

<https://doi.org/10.1038/s42003-026-09768-8>

iAODE for benchmarking and continuum modeling of single-cell chromatin accessibility

Zeyu Fu^{1,3}✉, Chunlin Chen^{2,3}, Song Wang¹, Junping Wang¹✉ & