle genes. Differential gene expression (DEG) analysis revealed that six surface markers—*UNC5B*, *TLR3*, *PCDH19*, *PROCR*, *SLITRK6*, and *SEMA6B*—were enriched in the pre-treatment states of these high-shift cells. These markers may serve as early predictors of phenotypic transitions in response to palbociclib treatment and as potential targets for antibody-drug conjugates (ADCs).²² More broadly, PROFET provides a generalizable approach for reconstructing gene expression trajectories from time-series scRNA-seq data across phenotypically heterogeneous cell populations.

RESULTS

Overview of PROFET

To reconstruct complete trajectories from discrete scRNA-seq measurements, PROFET employs a two-step framework to infer the underlying temporal dynamics using force-based generative models, on whose output a deep neural network is then trained (Figure 1A). The first step of PROFET is based on the generative particle algorithm (GPA),^{23,24} which interpolates gene expression patterns by propagating initial cell states (i.e., the source) along the gradient of a potential landscape (Figure 1B; Notes S1 and S2). Unlike OT-based formulations that rely on W2-based interpolation, GPA defines trajectories as a Wasserstein gradient flow of a Lipschitz-regularized KL divergence with respect to the target distribution. The Lipschitz-regularized KL divergence is well-defined on a broad class of distributions, imposing no structural or moment conditions on the target—in particular, requiring no finite second moment assumption—and yields a well-defined potential function for the gradient flow.¹⁵ The Lipschitz continuity of the potential induces finite-speed dynamics, with the Lipschitz bound serving as a speed limit, ensuring numerical stability of the forward Euler discretization via the Courant-Friedrichs-Lewy (CFL) condition (STAR Methods; Note



(legend on next page)

S2). Furthermore, the dissipative structure of the Wasserstein gradient flow naturally admits a greedy sequential implementation, where each time step solves an independent local variational problem. This enables adaptive time horizons (see [STAR Methods](#), “stopping time for GPA”), providing sufficient resolution to capture transport through complex geometries—in contrast to control-theoretic formulations of dynamic OT, where the time horizon and integration resolution are fixed a priori and typically implemented as amortized flow maps. Collectively, these properties endow PROFET with both theoretical stability guarantees and empirical robustness against trajectory collapse or divergence during training.

In the second step, the outputs from GPA are used to train a neural network that learns a continuous, time-dependent vector field defined on a Eulerian grid using a force-matching scheme,²⁵ bridging the trajectory-based (Lagrangian) and field-based (Eulerian) representations of cell dynamics²⁶ ([STAR Methods](#); [Note S3](#)). In terms of training methodology, this step amortizes the piecewise particle trajectories from step 1 into a globally smooth velocity field in a simulation-free manner. As a result, the complexity of the training signal scales naturally with the data geometry determined in step 1. This procedure enforces temporal consistency across subintervals and yields a continuous vector field that can be integrated from arbitrary initial conditions. The resulting object is therefore a time-dependent flow map, analogous to those obtained in dynamic OT methods, but derived through a fundamentally different and more robust route. This design is biologically motivated: in many systems, particularly under perturbations (e.g., drug treatment or mutation), the underlying regulatory landscape is non-stationary and evolves over time.

Together, the principled gradient-flow formulation and the two-step training pipeline enable robust modeling of complex biological dynamics—including multimodal splitting and merging processes, the emergence and disappearance of cell types, and strong external perturbations such as drug treatments—where trajectories become highly nonlinear and time dependent, and accurate mode prediction is biologically critical. The amortized velocity field further enables downstream applications such as fate prediction, gene dynamics reconstruction, and phenotypic shift analysis from arbitrary initial conditions.

For PROFET development, we used synthetic data to guide hyperparameter selection, including the convergence criteria for particle dynamics, the Lipschitz constraint, and the dimensionality of the latent space. To this end, we generated synthetic single-cell expression trajectories using a stochastic gene regulatory network model²⁷ of the epithelial-mesenchymal transition (EMT), producing temporally sparse observations that mimic

time-series scRNA-seq measurements ([STAR Methods](#); [Figure S1A](#)). PROFET was trained on a subset of time points and evaluated by comparing the reconstructed states with the withheld intermediate states, allowing us to identify stable parameter ranges that produced robust and non-divergent trajectories ([Figure S1B](#); [Note S4](#)). The synthetic data were used exclusively for hyperparameter optimization. For each biological dataset analyzed subsequently, PROFET was trained directly on the time-series scRNA-seq data using a subset of time points to learn temporal dynamics and reconstruct withheld intermediate states ([Figure S1C](#)).

Building on this foundation, we performed three key downstream analyses to interpret PROFET outputs in biologically meaningful ways ([STAR Methods](#)). First, to reconstruct single-gene dynamics, we applied an inverse transformation to map the inferred single-cell trajectories—originally reconstructed in a low-dimensional principal component (PC) space—back to the full gene expression space defined by all input genes. This step enables recovery of continuous gene expression dynamics for every gene in every individual cell ([Figure 1C](#)). Second, we implemented two strategies to classify predicted trajectories. The first assigns trajectories to predefined cell subpopulations based on clustering cells at a specific reference time point, allowing interpretation in terms of fate or origin ([Figure 1D](#) illustrates this classification applied to terminal fates). The second strategy quantifies the magnitude of cell-state transitions across time by computing trajectory-wise Euclidean distances, which are then thresholded into distinct transition groups. Finally, by integrating the above classifications, we performed time-resolved differential gene expression analysis across sub-trajectories. This approach enabled us to identify not only which marker genes distinguish different trajectories but also when these genes begin to diverge. These analyses uncover temporal transcriptional programs that underlie divergent treatment responses ([Figure 1E](#)).

Model validation and benchmarking across four biological datasets

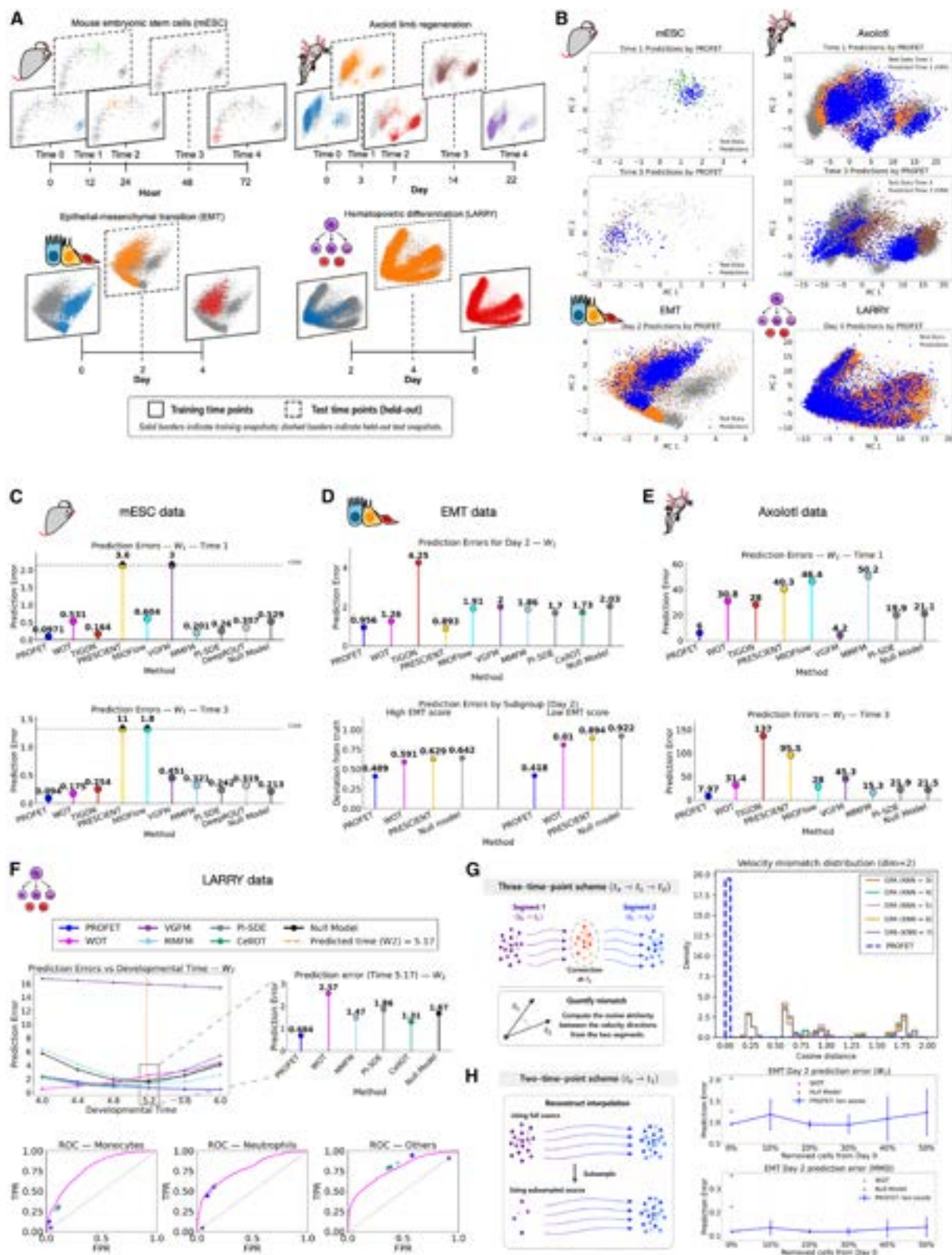
For validation and benchmarking, we compared PROFET against ten representative trajectory inference methods broadly rooted in OT, realized through different mathematical models and categorized into four distinct training methodologies—static transport maps, adjoint-based (neural ordinary differential equation [neural ODE]), DTO, and simulation-free (flow matching)—each with different individual features. WOT⁶ and CellOT¹³ compute static transport maps between pairs of time points, where WOT employs unbalancedness. TrajectoryNet⁷ and TIGON⁹ are adjoint-based methods that learn continuous dynamics via neural

Figure 1. Reconstruction of continuous single-cell gene expression dynamics from discrete scRNA-seq data

(A) Study objective and approach. Left: existing scRNA-seq technologies capture only static snapshots of cell states, limiting the ability to observe continuous trajectories. Right: to address this limitation, we developed PROFET to reconstruct continuous gene expression dynamics from sampled scRNA-seq time points (colored), while masking intermediate time points (gray) for validation.

(B) Two-step architecture of PROFET. Step 1: apply a generative particle algorithm (GPA) to infer gradient flows interpolating between observed distributions at adjacent time points. Step 2: use a neural network to perform force matching on these flows and learn a globally smooth, time-dependent vector field that models cell dynamics.

(C–E) Three key downstream analyses enabled by our method. (C) Prediction of continuous single-cell gene expression dynamics. (D) Classification of predicted trajectories into terminal fates, enabling tracing fate-assigned cells to their origins. (E) Identification of diverse temporal gene expression patterns associated with distinct fates, illustrated by three representative genes: left panel shows early divergence, middle shows late divergence, and right shows no divergence.



(legend on next page)

ODEs with backpropagation through trajectories, incorporating unbalancedness. Among DTO methods, MIOFlow⁸ learns nonlinear data manifolds, PRESCIENT¹⁶ employs a probabilistic energy-based formulation with unbalancedness, DeepROUT¹² combines unbalancedness with stochasticity, linking to the Schrödinger bridge, and PI-SDE¹¹ uses physics-informed stochastic differential equation (SDE) formulations. MMFM¹⁴ and VGFM¹⁰ adopt simulation-free flow matching, with VGFM additionally incorporating unbalancedness. A comprehensive comparison of these methods, including their mathematical frameworks, training strategies, and methodological features, is provided in Tables S1 and S2, while the methodological novelties of PROFET are discussed in detail in Note S5. We also included a null model, defined as random coupling between cell states across time points, which serves as a baseline without learned temporal structure.

We systematically benchmarked PROFET against these methods across four biological datasets spanning diverse biological systems and a broad range of nonlinear biological dynamics, including *in vitro* datasets profiling mouse embryonic stem cell (mESC) differentiation^{28,29} and transforming growth factor β (TGF- β)-induced EMT in the human MCF10A cell line,³⁰ as well as *in vivo* axolotl limb regeneration data³¹ and the LARRY mouse hematopoiesis dataset.³² To ensure consistent evaluation, the same training-and-testing strategy was applied across datasets by withholding intermediate time points during training and evaluating each method based on its ability to reconstruct the held-out cellular distributions. The mESC dataset contains five time points: 0, 24, and 48 h were used for training and 12 and 36 h were withheld for testing. This dataset exhibits a curved, detour-like trajectory in gene-expression space rather than a direct linear progression. The EMT dataset contains three time points: days 0 and 4 were used for training and day 2 was withheld for testing. It exhibits redistribution of cell states, transitioning from a bimodal structure at day 0 to a unimodal structure at day 4. The axolotl dataset contains five time points: days 0, 7, and 22 were used for training and days 3 and 14 were withheld for testing. This dataset displays oscillatory

dynamics; the number of modes increased from days 0 to 7 and subsequently decreased from days 7 to 22. Finally, the LARRY dataset contains three time points: days 2 and 6 were used for training and day 4 was withheld for testing. It captures mouse hematopoiesis, in which hematopoietic progenitor cells progressively diverge into multiple mature blood-cell lineages over time (Figure 2A).

Across the benchmark datasets, PROFET accurately predicted the held-out test time points (Figure 2B) and consistently achieved strong performance, with 2.6–12.5 $\times$ lower W2 error than competing methods (geometric mean of error ratios across all datasets and time points; Table S3). Although some methods performed competitively on individual datasets or metrics, their performance varied substantially across biological systems (Figures 2C–2E). In contrast, PROFET maintained stable performance across all benchmark datasets. In the mESC dataset, PROFET achieved the lowest prediction error at both held-out time points (12 and 36 h) across all evaluated metrics, including W2 distance, Sinkhorn divergence, and maximum mean discrepancy (MMD). The second-best method, TIGON, exhibited approximately 1.7-fold higher error, whereas all remaining methods showed more than 2-fold higher errors relative to PROFET (Figure 2C; Figure S2).

Moving from the mESC dataset to the EMT dataset, the complexity of the underlying dynamics increases, as the cellular distribution transitions from a bimodal to a unimodal structure over time. The number of modes is biologically meaningful: by computing an EMT score for each cell (STAR Methods), we identified two distinct subpopulations ($p = 0.0001$; Wilcoxon rank-sum test; Figure S3). In this setting, although PRESCIENT achieves comparable W2 error to PROFET (1.06 versus 0.89; Figures 2D, upper and S4), it fails to capture the underlying multimodal structure. Specifically, it predicts a single dominant mode, overestimating the abundance of the mesenchymal subpopulation by more than 60% relative to the ground truth while underestimating the epithelial subpopulation by approximately 90%. Similar behavior is observed for other methods, which also

Figure 2. Validation and benchmarking of PROFET

- (A) Overview of the four time-series scRNA-seq datasets, indicating the time points used for model training and those reserved for testing.
- (B) Principal component (PC) plots of the mESC, EMT, axolotl, and LARRY scRNA-seq datasets comparing PROFET predictions with the experimentally observed ground-truth cell-state distributions at the held-out test time points.
- (C) mESC dataset: quantitative prediction errors measured as Wasserstein-2 (W2) distances between the predicted cell-state distributions and the experimentally observed ground-truth distributions at time 1 (upper row) and time 3 (lower row).
- (D) EMT dataset: the upper row shows quantitative prediction errors between the predicted and experimentally observed cell-state distributions at the held-out intermediate time point (day 2), measured using the W2 metric. The lower row quantifies the absolute deviations in the predicted proportions of the two EMT subpopulations relative to the ground truth at day 2, corresponding to overestimation of the high-EMT subgroup and underestimation of the low-EMT subgroup.
- (E) Axolotl dataset: quantitative prediction errors measured as W2 distances between the predicted cell-state distributions and the experimentally observed ground-truth distributions at time 1 (upper row) and time 3 (lower row).
- (F) LARRY dataset: the upper row shows prediction errors across developmental times from days 4 to 6, measured using the W2 metric. The vertical dashed line indicates the predicted developmental time, estimated from the ratio of W2 distances between the days 2–4 and 4–6 intervals; the inset highlights prediction errors for competing methods at this estimated time point. The lower row shows prediction performance for the three cell fates using receiver operating characteristic (ROC) curves, where TPR (true positive rate) measures the fraction of correctly identified cells belonging to a given fate, and FPR (false positive rate) measures the fraction of incorrectly assigned cells from other fates.
- (G) Ablation analysis for three-time point interpolation: the left panel illustrates the computation of vector field mismatch at the connected intermediate time point using cosine distance. The right panel shows the distribution of cosine distances for PROFET and GPA-only reconstructions. For GPA-only, the analysis depends on the choice of K-nearest neighbors (KNN); details are provided in Note S7.
- (H) Ablation analysis for two-time point interpolation: the left side illustrates subsampling cells from the source time point to evaluate robustness under reduced training data. The right side shows prediction performance under source-cell subsampling from 10% to 50%, compared with the baseline WOT and null models trained using the full dataset. To account for stochastic variability in the subsampling procedure, each subsampling level was repeated across ten independent random seeds, and error bars represent the standard deviation across the ten independent runs.

show a strong bias toward the dominant mode (Figure 2D, lower). In contrast, PROFET more accurately reconstructs both populations, with estimated abundances closer to the ground truth (40% overestimation for the mesenchymal subpopulation and 30% underestimation for the epithelial subpopulation; Figure 2D, lower). Notably, at the held-out intermediate time point, PROFET retained approximately 70% of the epithelial subpopulation, whereas all competing methods retained only ~10%, representing an approximately 7-fold improvement in recovering the transient minority population.

In the axolotl dataset—the most complex system analyzed here, featuring oscillatory dynamics with both divergence and convergence of cell states (Figure S5)—PROFET outperforms all methods except VGFM. Specifically, PROFET achieves W2 errors approximately 3- to 8-fold lower than competing models at both withheld time points (time 1, day 7; time 3, day 14) across all distributional metrics (Figures 2E and S5). Although VGFM shows comparable performance at time 1 (0.9-fold relative to PROFET), its error grows to approximately 5-fold that of PROFET at time 3. This degradation likely arises from VGFM's tendency to evolve too slowly, preventing convergence to the target distribution within the required temporal window (Figure S5).

Furthermore, we performed two evaluations using the LARRY dataset. First, for interpolation benchmarking, we leveraged the three available time points (days 2, 4, and 6) in the hematopoietic system. Following the same setup as for the EMT dataset, we trained the models using days 2 and 6 and evaluated predictions on the held-out intermediate time point (day 4). Previous studies have shown that cells at days 4 and 6 are already close to mature states, and day 4 approaches a quasi-stationary distribution. To account for this observation, we further estimated an effective developmental time for the held-out distribution by comparing relative distances between distributions (day 2 → day 4 and day 6 → day 4) across multiple metrics, yielding consistent estimates (day 5.17 using W2 and day 5.66 using MMD). Across this broad range of estimated intermediate times (days 5 and 6), PROFET consistently achieved the lowest prediction error among all methods (Figures 2F, upper and S6), demonstrating robustness to uncertainty in temporal alignment. Second, leveraging the lineage barcode information unique to the LARRY dataset, we evaluated the cell-to-cell matching performance of PROFET and the other methods. Following the original study, we assessed each method's ability to recover progenitor-fate relationships for major hematopoietic lineages (e.g., neutrophils, monocytes, and others) using the true positive rate (TPR) and the area under the receiver operating characteristic (ROC) curve (AUC) metrics (STAR Methods). Although lineage matching is not the primary objective of PROFET, we found that it achieves performance comparable to that of established approaches such as WOT, as well as recent methods, including CellOT, VGFM, and MMFM (Figure 2F, lower). These results indicate that, despite not being explicitly designed for lineage reconstruction, PROFET preserves biologically meaningful relationships between progenitor and descendant cells.

This analysis establishes that alternative methods exhibit clear limitations when applied to temporally sparse and highly nonlinear biological datasets. In particular, MMFM, WOT, TIGON, and PRESCIENT underestimate the number of cellular

subpopulations present at transient states, such as day 2 in the EMT dataset. Capturing these cellular subpopulations is biologically essential, as they reflect heterogeneous cell identities and state transitions. In addition, VGFM and PRESCIENT show deficiencies in aligning learned dynamics with the underlying temporal scale. VGFM tends to evolve too slowly, often failing to reach even the first intermediate time point and not converging to the target distribution, whereas PRESCIENT exhibits the opposite behavior, overshooting the target distribution (e.g., in the mESC dataset, Figure S2). These mismatches indicate limitations of these methods in accurately modeling the temporal progression of cellular dynamics. Furthermore, PI-SDE, DeepROUT, and MIOFlow perform poorly on sparse and highly nonlinear datasets, such as mESC and axolotl (Figures 2C, 2E, S2, and S5), frequently producing collapsed or distorted trajectory reconstructions. CellOT, which is designed for pairwise transport between two time points under a W2 OT framework, achieves reasonable performance in two-time point settings, similar to WOT. However, when evaluated on interpolation tasks involving three-time point datasets (Figures 2D for EMT and 2F for LARRY), CellOT exhibits lower accuracy in predicting the masked intermediate time point compared with PROFET. Finally, TrajectoryNet, a continuous normalizing flow (CNF)-based model, fails to converge across the benchmark datasets, using both the parameter settings recommended in the original study and additional hyperparameter configurations, and generates implausible trajectories that deviate from the withheld intermediate distributions (Note S6).

Ablation analysis, robustness, and computational scalability

To further assess the robustness and component-wise contributions of PROFET, we performed systematic ablation and sensitivity analyses. For the ablation analysis, we evaluated the contribution of the neural force-matching step (step 2) to the GPA component (step 1) under two data schemes: (1) datasets with more than two observed training time points (e.g., the mESC dataset) and (2) datasets with only two observed training time points (e.g., the EMT dataset). This distinction is important because the force-matching step is expected to provide different benefits depending on the amount of temporal information available during training.

In multi-time point settings, the primary role of the neural force-matching step, step 2, is to ensure temporal consistency across adjacent trajectory segments. Using only the GPA component, step 1, trajectories are reconstructed independently between neighboring time points, which can result in inconsistencies in the inferred velocity fields at shared intermediate states. To illustrate this effect, we evaluated the mESC dataset, where trajectories are independently reconstructed by GPA between time points 0–2 and 2–4. The intermediate time point (time 2) therefore serves as the connection point between the two GPA-derived trajectory segments. To quantify the degree of inconsistency between these independently reconstructed segments, we compared velocity vectors inferred from the two adjacent segments at this shared time point (Figure 2G, left). Because GPA does not provide explicit cell-to-cell correspondences between independently reconstructed segments, we estimated local correspondences using a K-nearest-neighbor

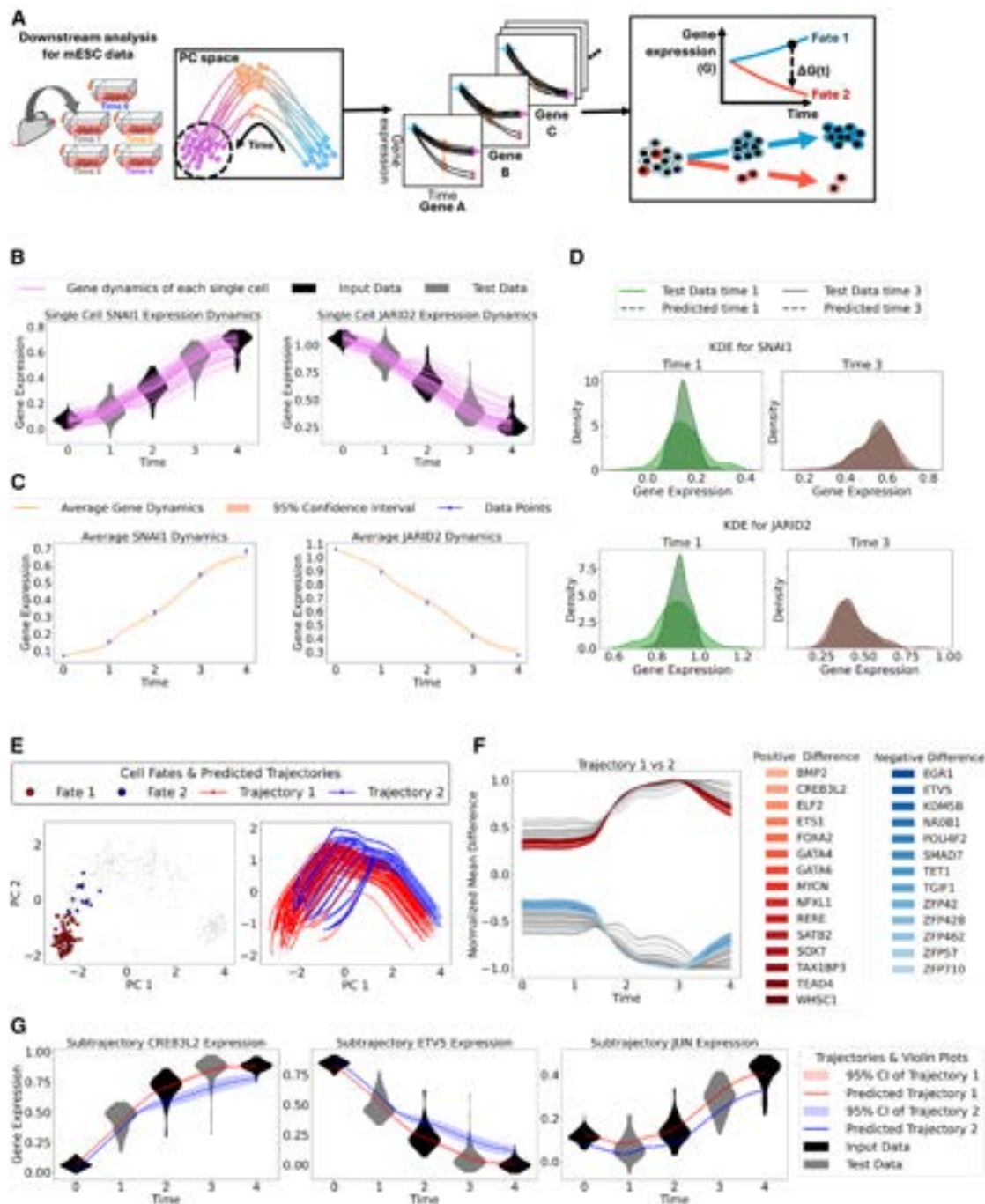


Figure 3. Downstream analysis of reconstructed single-cell trajectories for mESC data

(A) Overview of the downstream analysis pipeline for mESC differentiation. The process begins with PROFET-reconstructed single-cell trajectories in PCA space. These trajectories are then inverse-mapped back to the original gene expression space to recover dynamic gene expression profiles for each gene in each cell. Subsequently, we investigate gene expression differences across distinct sub-trajectories to identify the timing and magnitude of divergence associated with different terminal fates. $\Delta G(t)$ denotes the temporal difference in mean gene expression between sub-trajectories.

(B) Single-cell, single-gene dynamics plotted as gene expression levels (y axis) over time (x axis), with each green curve representing an individual cell trajectory. The predicted trajectories are compared against real data distributions using violin plots (gray for test data, black for training data). The violin plots summarize the distributions of gene expression across all cells at the corresponding time points.

(C) Predicted average gene expression trajectories (orange curves with shaded regions indicating 95% confidence intervals across reconstructed cell trajectories) compared with actual gene expression levels from real data (blue dots indicating the mean, with error bars representing standard deviation).

(D) Comparison of predicted and actual gene expression distributions at day 1 (green) and day 3 (brown). Predicted distributions are shown as dashed lines, with real data distributions shown as solid lines. The distributions summarize gene expression across all cells at the corresponding time points.

(legend continued on next page)

(KNN) approach and computed the cosine distance between matched velocity vectors (Note S7). We found that step 2 reduces these velocity mismatches by integrating local GPA-derived flows into a unified continuous-time vector field, lowering cosine distances and improving the alignment of the inferred dynamics (Figure 2G, right). These results indicate that step 2 resolves local inconsistencies and produces smoother, more temporally coherent trajectories than using step 1 alone.

In two-time point settings, step 1 alone is sufficient to generate interpolation trajectories, allowing us to isolate the contribution of step 2 to data efficiency, robustness, and generalization. To assess data efficiency, we performed systematic cell-subsampling experiments, in which only 10%–50% of the source cells were retained, while preserving the overall distribution of cell states (Figure 2H, left). Under this setting, step 1 learns trajectories only for the selected subset of cells. Step 2 then learns a global continuous-time vector field from these trajectories, enabling the model to generalize beyond the sampled cells and reconstruct dynamics throughout the broader state space. Across all subsampling levels, the full PROFET framework maintained prediction accuracy comparable to that obtained using the complete dataset, demonstrating reduced sensitivity to sample size and improved robustness under limited-data conditions (Figure 2H, right). Each subsampling level was repeated across ten independent random subsamples to account for stochastic variability. Together, these results demonstrate that the force-matching step (step 2) not only improves temporal consistency but also enhances the stability and generalization of the learned dynamics.

In addition to the ablation analysis, we performed sensitivity analyses across the key hyperparameters of PROFET. Specifically, we varied the GPA stopping criterion, the dimensionality of the latent space, and the Lipschitz constraint parameter. Across both the EMT and mESC datasets, PROFET maintained stable performance over a broad range of parameter settings, indicating that the framework is robust to moderate variations in model configuration (Figure S7). In particular, performance remained stable across Lipschitz constraint values ranging from 1 to 5 in the EMT dataset, whereas only a mild degradation was observed at larger values in the mESC dataset (Figure S7). Overall, these results demonstrate that PROFET is not highly sensitive to hyperparameter selection and retains strong predictive performance across diverse settings. Detailed results for the sensitivity analyses are provided in Figure S7 and Note S7.

To assess computational efficiency and scalability, we further evaluated runtime and memory usage across datasets of varying sample sizes and dimensions (Note S8). PROFET's runtime scales moderately with both dimension (487 to 1,085 s from $d = 2$ to $d = 16$) and sample size (487 to 2,743 s from $n = 278$ to $n = 34,317$), while peak memory remains constant across dimensions (~ 0.36 GB) and grows to 3.95 GB only on the largest

dataset—well below that of methods that retain full computational graphs (DeepRUOT: 7.41 GB, MIOFlow: 6.70 GB, VGFM: 6.83 GB). On the largest benchmark, the LARRY dataset, PROFET was 13 $\times$ faster than TIGON and 34 $\times$ faster than TrajectoryNet. Despite taking finer, geometry-adaptive integration steps, PROFET's runtime remains comparable to DTO methods (PRESCIENT, PI-SDE, and DeepRUOT) because each forward Euler step is a greedy local computation with no backward pass through the trajectory. Methods that reduce to a single integration step per time interval (MIOFlow and VGFM) or bypass simulation entirely (MMFM) are faster but sacrifice temporal smoothness or modeling flexibility (Table S4).

Predicting gene expression and differentiation pathways from single-cell trajectories

PROFET enables reconstruction of continuous single-cell gene expression trajectories for all genes in a multi-time point scRNA-seq dataset. We first set out to analyze the mESC differentiation dataset. To this end, we applied an inverse principal component analysis (PCA) transformation to project the reconstructed trajectories from the low-dimensional PCA space back to the original high-dimensional gene expression space, defined by all genes in the input data (STAR Methods). By aligning the trajectories with subgroup classifications, PROFET then enables analysis of gene expression differences across distinct sub-trajectories (Figure 3A).

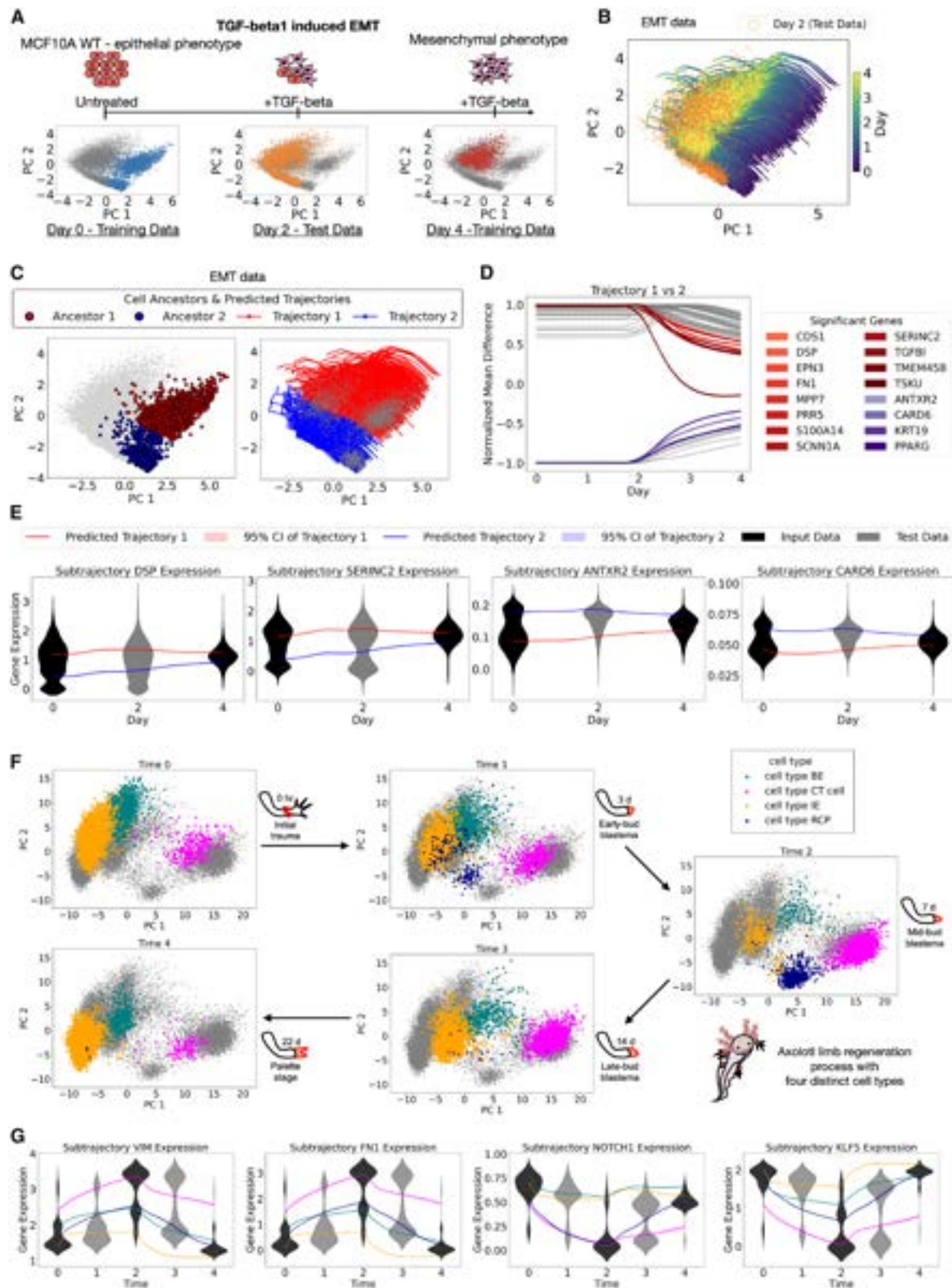
By performing this downstream analysis, we recovered the gene expression dynamics for all genes across individual cells in the mESC differentiation dataset. We highlight two representative genes, *JARID2* and *SNAI1*, both of which are known regulators of embryonic development^{33,34} (Figure 3B). Using three time points (time points 0, 2, and 4) for model input, PROFET successfully reconstructs full gene expression trajectories for individual cells, validated by accurate predictions at withheld time points (time points 1 and 3, treated as test data). To quantify prediction accuracy, we compared the mean gene expression along the reconstructed trajectories to the experimental data, finding that, for both test time points 1 and 3, the experimental means (with error bars) consistently fall within the 95% confidence intervals of our predicted means (Figure 3C). We further compared the full distributions of predicted and measured gene expression, observing no statistically significant differences for both genes (one-sided permutation test, $p > 0.05$ and STAR Methods; Figure 3D). Complete results for all other genes are provided in Figures S8–S10 and Table S5.

We next sought to investigate whether single-cell trajectories diverge into distinct cell fates during mESC differentiation. We performed cluster analysis on the final time point of the scRNA-seq dataset (STAR Methods), leading to the identification of two clusters representing different fate subpopulations. Differential gene expression analysis suggested that one trajectory group (fate 1) was enriched for endoderm- or extraembryonic

(E) Classification of differentiation fates and ancestral trajectory tracing. Left panel: two distinct cell fate subgroups at day 4, identified through clustering analysis. Right panel: ancestral trajectories of the two subgroups traced back to earlier time points.

(F) Temporal quantification of gene expression divergence between fate groups. Normalized mean differences are plotted over time, with genes exceeding a divergence threshold of 0.5 highlighted. Each curve represents one gene. Positive values indicate higher expression in trajectory group 1, and negative values indicate higher expression in trajectory group 2.

(G) Examples of gene expression divergence in the two subpopulations. Three representative genes are shown, with average gene expression trajectories plotted separately for each subgroup (shaded regions indicate 95% confidence intervals across reconstructed cell trajectories).



(legend on next page)

endoderm-associated genes,³⁵ including *Gata4*, *Gata6*, *Sox7*, *Foxa2*, and *Bmp2*, whereas the other trajectory group (fate 2) retained higher expression of pluripotency- or early stem state-associated genes,³⁶ such as *Zfp42/Rex1*, *Nr0b1*, *Kdm5b*, *Tet1*, and *Zfp57* (Figure 3E, left). Leveraging the reconstructed single-cell trajectories, we assigned the final state of each trajectory to one of the two fate clusters by computing its distance to the cluster centroids and selecting the closer one (STAR Methods). This approach enabled us to group trajectories into two fate-specific subpopulations (Figure 3E, right).

To identify genes that diverge between the two sub-trajectories, we computed the temporal difference in mean gene expression, normalized by the maximum difference between the groups from days 0 to 4 (STAR Methods). Genes with a normalized difference greater than 0.5 were classified as divergent (Figure 3F). The continuity of reconstructed trajectories also enabled us to pinpoint when these divergences emerged with high temporal resolution. We identified genes whose expression began to diverge between time points 1 and 2, shedding light on the timing of fate decisions (Figure 3F). Two representative genes, *CREB3L2* and *ETV5*, showed changes exceeding 0.5, in contrast to *JUN*, which exhibited a smaller change of 0.1. Their mean expression profiles across the two trajectory groups (Figure 3G) highlight distinct temporal patterns of gene expression divergence. Complete results for all other genes are provided in Figure S11.

Predicting cell-state convergence in EMT dynamics

We then performed downstream analyses on the reconstructed trajectories from the EMT dataset. This dataset exhibits cell-state convergence—transitioning from two distinct subpopulations at day 0 to a single population by day 4 (Figure 4A). However, with only discrete time points available in scRNA-seq data, the complete trajectory and timing of this convergence remain unclear. We thus used PROFET to reconstruct the full trajectories of cell states by interpolating between the initial (day 0) and final (day 4) distributions while withholding the intermediate time point (day 2) for validation (Figures 4B and S12A–S12D).

We again applied an inverse PCA transformation to reconstruct continuous single-cell, single-gene trajectories for all genes in the EMT dataset. Analysis of two example genes, *DSP*³⁷ and *SERINC2*,³⁸ showed that their trajectories converged from two distinct modes at day 0 into a single mode by day 4 (Figure S12E). PROFET also captured the intermediate bimodal

structure at day 2, which had been withheld as test data (Figure S12F). One-sided permutation tests revealed no statistically significant differences between the predicted and observed distributions on day 2 ($p > 0.05$), supporting the accuracy of PROFET (STAR Methods). Complete results for all other genes are provided in Figures S13 and S14 and Table S6.

Unlike the mESC dataset, where trajectory divergence was observed, the EMT dataset reveals evidence of convergence in cellular states. To further explore this observation, we performed clustering analysis on the day 0 data and identified two ancestral subpopulations (Figure 4C, left). By classifying each reconstructed trajectory based on its proximity to these ancestral clusters, we traced their respective differentiation paths toward a common final state (Figure 4C, right).

To identify genes associated with the convergence process, we calculated the normalized difference in mean gene expression between the two trajectory groups, as described in the previous section. Genes whose expression differences declined by more than 0.5—representing a reduction of over 50% from their maximal divergence—between days 0 and 4 were considered strongly associated with convergence (Figure 4D). A key advantage of PROFET, which reconstructs continuous trajectories using force matching (STAR Methods), is its ability to precisely localize the timing of convergence. We observed that a major reduction in inter-trajectory expression differences occurred after day 2, as illustrated by four representative genes. Each of these genes exhibited a decline of over 50% in mean expression differences between the two sub-trajectories from days 2 to 4, highlighting their involvement in the convergence process in the later phase of EMT (Figure 4E).

We next set out to investigate temporal phenotypic switching across diverse cell types in the axolotl dataset. Unlike the *in vitro* EMT dataset, the axolotl limb regeneration data include four pre-annotated cell types—basal epidermis (BE), intermediate epidermis (IE), connective tissue (CT), and *Prdx2*⁺ blastema cells (RCP)—originally defined in the source study³¹ (Figure 4F). Divergence among these four cell types is most prominent at the mid-bud blastema stage (time 2, day 7; Figure 4F). However, the full gene-expression dynamics of these cell types, including the early EMT-like phase following injury (from days 0 to 7) and the later mesenchymal-to-epithelial transition (MET)-like phase (from days 7 to 22), were not resolved in the original analysis.³¹

Using PROFET, we reconstructed continuous trajectories across these stages in PC space and applied the same

Figure 4. Downstream analysis of reconstructed single-cell trajectories for EMT and axolotl data

- (A) Illustration of EMT scRNA-seq data.
- (B) Application of PROFET to reconstruct continuous trajectories using only data from time points 0 and 4, with time point 2 held out for validation. The colormap represents the predicted time along the trajectory, ranging from time 0 to time 4, and the test data are highlighted by orange circles.
- (C) Classification of ancestral cell subgroups and forward trajectory predictions. Left: two distinct cell subpopulations at time 0, identified via clustering analysis. Right: forward-predicted trajectories of these two subpopulations.
- (D) Normalized mean differences in gene expression between the two trajectory groups are plotted over time. Each curve represents one gene. Positive values indicate higher expression in group 1, while negative values indicate higher expression in group 2. Genes exhibiting a drop greater than 0.5 in normalized mean difference at specific time points are highlighted, indicating pronounced convergence in expression levels between the two trajectories.
- (E) Representative examples of gene expression convergence. Four genes are shown, with average gene expression forward trajectories plotted separately for each ancestral subpopulation. Shaded regions indicate 95% confidence intervals across reconstructed cell trajectories.
- (F) Illustration of the axolotl scRNA-seq dataset, highlighting the cell-state distributions of four cell types across five time points (days 0, 3, 7, 14, and 22).
- (G) Representative examples of gene expression divergence. Four genes are shown, with average gene expression trajectories plotted separately for each cell type identified at the mid-bud blastema stage (day 7). Shaded regions indicate 95% confidence intervals across reconstructed cell trajectories. The color of each curve corresponds to the cell-type legend shown in (F).

downstream analyses described previously to map the trajectories back to the original gene-expression space, obtaining dynamic expression profiles for every gene. By examining EMT- and stemness-related genes identified from the Molecular Signatures Database (MSigDB),³⁹ we predicted complete EMT- and stemness-associated gene dynamics spanning the EMT-to-MET transition (Figures 4G, S15, and S16). Four representative genes are shown in Figure 4G: *VIM* and *FN1*, markers of mesenchymal cells, and *NOTCH1* and *KLF5*, associated with stemness. PROFET disentangled their gene-expression trajectories across all four cell types, revealing clear divergence during the EMT phase (from time 0 to 2) and convergence during the MET phase (from time 2 to 4).

Tracing phenotypic shifts associated with palbociclib therapy in HR+ breast cancer

We then sought to deploy PROFET to investigate heterogeneity in treatment response to palbociclib, a CDK4/6 inhibitor used in HR+ breast cancer that is known to induce variable treatment responses.^{19,40} This analysis was performed across four datasets: one *in vitro* cell line model and three patient-derived samples. For the *in vitro* model, we analyzed scRNA-seq data from an MCF7 breast cancer cell line in which palbociclib resistance was established by escalating drug doses over a 16-week period¹⁷ (STAR Methods). The dataset includes both parental and resistant cells (Figure 5A).

In parallel, we analyzed longitudinal scRNA-seq data from three HR+ metastatic breast cancer patients. The first patient (PA3) was part of the study by Luo et al.¹⁸ Her initial biopsy was collected prior to palbociclib treatment, and two additional biopsies were both obtained 11 months after initiating palbociclib plus endocrine therapy. These latter samples were collected from ascitic fluid and confirmed disease progression; in our analysis, they were combined as a single post-treatment sample. The second (862) and third (887) patients were part of the study by Klughammer et al.¹⁹ For patient 862, biopsies were collected before and after 200 days of treatment with endocrine therapy plus palbociclib; the second biopsy was performed for research purposes. Patient 887 initially responded to therapy following the first biopsy, but clinical progression was observed prior to the second biopsy, which was collected at day 220. The patient started receiving palbociclib approximately 2 weeks before the second biopsy (Figure 5A; STAR Methods).

PROFET was developed and validated on datasets with lower dimensionality: a synthetic dataset with 26 genes, the mESC dataset with 100 genes, and the EMT dataset with 72 genes. In contrast, the four palbociclib-treated samples were assayed with genome-wide scRNA-seq, necessitating dimensionality reduction prior to analysis. However, due to high noise levels in these datasets, the first two PCs captured only 5%–10% of the total variance across samples, making direct PCA-based reduction insufficient.

To address this issue, we adopted a two-step dimensionality reduction strategy. First, we preselected 121 genes previously implicated in palbociclib response, including genes involved in cell-cycle regulation, MYC and estrogen receptor (ER) signaling, growth factor pathways, the Hippo pathway, and inflammatory processes^{41,42} (Table S7). In the second step, we applied PCA to the gene expression matrix restricted to this curated gene

set. In the resulting PCA space, we observed that the *in vitro* dataset exhibited the most distinct separation between pre- and post-treatment samples, while patient-derived datasets showed less clear separation, indicating heterogeneous responses to palbociclib (Figure 5B). To quantify these differences, we leveraged the continuous single-cell trajectories reconstructed by PROFET. For each cell, we computed the extent of phenotypic shift as the Euclidean distance between its inferred pre-treatment and post-treatment states in the gene expression space (STAR Methods). As expected, the *in vitro* dataset exhibited the largest average shift with low variability, while the patient samples displayed broader, more heterogeneous distributions (Figure 5B): the coefficients of variation (CVs) were 0.10 for the *in vitro* data compared with 0.43, 0.30, and 0.26 for patients PA3, 862, and 887, respectively (STAR Methods; Table S8), confirming greater phenotypic variability in the clinical setting.

Based on the relatively homogeneous treatment response observed in the MCF7 *in vitro* dataset—specifically in genes relevant to palbociclib treatment—we used this dataset as a reference to dissect heterogeneity in resistant cell states in patient samples. We focused on identifying subpopulations of patient-derived cells that underwent substantial phenotypic shifts analogous to those observed *in vitro*. To this end, we examined the distributions of phenotypic shift magnitudes and identified three major modes using local minima (STAR Methods; Figure 5B), allowing us to stratify cells in each patient sample into three subgroups: those marked by high, medium, and low phenotypic shifts, respectively. This classification, visualized in PCA space (Figure 5C), supports our hypothesis of heterogeneous treatment responses and enables quantification of which cells undergo large state transitions—an analysis not accessible from unpaired scRNA-seq data alone. This single-cell-level classification provides a framework for dissecting phenotypic heterogeneity in response to palbociclib and offers a powerful strategy for identifying candidate cell populations likely to undergo substantial phenotypic shifts in response to treatment.

Trajectory-based dissection of treatment response heterogeneity and early biomarker discovery

Using these trajectory-based subgroups, we next aimed to identify genes with the largest expression changes within the high-shift subgroup and to further pinpoint surface markers uniquely enriched in pre-treatment cells within this group (Figure 6A). We performed DEG analysis for all four datasets by comparing average gene expression levels between post-treatment samples and pre-treatment samples (STAR Methods). This analysis revealed that most cell-cycle-related genes were consistently downregulated following treatment across all datasets (Figure 6B). In particular, several members of the centromere protein family—*CENPE*, *CENPF*, *CENPM*, and *CENPV*—were significantly downregulated in the MCF7 cell line following palbociclib treatment ($-2.41 < \log_2$ fold change < -1.36 ; $p < 5e-10$, two-sided Welch's *t* test) and showed similar downregulation trends in the patient samples (Figure 6B). Complete DEG results are provided in Table S9.

In contrast, we observed divergent expression patterns of ER signaling-related genes between the cell line and patient samples PA3 and 862, both of whom received combination therapy with palbociclib and endocrine treatment (Figure 5A). In the cell

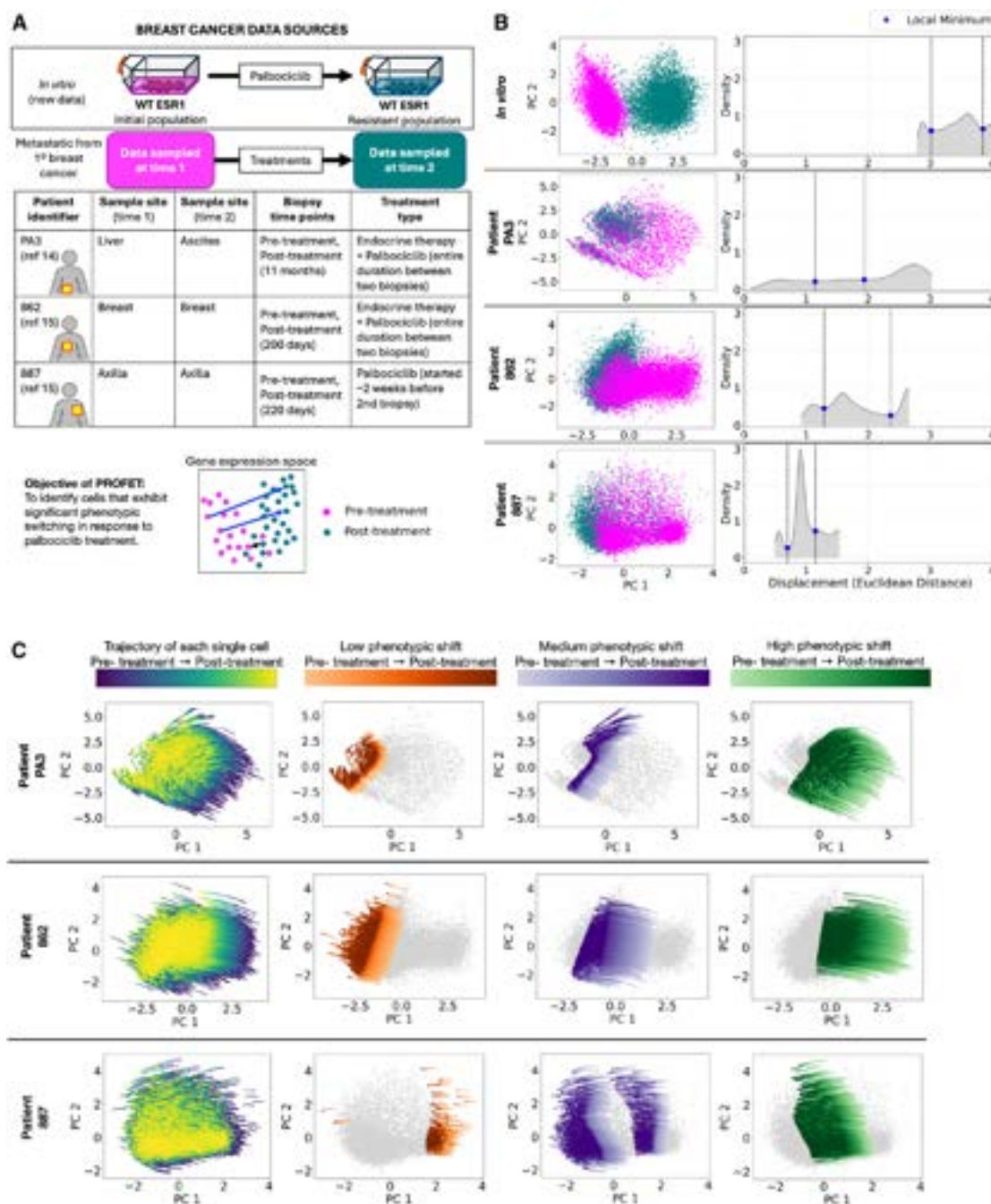


Figure 5. Reconstructed single-cell trajectories for palbociclib resistance

(A) Overview of the datasets analyzed (upper table): one *in vitro* MCF7 breast cancer scRNA-seq dataset (GEO: GSE326012) and three patient-derived datasets—one from Luo et al.¹⁸ and two from Klughammer et al.¹⁹ The lower panel illustrates the objective of applying PROFET to these datasets: to identify single cells that undergo significant phenotypic switching in response to palbociclib treatment.

(B) Each row represents one dataset. Left: PCA projection of scRNA-seq data, with pre-treatment cells in magenta and post-treatment cells in teal. Right: kernel density estimates (KDEs) of phenotypic shifts, defined as the Euclidean distances between initial and final cell states for each single-cell trajectory inferred by PROFET.

(C) Trajectories of resistance in HR+ breast cancer. Each row corresponds to one dataset. The first panel presents the full set of continuous single-cell trajectories reconstructed by PROFET. The subsequent three panels display subsets of trajectories grouped by their magnitudes of phenotypic shift, quantified as the Euclidean distances between initial and final cell states, as defined in (B).

line data, *AR*, *FOXA1*, *GHR*, *IGFBP4*, and *PGR* were upregulated, whereas these same genes were downregulated in the patient samples. *JUN* and *MYB* were downregulated in both the cell line and patient datasets, with more pronounced downregulation observed in the patient samples (Figure 6B for the full fold-change heatmap and tables). These results indicate that the profound downregulation of ER signaling-relevant genes is specific to the patient samples, while centromere genes were consistently downregulated in both the cell line and patient data. The well-established effects of these treatments—in which palbociclib primarily inhibits cell-cycle-related genes^{19,43} and endocrine therapy downregulates ER signaling-related genes^{44,45}—validate our observed trajectories, aligning with the treatment regimens applied to each sample (Figure 5A).

Because this DEG analysis reflects only population-average changes, it does not resolve the specific subgroups within each tumor that exhibit these phenotypic changes induced by the combination therapy. To further dissect this treatment response heterogeneity, we used PROFET to reconstruct gene expression dynamics at the single-cell, single-gene level. Here, we present results for two representative cell-cycle-related genes (*CENPF* and *CCNA2*) and two ER signaling-related genes (*MYB* and *IGFBP4*) (Figures 6C and 6D). Complete results for additional genes are provided in Figures S17–S20. In the *in vitro* dataset, *CENPF* and *CCNA2* showed uniform downregulation, *MYB* expression remained consistently steady, and *IGFBP4* was uniformly upregulated (Figure 6C). In contrast, the patient datasets displayed substantial variability, with the high-phenotypic-shift subgroup exhibiting more pronounced downregulation of all four genes compared with the low-shift subgroup (Figure 6D).

We then sought to identify early markers of the subpopulation of cells undergoing high phenotypic shifts in response to palbociclib resistance. To identify such markers, we performed DEG analysis of surface marker genes between the high- and low-phenotypic-shift groups at the pre-treatment time point.³⁷ This analysis revealed six genes—*UNC5B*, *TLR3*, *PCDH19*, *PROCR*, *SLITRK6*, and *SEMA6B*—that were consistently enriched in the high-shift subgroup across all three patient datasets (Figure 6E).

Among these genes, *UNC5B* has been reported to be significantly upregulated in breast cancer and is correlated with poor overall survival in breast cancer patients.⁴⁶ Activation of *TLR3* has been shown to increase cancer stem cell (CSC) markers, which are associated with breast cancer progression.⁴⁷ *PROCR* is known to activate several signaling pathways in breast cancer cells, including the extracellular signal-regulated kinase (ERK), phosphatidylinositol 3-kinase (PI3K)-Akt-mechanistic target of rapamycin (mTOR), and RhoA-Rho-associated coiled-

coil-containing protein kinase (ROCK)-p38 pathways.⁴⁸ Although *PCDH19* has not previously been implicated in metastatic breast cancer, it belongs to the protocadherin family; another member of this family, *PCDHA1*, has been associated with poor prognosis and immune cell infiltration in breast cancer.⁴⁹ Although *SEMA6B* and *SLITRK6* have not yet been linked to breast cancer, *SLITRK6* has been identified as a novel biomarker in hepatocellular carcinoma,⁵⁰ while *SEMA6B* has been associated with poor prognosis and an immunosuppressive tumor microenvironment in colorectal cancer⁵¹—suggesting that, based on our findings, both genes may serve as candidate markers of disease progression in HR+ metastatic breast cancer.

Taken together, these studies demonstrate that overexpression of these genes is strongly associated with more aggressive breast cancer and poorer prognosis. Building on this evidence, our findings—based on reconstructed phenotypic trajectories before and after treatment—further suggest that these genes may serve as novel surface markers capable of predicting which pre-treatment cells are likely to undergo profound phenotypic changes in response to palbociclib and endocrine therapy. In both the *in vitro* dataset and patient samples PA3¹⁸ and 887,¹⁹ where treatment resistance developed, these markers show promise for the early identification of treatment-resistant cells, paving the way for more targeted and proactive therapeutic strategies.

DISCUSSION

Here, we introduce PROFET, a computational framework for reconstructing continuous, nonlinear single-cell trajectories from limited and discrete scRNA-seq time points. PROFET addresses a fundamental limitation of traditional OT approaches, which interpolate between observed cellular distributions through cost-minimizing W2 geodesics. Although such W2-based methods provide efficient displacement between snapshots, many biological systems exhibit multimodal distributions, branching trajectories, and rare outlier cell states that remain challenging for W2-based trajectory inference. First, the W2 distance requires finite second moments; recent work showed that W2-regularized objectives in general failed to learn distributions lacking this property,¹⁵ whereas W1-regularized objectives succeeded. Second, dynamic OT formulations penalize kinetic energy, biasing inferred trajectories toward displacement-minimizing paths that can shortcut across branching or merging structures commonly observed in cell differentiation trajectories, rather than follow them.⁷ Even OT on learned nonlinear latent manifolds can fail to recover a smooth geometry from limited samples, and may produce unreliable interpolations when

(C) Representative examples of gene expression dynamics for individual cells in the *in vitro* MCF7 breast cancer dataset (GEO: GSE326012). Each magenta curve represents one reconstructed single-cell trajectory. Violin plots show the distributions of observed gene expression across all cells at each time point used as input for trajectory inference.

(D) Representative gene dynamics for individual cells across three patient datasets. Violin plots show real expression data; colored lines represent predicted trajectories from three subgroups of cells defined by low, medium, and high phenotypic shift levels, as described in Figure 5.

(E) Differential expression analysis of surface marker genes in pre-treatment cells from patient PA3 (Luo et al.¹⁸), comparing high versus low phenotypic shift subgroups. Statistical significance was assessed using a two-sided Welch's *t* test. Red dots indicate significantly differentially expressed genes ($p < 0.05$, $|\log_2 \text{fold change}| > 1.1$). Blue-labeled genes represent significantly upregulated genes (positive $\log_2 \text{fold change}$) in the high phenotypic shift subgroup of PA3 and also in the corresponding subpopulations of the other two patients (862 and 887) from Klughammer et al.¹⁹ The accompanying Venn diagram summarizes the overlap of differentially expressed surface-marker genes across all three datasets.

distinct cell populations from different time points overlap in space.⁵² Third, the control-theoretic formulation of dynamic OT is naturally implemented in amortized flow maps, where the time horizon and integration resolution are fixed a priori and may be insufficient for capturing complex geometries. Together, these limitations make it difficult for W2-based methods to learn realistic flows when the underlying biological dynamics involve complex geometry.

To address these limitations, we formulate trajectory inference as a Wasserstein gradient flow of the Lipschitz-regularized KL divergence—mathematically equivalent to a W1 proximal regularized KL divergence. The framework discretizes the Wasserstein gradient flow via forward Euler: at each step, the variational dual formulation of the Lipschitz-regularized KL functional is optimized, and particle positions are updated by taking the gradient of the dual potential greedily. The number of steps adapts to the complexity of the data geometry—more steps for branching or merging dynamics, fewer for simpler transitions. This discretization depth determines the resolution of the particle trajectories from which a neural velocity field is subsequently learned in a simulation-free manner, so the complexity of the training signal scales naturally with the data geometry. A subsequent force-matching step then amortizes these piecewise particle trajectories into a globally smooth, time-dependent velocity field that captures the entire cell-state transition process, ensuring temporal consistency across sub-intervals. This two-step framework enables modeling of highly nonlinear and time-inhomogeneous biological dynamics, including multimodal splitting and merging processes, transient cellular states, and strong perturbation responses such as drug treatment.

To further position PROFET within the rapidly growing landscape of trajectory inference methods, we constructed a structured comparison of representative approaches across mathematical framework and training methodology. This comparison highlights that PROFET is uniquely grounded in a variational formulation of Wasserstein gradient flow. Although PRESCIENT also assumes gradient-flow dynamics, it does not directly optimize an energy functional, instead learning a time-independent potential through Sinkhorn-based distribution matching. In contrast, most other dynamics-based approaches rely on OT either as a primary training objective or, in the case of MMFM, for sample coupling during spline-based trajectory construction. Methodologically, PROFET combines greedy (backpropagation-free) optimization with simulation-free force matching, avoiding both the backpropagation through unrolled dynamics required by adjoint-based methods and DTO-based methods. These distinctions contribute to PROFET's scalable and flexible modeling of complex biological dynamics.

Consistent with these theoretical distinctions, we comprehensively benchmarked PROFET against widely used and recently developed trajectory inference methods for scRNA-seq data and observed strong performance across diverse datasets. Our benchmarking and validation are conducted primarily on real datasets, spanning a range of biological systems (*in vitro* and *in vivo*; mouse, axolotl and human), including EMT, stem cell differentiation, regeneration, and hematopoiesis. These datasets exhibit diverse and complex dynamical patterns, including nonlinear trajectories (mESC), merging dynamics (EMT), combined splitting

and merging behaviors (axolotl), and lineage divergence (LARRY), which together reduce bias toward any specific modeling assumption and allow us to assess the general applicability of PROFET. To quantify performance, we employ multiple complementary evaluation metrics, including W2 distance, Sinkhorn divergence, and MMD, to assess distributional accuracy. Across the benchmark datasets, PROFET consistently achieved strong performance, with 2.6–12.5× lower W2 error than competing methods. In addition, we incorporate biologically motivated analyses—such as the ability to capture multimodal structure (e.g., emergence and disappearance of cell populations)—to provide a more comprehensive evaluation of model performance.

Given PROFET's strength in analyzing systems with phenotypic heterogeneity, we applied it to study variability in response to palbociclib, a CDK4/6 inhibitor used in HR+ breast cancer. In an *in vitro* breast cancer cell line treated with escalating doses of palbociclib, we observed a relatively uniform response characterized by downregulation of cell cycle genes and eventual drug resistance. In contrast, patient samples treated with combined palbociclib and endocrine therapy exhibited greater heterogeneity; some cells showed strong downregulation of both cell cycle and ER signaling genes, while others remained largely unresponsive. Using PROFET to reconstruct single-cell trajectories, we identified a subset of pre-treatment cells that later underwent pronounced phenotypic shifts. Differential expression analysis revealed surface markers—including *UNC5B*, *TLR3*, *PCDH19*, *PROCR*, *SEMA6B*, and *SLITRK6*—that were enriched in this high-shift group, suggesting their potential as early predictors of treatment-associated transitions. Both the *in vitro* model and two patient samples (PA3 and 887) showed signs of treatment resistance or disease progression, implicating these markers in the emergence of resistance. These findings point to the utility of these surface proteins not only for early response prediction in HR+ breast cancer but also as promising targets for ADCs, particularly *SLITRK6*, which is currently under clinical evaluation in ADC-based therapies.⁴³

Despite outperforming existing trajectory analysis tools and demonstrating strong utility in biological applications, our approach has several limitations. First, although we have performed a comparison with a broad set of widely used methods and evaluated prediction accuracy using multiple complementary distributional metrics (e.g., W2, Sinkhorn divergence, and MMD), no single quantitative metric can fully capture biological relevance. To address this limitation, we complemented these unbiased distribution-based evaluations with biologically grounded downstream analyses, such as cell-type/subpopulation prediction, to assess whether accurate distributional reconstruction also translates into meaningful biological interpretation. Second, PROFET is specifically designed for time-resolved scRNA-seq data and requires multiple observed time points. It is not intended for single-snapshot datasets, which are typically analyzed using pseudotime inference or stationary OT methods. These approaches rely on additional assumptions (e.g., predefined root states) and aim to infer latent progression rather than reconstruct dynamics in physical time, thus addressing a distinct problem. Third, PROFET is designed for interpolation within observed time intervals and does not perform extrapolation beyond the provided time range. Predicting future states

would require additional assumptions about the underlying dynamics and remains an important direction for future work.

In summary, PROFET offers a biologically grounded framework for reconstructing continuous, time-inhomogeneous, and nonlinear single-cell trajectories from discrete time-series scRNA-seq data. This characteristic makes it particularly valuable for analyzing systems with phenotypic heterogeneity, especially in the context of cell-state changes and treatment response. By enabling high-resolution inference of gene expression dynamics and fate transitions, PROFET offers a powerful and robust approach for uncovering hidden heterogeneity, tracing phenotypic plasticity, and informing time-sensitive interventions in both developmental and disease contexts.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Franziska Michor (michor@jimmy.harvard.edu).

Materials availability

This study did not generate new unique reagents or materials.

Data and code availability

- This study analyzes existing, publicly available scRNA-seq datasets. Accession numbers for all datasets are provided in the [key resources table](#).
- All original code used to implement PROFET, perform downstream analyses, benchmark competing methods, and generate the figures reported in this study has been deposited in a public repository and is publicly available as of the date of publication. The archived version of the code is available through Zenodo, and the corresponding GitHub repository (<https://github.com/HyeminGu/PROFET>) and Zenodo DOI are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this study is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

Y.-C.C. and F.M. thank the Center for Cancer Evolution at Dana-Farber Cancer Institute, Jacob Geisberg, Simona Cristea, and members of the Michor lab for their support and helpful discussions. Y.-C.C. further acknowledges insightful conversations with Helen Byrne (University of Oxford), Tom Chou (UCLA), and Hong Qian (University of Washington). M.A.K. and H.G. are partially funded by AFOSR grant FA9550-21-1-0354 and NSF grant DMS-2307115. W.W. is grateful for the support from the National Institute of General Medical Sciences (grant T32GM144273). The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. D.L.A. acknowledges support from the DF/HCC Breast Specialized Program in Research Excellence (SPORE) grant 1P50CA168504, Cancer Couch Foundation, Hope Scarves, the Jamieson Family Fund for Early Career Breast Cancer Researchers, and the Saverin Family Fund. H.L. acknowledges funding from the Center for Theoretical Biological Physics through NSF PHY-2019745 and NSF DMS-245957. C.G. was supported by fellowships from the Associazione Italiana Ricerca sul Cancro (AIRC) and the American-Italian Cancer Foundation (AICF) during her postdoctoral training in R.J.'s lab. Y.Z. is supported by the National Natural Science Foundation of China (82372750), the CAMS Innovation Fund for Medical Sciences (2023-I2M-3-003 and 2022-I2M-1-008), and the National Key R&D Program of China (2023YFC2507000).

AUTHOR CONTRIBUTIONS

Y.-C.C. and H.G. contributed equally to this work. Y.-C.C., H.G., T.O.M., S.T., Y.W., Y.P., H.L., M.A.K., and F.M. conceived and designed the study. Y.-C.C.,

H.G., T.O.M., S.T., Y.P., M.A.K., and F.M. developed the methodology. Y.-C.C., H.G., and M.B. developed the computational software. Y.-C.C., H.G., C.G., Y.W., Y.Z., H.L., R.J., M.A.K., and F.M. validated the results. Y.-C.C., H.G., W.W., S.T., M.B., and Y.W. performed the computational and statistical analyses. C.G., D.R., D.L.A., and R.J. generated experimental data and provided biological resources. Y.-C.C., H.G., S.T., C.G., D.R., D.L.A., Y.W., H.L., R.J., and F.M. curated and interpreted the data. Y.-C.C., H.G., W.W., S.T., M.B., M.A.K., and F.M. prepared the figures and visualizations. Y.-C.C., H.G., R.J., M.A.K., and F.M. wrote the original manuscript draft. All authors reviewed, edited, and approved the manuscript. T.O.M., H.L., R.J., M.A.K., and F.M. supervised the research. Y.-C.C., H.G., R.J., M.A.K., and F.M. coordinated the project. R.J., M.A.K., and F.M. acquired funding for the study.

DECLARATION OF INTERESTS

F.M. is a co-founder of and holds equity in Harbinger Health, holds equity in Zephyr AI, and serves as a consultant for both companies. She also serves on the board of directors of Recursion Pharmaceuticals. She declares that none of these relationships are directly or indirectly related to the content of this manuscript. M.A.K. is a co-founder of and holds equity in Archimedean Inc. He declares that this relationship is not directly or indirectly related to the content of this manuscript. R.J. has received research funding from Lilly and Pfizer and has served on advisory boards or received consulting fees from Lilly, AstraZeneca, Novartis, Pfizer, GE Health, and Exscientia.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Synthetic EMT Data
 - Preprocessing of mESC data
 - Preprocessing of EMT data induced by TGF- β
 - Preprocessing of Axolotl Regeneration Data
 - Preprocessing of breast cancer cell line data
 - Preprocessing of breast cancer patient data
- **METHOD DETAILS**
 - Gradient flow simulation with Generative Particles Algorithm (GPA)
 - The role of Lipschitz regularization in empirical gradient flow modeling
 - Handling high-dimensionality via latent-space modeling
 - Stopping time for GPA
 - Extending GPA to capture nonlinear temporal dynamics
 - Force-matching for learning a unified velocity field
 - Overview of the two-step framework
 - Model architecture and training algorithm
 - Downstream Analysis Using PROFET
 - Setup of Benchmarked Methods
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Computing distributional distances between predicted and test-time distributions
 - Permutation testing of predicted versus observed distributions
 - Displacement-based quantification of phenotypic shift
 - KDE-based stratification of cellular subpopulations
 - Assessing EMT Score

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cels.2026.101710>.

Received: December 17, 2025

Revised: June 2, 2026

Accepted: July 30, 2026

REFERENCES

- Purvis, J.E., and Lahav, G. (2013). Encoding and Decoding Cellular Information through Signaling Dynamics. *Cell* 152, 945–956. <https://doi.org/10.1016/j.cell.2013.02.005>.
- Skylaki, S., Hilsenbeck, O., and Schroeder, T. (2016). Challenges in long-term imaging and quantification of single-cell dynamics. *Nat. Biotechnol.* 34, 1137–1144. <https://doi.org/10.1038/nbt.3713>.
- Mazutis, L., Gilbert, J., Ung, W.L., Weitz, D.A., Griffiths, A.D., and Heyman, J.A. (2013). Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* 8, 870–891. <https://doi.org/10.1038/nprot.2013.046>.
- Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. <https://doi.org/10.1038/nature21350>.
- Wagner, D.E., and Klein, A.M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* 21, 410–427. <https://doi.org/10.1038/s41576-020-0223-2>.
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* 176, 928–943.e22. <https://doi.org/10.1016/j.cell.2019.01.006>.
- Tong, A., Huang, J., Wolf, G., Van Dijk, D., and Krishnaswamy, S. (2020). TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics. In *Proceedings of the 37th International Conference on Machine Learning (PMLR)*, pp. 9526–9536. <https://proceedings.mlr.press/v119/tong20a.html>.
- Huguet, G., Magruder, D.S., Tong, A., Fasina, O., Kuchroo, M., Wolf, G., and Krishnaswamy, S. (2022). Manifold Interpolating Optimal-Transport Flows for Trajectory Inference. *Advances in Neural Information Processing Systems* 35, 29705–29718. <https://doi.org/10.52202/068431-2154>.
- Sha, Y., Qiu, Y., Zhou, P., and Nie, Q. (2024). Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nat. Mach. Intell.* 6, 25–39. <https://doi.org/10.1038/s42256-023-00763-w>.
- Wang, D., Jiang, Y., Zhang, Z., Gu, X., Zhou, P., and Sun, J. (2025). Joint Velocity-Growth Flow Matching for Single-Cell Dynamics Modeling. In *Proceedings of the Thirty-ninth Annual Conference on Neural Information Processing Systems* <https://openreview.net/forum?id=aXAkNlbnGa>.
- Jiang, Q., and Wan, L. (2024). A physics-informed neural SDE network for learning cellular dynamics from time-series scRNA-seq data. *Bioinformatics* 40, ii120–ii127. <https://doi.org/10.1093/bioinformatics/btae400>.
- Zhang, Z., Li, T., and Zhou, P. (2025). Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. In *Proceedings of the Thirteenth International Conference on Learning Representations* <https://openreview.net/forum?id=gQlx3Mtru>.
- Bunne, C., Stark, S.G., Gut, G., del Castillo, J.S., Levesque, M., Lehmann, K.-V., Pelkmans, L., Krause, A., and Rätsch, G. (2023). Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* 20, 1759–1768. <https://doi.org/10.1038/s41592-023-01969-x>.
- Rohbeck, M., De Brouwer, E., Bunne, C., Huetter, J.-C., Biton, A., Chen, K.Y., Regev, A., and Lopez, R. (2025). Modeling Complex System Dynamics with Flow Matching Across Time and Conditions. In *Proceedings of the Thirteenth International Conference on Learning Representations* <https://openreview.net/forum?id=hwnObmOTrV>.
- Chen, Z., Gu, H., Katsoulakis, M.A., Rey-Bellet, L., and Zhu, W. (2025). Robust generative learning with Lipschitz-regularized α -divergences allows minimal assumptions on target distributions. *Inf. Inference* 14, iaaf028. <https://doi.org/10.1093/imaia/iaaf028>.
- Yeo, G.H.T., Saksena, S.D., and Gifford, D.K. (2021). Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nat. Commun.* 12, 3222. <https://doi.org/10.1038/s41467-021-23518-w>.
- Guarducci, C., Abravanel, D., Russo, D., Prasad, K., Fu, J., Nagy, Z., Rao, V., Nardone, A., Kuang, Y., Feit, A., et al. (2026). ESR1 mutations and CDK4/6 inhibitor choice shape clonal selection and adaptive cell states during acquired resistance. Published online June 27, 2026. *Genome Med.* <https://doi.org/10.1186/s13073-026-01690-2>.
- Luo, L., Yang, P., Mastoraki, S., Rao, X., Wang, Y., Kettner, N.M., Raghavendra, A.S., Tripathy, D., Damodaran, S., Hunt, K.K., et al. (2025). Single-cell RNA sequencing identifies molecular biomarkers predicting late progression to CDK4/6 inhibition in patients with HR+/HER2-metastatic breast cancer. *Mol. Cancer* 24, 48. <https://doi.org/10.1186/s12943-025-02226-9>.
- Klughammer, J., Abravanel, D.L., Segerstolpe, Å., Blosser, T.R., Goltsev, Y., Cui, Y., Goodwin, D.R., Sinha, A., Ashenberg, O., Slyper, M., et al. (2024). A multi-modal single-cell and spatial expression map of metastatic breast cancer biopsies across clinicopathological features. *Nat. Med.* 30, 3236–3249. <https://doi.org/10.1038/s41591-024-03215-z>.
- Finn, R.S., Martin, M., Rugo, H.S., Jones, S., Im, S.-A., Gelmon, K., Harbeck, N., Lipatov, O.N., Walshe, J.M., Moulder, S., et al. (2016). Palbociclib and Letrozole in Advanced Breast Cancer. *N. Engl. J. Med.* 375, 1925–1936. <https://doi.org/10.1056/NEJMoa1607303>.
- Álvarez-Fernández, M., and Malumbres, M. (2020). Mechanisms of Sensitivity and Resistance to CDK4/6 Inhibition. *Cancer Cell* 37, 514–529. <https://doi.org/10.1016/j.ccell.2020.03.010>.
- Dumontet, C., Reichert, J.M., Senter, P.D., Lambert, J.M., and Beck, A. (2023). Antibody–drug conjugates come of age in oncology. *Nat. Rev. Drug Discov.* 22, 641–661. <https://doi.org/10.1038/s41573-023-00709-2>.
- Birrell, J., Dupuis, P., Katsoulakis, M.A., Pantazis, Y., and Rey-Bellet, L. (2022). (f, Γ)-Divergences: Interpolating between f-Divergences and Integral Probability Metrics. *J. Mach. Learn. Res.* 23, 1–70. <http://jmlr.org/papers/v23/21-0100.html>.
- Gu, H., Birmppa, P., Pantazis, Y., Rey-Bellet, L., and Katsoulakis, M.A. (2024). Lipschitz-Regularized Gradient Flows and Generative Particle Algorithms for High-Dimensional Scarce Data. *SIAM J. Math. Data Sci.* 6, 1205–1235. <https://doi.org/10.1137/23M1587841>.
- Noid, W.G. (2023). Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models. *J. Phys. Chem. B* 127, 4174–4207. <https://doi.org/10.1021/acs.jpcc.2c08731>.
- Kundu, P.K., Cohen, I.M., Dowling, D.R., and Capece, J. (2024). *Fluid Mechanics* (Elsevier).
- Tripathi, S., Kessler, D.A., and Levine, H. (2023). Minimal frustration underlies the usefulness of incomplete regulatory network models in biology. *Proc. Natl. Acad. Sci. USA* 120, e2216109120. <https://doi.org/10.1073/pnas.2216109120>.
- Hayashi, T., Ozaki, H., Sasagawa, Y., Umeda, M., Danno, H., and Nikaïdo, I. (2018). Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* 9, 619. <https://doi.org/10.1038/s41467-018-02866-0>.
- Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T., and Nikaïdo, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* 33, 2314–2321. <https://doi.org/10.1093/bioinformatics/btx194>.
- Deshmukh, A.P., Vasaikar, S.V., Tomczak, K., Tripathi, S., den Hollander, P., Arslan, E., Chakraborty, P., Soundararajan, R., Jolly, M.K., Rai, K., et al. (2021). Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. *Proc. Natl. Acad. Sci. USA* 118, e2102050118. <https://doi.org/10.1073/pnas.2102050118>.
- Li, H., Wei, X., Zhou, L., Zhang, W., Wang, C., Guo, Y., Li, D., Chen, J., Liu, T., Zhang, Y., et al. (2021). Dynamic cell transition and immune response landscapes of axolotl limb regeneration revealed by single-cell analysis. *Protein Cell* 12, 57–66. <https://doi.org/10.1007/s13238-020-00763-1>.
- Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* 367, eaaw3381. <https://doi.org/10.1126/science.aaw3381>.

33. Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624. <https://doi.org/10.1038/ncb2065>.
34. Lin, Y., Li, X.-Y., Willis, A.L., Liu, C., Chen, G., and Weiss, S.J. (2014). Snail1-dependent control of embryonic stem cell pluripotency and lineage commitment. *Nat. Commun.* 5, 3070. <https://doi.org/10.1038/ncomms4070>.
35. Zhang, C., Ye, X., Zhang, H., Ding, M., and Deng, H. (2007). GATA factors induce mouse embryonic stem cell differentiation toward extraembryonic endoderm. *Stem Cells Dev.* 16, 605–613. <https://doi.org/10.1089/scd.2006.0077>.
36. Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. *Development* 144, 1221–1234. <https://doi.org/10.1242/dev.142711>.
37. Chae, Y.K., Chang, S., Ko, T., Anker, J., Agte, S., Iams, W., Choi, W.M., Lee, K., and Cruz, M. (2018). Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci. Rep.* 8, 2918. <https://doi.org/10.1038/s41598-018-21061-1>.
38. Wang, X., Jiang, C., and Li, Q. (2024). Serinc2 Drives the Progression of Cervical Cancer Through Regulating Myc Pathway. *Cancer Med.* 13, e70296. <https://doi.org/10.1002/cam4.70296>.
39. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. <https://doi.org/10.1093/bioinformatics/btr260>.
40. Cheng, Y.-C., Stein, S., Nardone, A., Liu, W., Ma, W., Cohen, G., Guarducci, C., McDonald, T.O., Jeselsohn, R., and Michor, F. (2023). Mathematical Modeling Identifies Optimum Palbociclib-fulvestrant Dose Administration Schedules for the Treatment of Patients with Estrogen Receptor-positive Breast Cancer. *Cancer Res. Commun.* 3, 2331–2344. <https://doi.org/10.1158/2767-9764.CRC-23-0257>.
41. Asciolla, J.J., Wu, X., Adamopoulos, C., Gavathiotis, E., and Poulikakos, P.I. (2025). Resistance mechanisms and therapeutic strategies of CDK4 and CDK6 kinase targeting in cancer. *Nat. Cancer* 6, 24–40. <https://doi.org/10.1038/s43018-024-00893-z>.
42. Migliaccio, I., Guarducci, C., and Malorni, L. (2025). CDK4/6 Inhibitor Resistance in ER+ Breast Cancer. In *A Guide to Breast Cancer Research: From Cellular Heterogeneity and Molecular Mechanisms to Therapy*, T. Sørlie and R.B. Clarke, eds. (Springer Nature), pp. 475–493. https://doi.org/10.1007/978-3-031-70875-6_23.
43. Morrison, K., Challita-Eid, P.M., Raitano, A., An, Z., Yang, P., Abad, J.D., Liu, W., Lortie, D.R., Snyder, J.T., Capo, L., et al. (2016). Development of ASG-15ME, a Novel Antibody-Drug Conjugate Targeting SLITRK6, a New Urothelial Cancer Biomarker. *Mol. Cancer Ther.* 15, 1301–1310. <https://doi.org/10.1158/1535-7163.MCT-15-0570>.
44. Osborne, C.K., Wakeling, A., and Nicholson, R.I. (2004). Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action. *Br. J. Cancer* 90, S2–S6. <https://doi.org/10.1038/sj.bjc.6601629>.
45. Patel, R., Klein, P., Tiersten, A., and Sparano, J.A. (2023). An emerging generation of endocrine therapies in breast cancer: a clinical perspective. *NPJ Breast Cancer* 9, 20. <https://doi.org/10.1038/s41523-023-00523-4>.
46. Wu, S., Guo, X., Zhou, J., Zhu, X., Chen, H., Zhang, K., Lu, Y., and Chen, Y. (2020). High expression of UNC5B enhances tumor proliferation, increases metastasis, and worsens prognosis in breast cancer. *Aging (Albany, NY)* 12, 17079–17098. <https://doi.org/10.18632/aging.103639>.
47. Muresan, X.M., Bouchal, J., Culig, Z., and Souček, K. (2020). Toll-Like Receptor 3 in Solid Cancer and Therapy Resistance. *Cancers (Basel)* 12, 3227. <https://doi.org/10.3390/cancers12113227>.
48. Wang, D., Liu, C., Wang, J., Jia, Y., Hu, X., Jiang, H., Shao, Z.M., and Zeng, Y.A. (2018). Protein C receptor stimulates multiple signaling pathways in breast cancer cells. *J. Biol. Chem.* 293, 1413–1424. <https://doi.org/10.1074/jbc.M117.814046>.
49. Du, J., and Wu, K. (2023). PCDHA1 High Expression is Associated With Poor Prognosis and Correlated With Immune Cell Infiltration in Breast Cancer. *Clin. Breast Cancer* 23, 397–407. <https://doi.org/10.1016/j.clbc.2023.02.001>.
50. Liu, X., Liu, Y., Liu, Z., Zhang, Y., Ma, Y., Bai, J., Yao, H., Wang, Y., Zhao, X., Li, R., et al. (2021). Identification of SLITRK6 as a Novel Biomarker in hepatocellular carcinoma by comprehensive bioinformatic analysis. *Biochem. Biophys. Rep.* 28, 101157. <https://doi.org/10.1016/j.bbrep.2021.101157>.
51. Li, T., Yan, Z., Wang, W., Zhang, R., Gan, W., Lv, S., Zeng, Z., Hou, Y., and Yang, M. (2021). SEMA6B Overexpression Predicts Poor Prognosis and Correlates With the Tumor Immunosuppressive Microenvironment in Colorectal Cancer. *Front. Mol. Biosci.* 8, 687319. <https://doi.org/10.3389/fmolb.2021.687319>.
52. Hughes, N.W., Qu, Y., Zhang, J., Tang, W., Pierce, J., Wang, C., Agrawal, A., Morri, M., Neff, N., Winslow, M.M., et al. (2022). Machine-learning-optimized Cas12a barcoding enables the recovery of single-cell lineages and transcriptional profiles. *Mol. Cell* 82, 3103–3118.e8. <https://doi.org/10.1016/j.molcel.2022.06.001>.
53. Huang, B., Lu, M., Jia, D., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2017). Interrogating the topological robustness of gene regulatory circuits by randomization. *PLOS Comput. Biol.* 13, e1005456. <https://doi.org/10.1371/journal.pcbi.1005456>.
54. LeVeque, R.J. (2007). Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems (SIAM). <https://doi.org/10.1137/1.9780898717839>.
55. Leinonen, R., Sugawara, H., and Shumway, M.; International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. <https://doi.org/10.1093/nar/gkq1019>.
56. Byers, L.A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., Shen, L., Fan, Y., Giri, U., Tumula, P.K., et al. (2013). An Epithelial-Mesenchymal Transition Gene Signature Predicts Resistance to EGFR and PI3K Inhibitors and Identifies Axl as a Therapeutic Target for Overcoming EGFR Inhibitor Resistance. *Clin. Cancer Res.* 19, 279–290. <https://doi.org/10.1158/1078-0432.CCR-12-1558>.
57. Chakraborty, P., George, J.T., Tripathi, S., Levine, H., and Jolly, M.K. (2020). Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. *Front. Bioeng. Biotechnol.* 8, 220. <https://doi.org/10.3389/fbioe.2020.00220>.
58. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049. <https://doi.org/10.1038/ncomms14049>.
59. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* 42, 293–304. <https://doi.org/10.1038/s41587-023-01767-y>.
60. Tickle, T., Tirosh, I., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project (Klarman Cell Observatory, Broad Institute of MIT and Harvard). <https://doi.org/10.18129/B9.bioc.infercnv>.
61. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
62. Cimino, A., Halushka, M., Illei, P., Wu, X., Sukumar, S., and Argani, P. (2010). Epithelial cell adhesion molecule (EpCAM) is overexpressed in breast cancer metastases. *Breast Cancer Res. Treat.* 123, 701–708. <https://doi.org/10.1007/s10549-009-0671-z>.
63. Genevay, A., Peyré, G., and Cuturi, M. (2018). Learning Generative Models with Sinkhorn Divergences. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (PMLR)*, pp. 1608–1617. <https://proceedings.mlr.press/v84/genevay18a.html>.

64. Ercolessi, F., and Adams, J.B. (1994). Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* 26, 583–588. <https://doi.org/10.1209/0295-5075/26/8/005>.
65. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., and Le, M. (2023). Flow Matching for Generative Modeling. In *Proceedings of the Eleventh International Conference on Learning Representations* <https://openreview.net/forum?id=PqvMRDCJT9t>.
66. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., et al. (2013). API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.
67. Cai, Y., and Lim, L.-H. (2022). Distances Between Probability Distributions of Different Dimensions. *IEEE Trans. Inf. Theory* 68, 4020–4031. <https://doi.org/10.1109/TIT.2022.3148923>.
68. Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* 28, 129–137. <https://doi.org/10.1109/TIT.1982.1056489>.
69. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>.
70. McDermaid, A., Monier, B., Zhao, J., Liu, B., and Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief. Bioinform.* 20, 2044–2054. <https://doi.org/10.1093/bib/bby067>.
71. Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S., Trounev, A., and Peyré, G. (2019). Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics (PMLR)*, pp. 2681–2690. <https://proceedings.mlr.press/v89/feydy19a.html>.
72. Cuturi, M. (2013). Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, 26, pp. 2292–2300. https://proceedings.neurips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.
73. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., and Smola, A. (2012). A Kernel Two-Sample Test. *J. Mach. Learn. Res.* 13, 723–773. <http://jmlr.org/papers/v13/gretton12a.html>.
74. Flamary, R., Courty, N., Gramfort, A., Alaya, M.Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. (2021). POT: Python Optimal Transport. *J. Mach. Learn. Res.* 22, 1–8. <http://jmlr.org/papers/v22/20-451.html>.
75. Silverman, B.W. (2018). *Density Estimation for Statistics and Data Analysis*, First Edition (Routledge). <https://doi.org/10.1201/9781315140919>.
76. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Mouse embryonic stem cell (mESC) scRNA-seq dataset	Hayashi et al. ²⁸	GEO: GSE98664
MCF10A EMT scRNA-seq dataset	Deshmukh et al. ³⁰	BioProject ID: PRJNA698642
Axolotl limb regeneration scRNA-seq dataset	Li et al. ³¹	SRA: PRJNA589484
LARRY hematopoiesis scRNA-seq dataset	Weinreb et al. ³²	GEO: GSE140802
Breast cancer patient dataset (PA3)	Luo et al. ¹⁸	GEO: GSE262288
Breast cancer patient datasets (862 and 887)	Klughammer et al. ¹⁹	dbGap: phs002371
MCF7 palbociclib resistance scRNA-seq dataset	Guarducci et al. ¹⁷	GEO: GSE326012
Processed datasets used in this study	This study	Zenodo DOI: 10.5281/zenodo.21014564
Software and algorithms		
PROFET	This study	Zenodo DOI: 10.5281/zenodo.21015867
PROFET GitHub repository	This study	https://github.com/HyeminGu/PROFET
PyTorch	PyTorch developers	Version 2.4.0; https://github.com/pytorch/pytorch
TensorFlow	TensorFlow developers	Version 2.16.1; https://github.com/tensorflow/tensorflow
GeomLoss	GeomLoss developers	Version 0.2.6; https://github.com/jeanfeydy/geomloss
scikit-learn	scikit-learn developers	Version 1.2.1; https://scikit-learn.org
Matplotlib	Matplotlib developers	Version 3.10.3; https://matplotlib.org/
NumPy	Numpy developers	Version 1.23.3; https://numpy.org/
SciPy	SciPy developers	Version 1.16.0; https://scipy.org/
YAML	PyYAML developers	Version 6.0.2; https://pyyaml.org/project/PyYAML/
notebook	Project Jupyter	Version 6.4.12; https://jupyter.org/
pandas	pandas developers	Version 1.5.0; https://pandas.pydata.org/
Other		
Lead Contact	Franziska Michor	michor@jimmy.harvard.edu
bioRxiv preprint	This study	DOI: 10.1101/2025.06.27.662030

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Synthetic EMT Data

We generated a synthetic dataset mimicking an experimental setup wherein a population of epithelial cells is treated with an inducer of EMT such as TGF- β for a fixed duration of time. For this purpose, we started with a previously characterized 26-node gene regulatory network that is known to establish gene expression patterns characteristic of epithelial and mesenchymal cells and to govern the transition between these phenotypic states.⁵³ The gene regulatory network is described by a connection matrix J whose elements represent the regulatory interactions between genes:

$$J_{ij} = \begin{cases} +1 & \text{gene } j \text{ activates gene } i \\ -1 & \text{if gene } j \text{ inhibits gene } i \\ 0 & \text{gene } j \text{ does not regulate gene } i \end{cases} \quad (\text{Equation 1})$$

Given the connection matrix J , we can write a system of stochastic differential equations to describe the dynamics of the EMT gene expression vector $x \equiv \{x_i\}$, $1 \leq i \leq 26$:

$$\frac{dx_i}{dt} = g_i \left(\prod_{j, J_{ij} \neq 0} H^S(x_j, \lambda_{ij}, \theta_{ij}, n_{ij}) \right) - k_i x_i + D x_i \eta_i(t) \quad (\text{Equation 2})$$

Here, g_i and k_i are the production and degradation rates for gene i , respectively. The term $D x_i \eta_i(t)$ corresponds to multiplicative noise where $D = 1$ is the noise magnitude and $\eta_i(t)$ is a delta correlated noise term with zero mean and unit variance. H^S is the shifted Hill function:

$$H^S(x_j, \lambda_{ij}, \theta_{ij}, n_{ij}) = \lambda_{ij} + (1 - \lambda_{ij}) \left(\frac{1}{1 + \left(\frac{x_j}{\theta_{ij}} \right)^{n_{ij}}} \right) \quad (\text{Equation 3})$$

The parameters λ_{ij} , θ_{ij} , and n_{ij} together characterize the nature of the regulation of gene i by gene j :

1. λ_{ij} is the maximum fold change in the production rate of gene i that gene j can cause. Thus, $\lambda_{ij} > 1$ if $J_{ij} = +1$ and $0 < \lambda_{ij} < 1$ if $J_{ij} = -1$.
2. θ_{ij} is the threshold parameter: the regulatory edge from j to i is active if $x_j > \theta_{ij}$. The regulatory interaction is very weak otherwise.
3. n_{ij} is the Hill coefficient: H^S varies more steeply with x_j for higher values of n_{ij} .

Given a parameter set $(\{g_i\}, \{k_i\}, \{\lambda_{ij}\}, \{\theta_{ij}\}, \{n_{ij}\})$ and the initial condition x_0 , Equation 2 can be solved using the Euler-Maruyama method.⁵⁴

Preprocessing of mESC data

The dataset analyzed consists of single-cell RNA-sequencing data from mouse embryonic stem cells collected at five time points—0, 12, 24, 48, and 72 hours—comprising 92, 91, 93, 87, and 93 cells, respectively.^{28,29} To preprocess the data, we applied a log-transformation to the raw read counts, replacing each value x with $\log(1+x)$. Following the same gene selection strategy as in the original study,²⁹ we calculated the variance of each gene across all cells and selected the top 100 genes with the highest variance for downstream temporal analysis.

Preprocessing of EMT data induced by TGF- β

The single-cell RNA-sequencing dataset analyzed in this study was obtained from a previous publication on MCF10A cells undergoing TGF- β -induced epithelial-to-mesenchymal transition (EMT).³⁰ We utilized the processed gene expression matrix provided by the authors; raw sequencing reads are available through the NCBI Sequence Read Archive (BioProject accession PRJNA698642).⁵⁵ For dimensionality reduction, we selected a subset of 72 genes that overlap with a previously established list of 76 genes known to be strongly associated with the EMT gene expression signature.^{56,57} The dataset includes cells collected at three time points: day 0 ($n = 2,734$), day 2 ($n = 2,381$), and day 4 ($n = 1,147$).

Preprocessing of Axolotl Regeneration Data

The single-cell RNA-sequencing dataset analyzed in this study was obtained from a previously published study on Axolotl limb regeneration.³¹ We used the processed gene-expression matrix provided by the authors. For cell-type annotations, we adopted the exact annotation files from the original publication and confirmed their consistency with the classifications reported in the study.

Preprocessing of breast cancer cell line data

Samples were aligned with 10x Genomics Cell Ranger (v3.1.0) pipeline⁵⁸ with the cellranger-count command with parameter 'expect-cells=8000' to the reference genome refdata-cellranger-GRCh38-3.0.0 provided by 10x Genomics. Filtered feature barcode matrices were imported into R using the Read10X function from Seurat (v5.2.1)⁵⁹ and empty droplets and dead cells were filtered out by exploratory data analysis of QC metrics and manually choosing thresholds to keep cells with percent mitochondrial reads $< = 25\%$, minimum number of genes with nonzero reads per cell $> = 100$, and minimum total number reads per cell $> = 10000$. Normalized expression values were calculated with Seurat's NormalizeData function with default 'LogNormalize' options. The dataset comprises cells collected at two time points: pre-treatment (day 0, $n = 6,249$) and post-treatment (week 16, $n = 7,911$).

Preprocessing of breast cancer patient data

For the Klughammer dataset, we used the original cell annotations to select malignant epithelial cells for downstream analysis.¹⁹ Specifically, for patient 862, the dataset includes cells collected at two time points: pre-treatment (day 0, $n = 8,953$) and post-treatment (day 200, $n = 8,307$). For patient 887, cells were also collected at two time points: pre-treatment (day 0, $n = 5,536$) and post-treatment (day 220, $n = 4,638$).

For the Luo dataset, we focused on patient PA3 and followed the published inferCNV pipeline⁶⁰ to identify malignant cells. Prior to running inferCNV, we performed unsupervised clustering using Seurat,⁶¹ which identified 15 distinct cell clusters (Figure S21A). Based on the expression of known immune cell marker genes (Table S10), a subset of clusters—specifically clusters 2, 3, 4, 5, 7, 9, 11, 13, and 14 (Figure S21A)—was identified as immune cell populations and designated as reference normal cells for the inferCNV analysis. We then applied inferCNV to infer copy number variations (CNVs) across the remaining cells. The analysis revealed consistently abnormal CNV burden across all clusters except one (Figure S21B).

Further inspection of this exception group (cluster 1) revealed elevated expression of multiple collagen family genes (Figure S21C), suggesting these cells were fibroblasts. As a result, this fibroblast cluster was excluded, and the remaining clusters were retained as malignant epithelial cells. To validate this classification, we examined the expression of EPCAM, a well-established marker of malignant epithelial cells,⁶² and confirmed its overexpression in the inferred malignant population (Figure S21D). This confirmed both the malignant nature and epithelial identity of the selected cells. The dataset includes cells collected at two time points: pre-treatment (day 0, $n = 2,157$) and post-treatment (month 11, $n = 2,535$).

METHOD DETAILS

Gradient flow simulation with Generative Particles Algorithm (GPA)

We simulate data-driven trajectories using the Generative Particles Algorithm (GPA),²⁴ a deterministic method that iteratively transports particles from a source distribution toward a target. In our setting, each particle represents a cell's gene expression profile in reduced-dimensional space. GPA uses Lipschitz-regularized gradient flows to interpolate between arbitrary distributions—including empirical, discrete, or irregular ones—in a fully agnostic and mathematically justified manner. This provable flexibility¹⁵ is essential in single-cell analysis, where the underlying data structure is unknown.

At each time step, GPA estimates a potential function based on the current particle configuration and computes an update direction via its gradient. This approach leads to ODE-based dynamics:

$$\frac{dX_t}{dt} = v_t(X_t), \text{ with } v_t(x) = -\nabla \phi_t^{L,*}(x), X_0 \sim \rho_0, \quad (\text{Equation 4})$$

where $\phi_t^{L,*}$ is the time-dependent potential learned from the data. The underlying Wasserstein gradient flow formulation and the variational structure of the Lipschitz-regularized KL divergence are provided in detail in Note S1.

To solve this ODE numerically, GPA uses a forward Euler scheme over n_T steps of size Δt , with a greedy optimization at each step:

$$Y_{n+1}^i = Y_n^i - \Delta t \nabla \phi_n^{L,*}(Y_n^i), \phi_n^{L,*} = \arg \max_{\phi \in \text{Lip}_L(\mathbb{R}^d)} \left\{ \frac{1}{N} \sum_i \phi(Y_n^i) - \frac{1}{M} \sum_j f^*(\phi(X_j^i)) \right\}, \quad (\text{Equation 5})$$

starting from samples $\{Y_0^i\}_{i=1}^N \sim \rho_0$. The inner optimization approximates the KL divergence between the current particle distribution and the target distribution using the variational representation of f -divergence, where $f(x) = x \log x$ and f^* denotes its Legendre transform.

At each iteration, GPA updates the particle positions and records the time-labeled velocities

$$(x, t, v) = (Y_n^i, n\Delta t, -\nabla \phi_n^{L,*}(Y_n^i)), i = 1, \dots, N; n = 0, \dots, n_T - 1. \quad (\text{Equation 6})$$

The evolution naturally slows as the particle distribution approaches the target, which is quantified by the decay of the expected kinetic energy $E[\|\nabla \phi_n^{L,*}(Y_n^i)\|^2] \rightarrow 0$ (Theorem 2, Lipschitz-regularized dissipation; Gu et al.²⁴).

The role of Lipschitz regularization in empirical gradient flow modeling

Lipschitz regularization on the KL divergence serves *three critical purposes* in modeling empirical gradient flows:

1. **Theoretical well-posedness:** Empirical distributions—being finite samples from unknown distributions on $\mathbb{R}^d$ —are generally not absolutely continuous with respect to each other. Consequently, the KL divergence between them may become ill-defined or divergent. The Lipschitz-regularized KL divergence resolves this issue by defining a well-posed variational objective even when absolute continuity fails. This approach ensures that the first variation and corresponding flow velocity remain well-defined. See Note S2 for a detailed discussion of this issue.
2. **Velocity control:** Lipschitz continuity ensures that the velocity field induced by the flow has a finite speed bound, specifically bounded by the Lipschitz constant L . In the limit as $L \rightarrow \infty$, the Lipschitz-regularized divergence recovers the unregularized KL divergence and its associated gradient flow, confirming that the regularization is a consistent approximation that introduces no fundamental distortion to the KL-based dynamics. In the unregularized case, the velocity grows without bound as the support gets unbounded (Note S9). For empirical distributions with compact support, a finite L suffices to approximate the KL-induced dynamics well.
3. **Numerical stability and stiffness control:** Bounding the velocity field by a finite Lipschitz constant L ensures that the transport speed is controlled, enabling stable time integration under the Courant–Friedrichs–Lewy (CFL) condition for Euler-type methods (2). However, larger L values allow the velocity to vary more rapidly across space, which can lead to stiff dynamics

that pose numerical instability, as demonstrated in Gu et al.,²⁴ Figures 2 and 3. This stiffness–stability trade-off motivates the use of moderate values of L that are large enough to capture meaningful dynamical behavior (e.g., convergence rate) but small enough to maintain stable, efficient simulations at practical step sizes. In effect, L acts as a dial that balances dynamical accuracy and numerical feasibility in GPA.

In our method, we set $L = 1$ to ensure numerical stability across diverse datasets, providing a universally robust baseline; a sensitivity analysis on L is provided in [Note S7](#). The step size Δt is then adjusted per dataset to accommodate variations in scale and complexity.

Handling high-dimensionality via latent-space modeling

Gene expression datasets are intrinsically high-dimensional, with each cell represented by expression levels across thousands of genes. To make learning tractable while preserving meaningful structure, we apply Principal Component Analysis (PCA) to project the data into a lower-dimensional latent space, where GPA is used to model the dynamics. For downstream analysis, we reconstruct the trajectories in the original space using the inverse PCA transform.

Across our 5 datasets — with ambient dimensions $D = 26, 100, 72, 117, 116, 115, 115$ corresponding to the number of genes in synthetic data, mESC data, EMT data, breast cancer cell line, patient PA3, patient 862, and patient 887, respectively — we reduce the dimension of each of them to $d = 2, 8, 32, \dots, D$ and observe that GPA consistently performs well up to moderate latent dimensions (e.g., $d \leq 32$), capturing temporal structure without overfitting or numerical instability. A sensitivity analysis on prediction errors across PCA dimensionalities is provided in [Note S7](#).

To justify latent modeling of high-dimensional biological applications, we provide a theoretical guarantee based on the Lipschitz-regularized divergence in [Note S10](#).

Stopping time for GPA

As the time horizon for gradient flows is infinite, numerical time-discretized algorithms such as forward Euler require a stopping criterion to determine when particle transport has sufficiently converged. The stopping time n_T determines the effective time horizon, which is then rescaled to the physical time scale between consecutive snapshots, thereby adjusting the resolution of the transport dynamics— more steps for complex dynamics such as branching or merging, fewer for simpler transitions. We use a threshold value of $D_f^{L|pL} < 0.05$ (with $f=KL$ and $L=1$). This threshold was calibrated on the synthetic dataset by computing the Sinkhorn divergence⁶³ between the generated and true distributions at held-out intermediate timepoints, and was found to provide optimal alignment for the synthetic dataset example. This threshold provides a practical and computationally efficient criterion for early stopping that balances convergence and overfitting to the endpoint distributions. A sensitivity analysis on stopping time is provided in [Note S7](#).

Extending GPA to capture nonlinear temporal dynamics

While GPA accurately simulates local transport between pairs of gene expression snapshots, many biological processes—such as drug response or lineage bifurcation—exhibit temporally nonlinear or branched dynamics that cannot be represented by a single global trajectory. To address this issue, we partition the total time horizon $[0, T]$ into subintervals and apply GPA independently within each, using the corresponding pair of snapshots.

This strategy yields a sequence of locally consistent gradient flows that adapt to complex temporal structures while preserving interpretability through KL-consistent dynamics. These local simulations serve as the foundation for constructing a unified, continuous-time model in the next stage of our framework.

Force-matching for learning a unified velocity field

To amortize the locally simulated gradient flows from multiple GPA subintervals into a global dynamic model, we adopt a force-matching approach.⁶⁴ Classically used to learn coarse-grained force fields from trajectory data in molecular simulations,²⁵ force-matching is here extended to the data-driven regime: instead of relying on predefined physical models, we use GPA-inferred particle trajectories—obtained separately on each subinterval—to supervise the learning of a continuous-time Eulerian velocity field. We require a globally defined velocity field over the entire time horizon to support downstream analyses—such as tracing fate bifurcations ([Figure 3E](#)), identifying genes with temporally localized divergence or convergence ([Figures 3F, 3G, 4H, and 4I](#)), and quantifying phenotypic shifts in patient-derived cells ([Figure 5C](#))—all of which rely on continuous integration from arbitrary time points.

We define a neural velocity field $v_\theta(x, s)$ over space and time and train it by minimizing the deviation between the learned velocity and the GPA-estimated transport velocity at randomly sampled time points across all particles:

$$\min_{\theta} E_{(x,s)} \left[\left| v_\theta(x, s) + \nabla \phi_s^{L,*}(x) \right|^2 \right]. \quad (\text{Equation 7})$$

Here, $\phi_s^{L,*}$ is the potential estimated from GPA on the subinterval that covers time s . To ensure temporal consistency across subintervals, the physical time variable $s \in [0, T]$ is defined by proportionally aligning each subinterval's local GPA simulation time $t \in [0, n_T^{(i)} \Delta t^{(i)}]$ to a shared physical time scale. All subintervals are reparameterized relative to the longest GPA simulation horizon $\max_i n_T^{(i)} \Delta t^{(i)}$, allowing the learned velocity field $v_\theta(x, s)$ to be defined on a unified time domain.

The resulting training process unifies the velocity information from all subintervals and fits a smooth vector field that generalizes beyond the discrete particle paths. This formulation distills the localized Lagrangian dynamics simulated by GPA into a global Eulerian flow model, ensuring smooth dynamics at junctions of adjacent time intervals where independently run GPA subintervals may produce mismatched velocities (Figure 2G) and enabling robust interpolation even under limited training data (Figure 2H).

Unlike conventional flow-matching methods,⁶⁵ our formulation uses empirical velocity targets derived from simulated particle dynamics. This approach enables the learned field to capture nonlinear biological transitions such as convergence, divergence, or bifurcations in cell states.

Overview of the two-step framework

Our method infers continuous-time gene expression dynamics from a small number of time-stamped single-cell snapshots by combining particle-based simulation and neural flow modeling in two successive stages.

In the first stage, we apply GPA independently to each consecutive pair of gene expression distributions $\{P_t, P_{t+1}\}$. GPA simulates discrete-time particle trajectories that approximate the Wasserstein gradient flow minimizing the Lipschitz-regularized KL divergence. At each iteration, GPA solves the greedy optimization problem to estimate the dual potential $\phi_n^{L,*}$ and updates particle positions via forward Euler steps as in Eq. (5). This process yields a time series of particle states and their corresponding velocity vectors.

In the second stage, we train a time-dependent velocity field $v_\theta(x, s)$ parameterized by a neural network to match the simulated velocities from GPA. This is done by minimizing the squared loss between predicted and simulated velocities over all time-labeled particle states. The result is a continuous vector field that generalizes beyond particle locations and interpolates temporal dynamics across the full time horizon.

Model architecture and training algorithm

To estimate the potential $\phi_n^{L,*}$, we use a fully connected feedforward network (MLP) with 3 hidden layers of 128 units each and ReLU activations, followed by a projection layer to enforce the 1-Lipschitz constraint via spectral normalization. Within each forward Euler step of the gradient flow, this potential is optimized for 5 inner gradient-descent (Adam) steps after which the computational graph is discarded before proceeding to the next step, avoiding backpropagation through the trajectory. The velocity field $v_\theta(x, s)$ is modeled as another MLP with 4 hidden layers of 64 units and Tanh activations, taking both position x and time s as inputs. This velocity field is trained via simulation-free force matching on the particle trajectories obtained from Step 1. Spectral normalization is applied to all layers to ensure Lipschitz continuity, which improves numerical stability and supports temporally smooth interpolation across time points. These architectures were selected based on preliminary experiments indicating reliable scalability to high-dimensional settings—up to ~ 100 dimensions in the GPA step and ~ 32 dimensions in the force-matching step. We further explored various hyperparameters including learning rates, iteration steps, network size, and regularization thresholds (Figures S22A–S22D). For the velocity field network, the chosen width and depth provided sufficient expressivity for modeling nonlinear dynamics while avoiding underfitting issues observed in larger architectures, which were prone to optimization instability and diminished learning efficacy.

The overall procedure is summarized in Note S11 and is illustrated in Figure 1.

Downstream Analysis Using PROFET

To extract biologically meaningful insights from the PROFET-inferred trajectories, we performed three main downstream analyses. First, we reconstructed the dynamic expression profiles of all genes provided in the input data for each individual cell. This step enables gene-level resolution of temporal behavior along each cell's trajectory. Second, although PROFET outputs trajectories at the single-cell level, we further enhanced interpretability by grouping these trajectories into subpopulations corresponding to distinct cellular behaviors or phenotypic states. Finally, by integrating trajectory classification with single-gene dynamics, we extended conventional differential gene expression analysis—typically performed at discrete time points—to a trajectory-aware framework. This approach allows us to identify key gene markers associated with divergent fate transitions, thereby uncovering regulators that may drive phenotypic bifurcations over time.

Reconstruction and evaluation of single-gene dynamics

To ensure stability and computational efficiency, PROFET was applied to reconstruct cell state trajectories in a reduced-dimensional space. Benchmarking against WOT and a null model demonstrated reliable performance with dimensions up to 30 (Figures S23 and S24). Accordingly, we first performed PCA on the gene expression matrix compiled across all input time points to reduce dimensionality. The resulting principal components were then used to represent the data in a lower-dimensional space for trajectory reconstruction. To recover gene-level information, the reconstructed trajectories were mapped back to the original gene expression space using the inverse PCA transformation,⁶⁶ thereby yielding predicted gene expression profiles for each individual cell.

To evaluate the accuracy of these reconstructions, we compared the predicted gene expression levels at held-out test time points with the actual expression data. We assessed whether the overall distribution of predicted cell states was statistically indistinguishable from that of the observed test data. We conducted permutation tests using three distributional distance metrics: Total Variation (TV) distance, Kullback-Leibler (KL) divergence, and Sinkhorn divergence.⁶⁷ For each metric, we first computed the observed distance between the predicted and observed distributions. Next, we pooled the two datasets and randomly permuted the group labels, recomputing the distance between two groups of equal size for each permutation. This process was repeated 1,000 times to generate a null distribution of distances under the hypothesis that both datasets originate from the same underlying distribution. A one-sided p-value was calculated as the fraction of permutations in which the permuted distance exceeded the observed distance.

This approach enabled us to quantitatively assess the alignment between predicted and actual cell state distributions over time. Results for the mESC and EMT datasets, including all p-values, are provided in [Tables S5](#) and [S6](#), respectively. Finally, we summarized the number of genes for which predicted and observed distributions were statistically indistinguishable ($p < 0.05$), based on one, two, or all three metrics ([Figure S25](#)).

Classification of sub-trajectories

In this study, we implemented two complementary strategies to classify predicted single-cell trajectories, providing distinct biological interpretations of cell fate dynamics. The first strategy involves classification based on cell subpopulations defined at a specific time point. These subpopulations can be identified using unsupervised clustering alone or in combination with supervised annotations such as known cell types or marker-based labels. In our analysis, we adopted an unsupervised approach using *k*-means clustering⁶⁸ to define subpopulations from the observed data. Specifically, we first reduced the dimensionality of gene expression matrices using PCA and then applied the *k*-means algorithm to the cell states at a chosen reference time point (e.g., the final day for mESC or the initial day for EMT). We used the *kmeans.fit* function from the scikit-learn Python package⁶⁹ to train the *k*-means model on the PCA-transformed gene expression profiles, which partitions the data into *k* clusters by minimizing within-cluster variance and learns the associated cluster centroids in the reduced feature space. The optimal number of clusters (*k*) was determined using the silhouette score ([Figure S26](#)).

Once the *k*-means model was trained on the observed cells, we then used it to assign predicted cell states from PROFET to the learned subpopulations. This was done using the *kmeans.predict* function, which computes the nearest centroid in Euclidean space for each predicted cell state at the reference time point and assigns it the corresponding cluster label. This strategy enables trajectory classification by linking each predicted cell to a biologically meaningful subpopulation derived from the data. For instance, in the mESC dataset, this allowed us to identify sub-trajectories that diverge into distinct terminal fates; in the EMT dataset, it enabled classification of sub-trajectories based on their inferred origins.

The second strategy classifies trajectories based on the extent of cell state transitions between two selected time points. For each trajectory, we computed the Euclidean distance between its predicted cell states at the early and late stages, using this distance as a quantitative proxy for the magnitude of phenotypic change or differentiation. Applying this procedure across all trajectories yielded a distribution of transition distances, which reflects the heterogeneity in dynamic behavior among cells. To stratify trajectories, we categorized them into low, intermediate, or high transition groups based on thresholds identified from local minima in the histogram of transition distances. This data-driven thresholding approach enables separation of distinct transition regimes. Unlike the first strategy, which aligns trajectories with known subpopulations at a fixed reference time point, this approach captures cell state transitions and highlights trajectories undergoing varying degrees of phenotypic change over time.

Differential analysis across sub-trajectories

By integrating the reconstructed gene expression dynamics along each predicted trajectory with the sub-trajectory classifications, we conducted two complementary types of differential analysis to identify gene expression differences associated with divergent cell fates. The first approach focused on quantifying temporal divergence in mean gene expression between subpopulations. For each gene, we computed the mean expression within each sub-trajectory group at every time point, then calculated the difference in means between the two groups. These differences were subsequently normalized by the maximum observed value across all time points, producing a relative divergence profile. This method allowed us to pinpoint when specific genes began to diverge between subpopulations, thereby highlighting key temporal features of gene regulation along distinct cell fates.

The second approach involved standard differential gene expression (DGE) analysis⁷⁰ at selected time points to identify gene markers that distinguish sub-trajectory groups. For each gene, we calculated the log2 fold change in expression between and assessed statistical significance using a two-sided Welch's t-test, which accounts for unequal variances between groups. These comparisons were performed at individual time points using sets of cells assigned to different sub-trajectories. By interpreting the results within the framework of sub-trajectory definitions—such as the degree of phenotypic transition or final cell identity—we were able to identify gene expression features associated with distinct developmental paths or treatment responses.

Setup of Benchmarked Methods

WOT

We employed the Waddington-OT (WOT) method⁶ to analyze cell-state transition probabilities over time in the scRNA-seq data, using normalized expression matrices and day annotations as inputs. Cell growth rates were estimated using a logistic function based on the cells' proliferation and apoptosis signatures derived from MSigDB gene sets.³⁹ These growth rates were incorporated into an unbalanced optimal transport formulation to model transitions between consecutive days, using parameters validated in the original study.⁶ Following the procedure described in the WOT tutorial, we then performed interpolation between time points. Specifically, an interpolating distribution was computed by first estimating a transport map between the two empirical distributions and then connecting each point in the first distribution to its mapped points in the second distribution according to the transport plan. Finally, a family of interpolations was constructed that continuously sweep from the first distribution to the second, generating a smooth trajectory across time.

TrajectoryNet

We employed the TrajectoryNet method⁷ to model continuous-time cellular dynamics using neural ODEs. The model was trained and evaluated on the same PCA basis as used in our framework to ensure a consistent coordinate representation across methods. The architecture of the neural network parameterizing the vector field was set to match ours, using three hidden layers with dimensions

specified as `-dims "128-128-128"`. All other hyperparameters were fixed to the default settings from the original implementation, and the model was trained without incorporating any growth rate estimation or biological priors. To ensure consistent temporal resolution, we linearly interpolated 100 evenly spaced time points along our physical time axis for trajectory generation and evaluation.

TIGON

We employed the MIOFlow framework,⁸ adapting the implementation provided in the original repository using the Dyngen example. For the PCA-based configuration, we used the same two-dimensional basis as in our experiments. When using PHATE or graph autoencoder (GAE) embeddings, the data were first projected into a 50-dimensional PCA space and then further reduced to two dimensions for visualization and analysis. Training parameters were set following the Embryoid Body Data configuration in the official notebook *01_mioflow.ipynb*, specifically: `n_local_epochs = 40` (phase 1), `n_epochs = 40` (phase 2), `n_post_local_epochs = 0` (phase 3), `reverse_schema = True`, and `reverse_n = 2`. The network architecture followed the default configuration of three layers with hidden dimensions of 16–32–16. Finally, trajectories were interpolated along the physical time axis using 100 evenly spaced time points.

PRESCIENT

We employed the PRESCIENT framework¹⁶ to model cell-state transitions in the same PCA coordinate system used across all methods for consistency. The model was trained with growth weights fixed to zero, ensuring that no external biological priors were introduced. The neural network parameterizing the potential function followed the default architecture of three layers with hidden dimensions of 32–32–32. To ensure consistent temporal resolution, we linearly interpolated 100 evenly spaced time points along our physical time axis for trajectory generation and evaluation.

MIOFlows

We employed the TIGON framework⁹ to model dynamic unbalanced optimal transport on the same PCA coordinate system used across all methods for consistency. The default neural network architecture, consisting of four layers with 16–16–16–16 hidden units, was used to parameterize the velocity field. To ensure consistent temporal resolution, we linearly interpolated 100 evenly spaced time points along our physical time axis for trajectory generation and evaluation.

CellOT

We applied the CellOT framework using the same two-dimensional PCA coordinate system used across all methods. The transport map was parameterized by an input convex neural network (ICNN) with four hidden layers of dimension 64–64–64–64 and a latent dimension of 50. The model was trained for 100,000 iterations with 10 inner discriminator iterations per step using the Adam optimizer ($\text{lr}=1\text{e-}4$, $\beta_1=0.5$, $\beta_2=0.9$) and a batch size of 256. Trajectories were generated by applying the learned transport map between consecutive observed time points.

DeepRUOT

We employed the DeepRUOT framework on the same PCA coordinate system used across all methods. The joint network (FNet) consists of velocity, growth, score, and diffusion sub-networks, each with three hidden layers of dimension 128–128–128 and LeakyReLU activation. Training proceeded in three phases: a velocity/growth pretraining phase using OT loss, a score model training phase (3,001 iterations) using the Schrödinger Bridge conditional flow matching objective with $\sigma=0.05$, and a joint fine-tuning phase. Trajectories were generated using stochastic differential equation (SDE) integration with torchsde at 100 evenly spaced time points along the physical time axis.

MMFM

We applied the Multi-Marginal Flow Matching (MMFM) framework on the same PCA coordinate system used across all methods. The vector field was parameterized by a network with three output layers of dimension 64 for the data, time, and condition embeddings, using SELU activation and Xavier initialization. The model was trained for 300 epochs with the Adam optimizer ($\text{lr}=1\text{e-}2$) and a cosine annealing learning rate schedule, using cubic interpolation and a flow variance of $\sigma=1.0$. Trajectories were generated at 100 evenly spaced time points using Euler integration.

VGFM

We applied the VGFM framework on the same PCA coordinate system used across all methods. The velocity field was parameterized by a network with three hidden layers of dimension 128–128–128 and LeakyReLU activation. Training consisted of two phases: a pretraining phase minimizing the VGFM objective for 2,000 epochs (Adam, $\text{lr}=1\times 10^{-3}$, UOT regularization $\text{reg}=0.5$, $\text{reg_m}=5$, batch size=60), followed by a refinement phase incorporating an OT loss for 30 epochs (Adam, $\text{lr}=1\times 10^{-4}$). Trajectories were generated at 10 evenly spaced time points along the physical time axis.

PI-SDE

We applied the PI-SDE framework on the same PCA coordinate system used across all methods. The drift was derived from a time-dependent potential network with two hidden layers of dimension 400–400 and softplus activation, while diffusion was set to a constant $\sigma=0.1$. The model was trained for 3,000 epochs using the Adam optimizer ($\text{lr}=5\times 10^{-3}$) with a step learning rate scheduler (step size=100, $\gamma=0.9$), Sinkhorn OT loss (blur=0.1, scaling=0.7), and gradient clipping at 0.1. Trajectories were generated at 100 evenly spaced time points along the physical time axis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Computing distributional distances between predicted and test-time distributions

To quantify how closely the predicted cell states aligned with the experimentally observed distributions at withheld time points, we computed distributional distances using three metrics: W2 distance, Sinkhorn divergence, and maximum mean discrepancy (MMD).

We approximated the W2 distance using the entropic optimal transport cost computed via the *SamplesLoss* function from the *GeomLoss* library.⁷¹ This regularized OT cost is defined as:

$$\mathcal{L}_\varepsilon(\mu, \nu) = \min_{\gamma \in \Pi(a, b)} \sum_{ij} \gamma_{ij} \|x_i - y_j\|^2 + \varepsilon \sum_{ij} \gamma_{ij} (\log \gamma_{ij} - 1), \quad (\text{Equation 8})$$

where γ is a transport plan between discrete empirical distributions $\mu = \sum_i a_i \delta_{x_i}$ and $\nu = \sum_j b_j \delta_{y_j}$ with marginals a and b , and $\varepsilon > 0$ controls the strength of entropy regularization. This formulation provides a smooth approximation of the W2 distance that is computationally stable and scalable.⁷² To obtain a proper divergence, we computed the Sinkhorn divergence by correcting for the bias introduced by the entropy term:

$$\text{Sinkhorn}_\varepsilon(\mu, \nu) = \mathcal{L}_\varepsilon(\mu, \nu) - \frac{1}{2} [\mathcal{L}_\varepsilon(\mu, \mu) + \mathcal{L}_\varepsilon(\nu, \nu)]. \quad (\text{Equation 9})$$

We tested a range of entropy regularization parameters $\varepsilon \in \{0.01, 0.1, 1, 10, 100\}$ to assess the sensitivity of this metric. Additionally, we computed MMD between predicted and observed distributions using a radial basis function (RBF) kernel:

$$\text{MMD}(X, Y) = \frac{1}{n^2} \sum_{ij} K(x_i, x_j) + \frac{1}{m^2} \sum_{ij} K(y_i, y_j) - \frac{2}{nm} \sum_{ij} K(x_i, y_j), \quad (\text{Equation 10})$$

where $K(x, y) = \exp(-\gamma \|x - y\|^2)$ is the RBF kernel with bandwidth $\gamma = 1.0 \text{ ADDIN ZOTERO_TEMP }^{72}$. All computations were performed in Python using the *GeomLoss*⁷³ and *POT (Python Optimal Transport)*⁷⁴ libraries.

Permutation testing of predicted versus observed distributions

To assess the agreement between predicted gene expression distributions and experimentally measured distributions at held-out time points, we conducted permutation-based hypothesis testing for each gene. Specifically, we evaluated whether the predicted and observed distributions could be considered samples from the same underlying distribution (null hypothesis), using three complementary distance metrics: total variation distance (TV), Kullback–Leibler divergence (KL), and Sinkhorn divergence. For each gene and each time point, we computed the observed value of the respective distance metric between the predicted and test distributions. We then permuted the pooled samples 1,000 times to construct a null distribution and estimated the empirical p -value as the proportion of permuted statistics greater than or equal to the observed one.

To evaluate the robustness of our predictions, we quantified the number of genes for which the null hypothesis could not be rejected ($p > 0.05$) under at least one, at least two, or all three distance metrics. In the mESC dataset (100 genes), at the held-out time point 1, 78 genes passed the test with at least one metric, 50 genes with at least two, and 37 genes with all three. At time point 3, these numbers increased to 100, 99, and 91, respectively. In the EMT dataset (72 genes), when predicting time point 2, 62 genes passed with at least one metric, 8 with at least two, and only 1 gene with all three. These results—summarized in [Figure S25](#), with all p -values reported in [Tables S5](#) and [S6](#)—highlight the consistency of PROFET’s predictions and underscore the value of using multiple statistical metrics to validate distributional similarity across time points.

Displacement-based quantification of phenotypic shift

To quantify the extent of phenotypic shifts inferred by PROFET, we computed the Euclidean displacement for each cell between its predicted pre-treatment and post-treatment states. Summary statistics—including the mean, standard deviation, interquartile range (IQR), coefficient of variation (CV), and entropy of the displacement distribution—were recorded for each dataset ([Table S8](#)). The full implementation can be found in the *plot_X1_hat_displacement_distribution* function, which loads the trained model and PCA parameters, computes displacements, saves statistical outputs as CSV files, and generates visualizations.

KDE-based stratification of cellular subpopulations

To identify subpopulations with distinct phenotypic shifts, we used kernel density estimation (KDE) to smooth the histogram of displacement values.⁷⁵ The density curves were computed using the Scott bandwidth method, and local minima were detected via peak-finding on the inverse KDE curve. The two smallest local minima (based on the x -axis values) were selected as cutoffs to stratify cells into three groups: low, medium, and high phenotypic shift. This analysis was performed independently for each patient dataset using the corresponding histogram of cell displacements. The KDE and local minima detection steps are implemented in a Python pipeline that reads per-sample histograms, applies KDE smoothing via *scipy.stats.gaussian_kde*, and uses *scipy.signal.find_peaks* to locate valleys in the density curve.⁷⁶

Assessing EMT Score

We computed the EMT score as a weighted sum of the expression levels of 76 EMT-related genes.⁵⁶ The weight assigned to each gene was determined by its correlation with the expression of CDH1 (E-cadherin). The resulting scores were normalized to have a mean of zero, such that negative values indicate a cell state closer to the epithelial (E) phenotype and positive values correspond to the mesenchymal (M) phenotype. To ensure consistent interpretation, we further rescaled the scores by multiplying them by -1 , thereby aligning their direction with the progression from the epithelial to mesenchymal state.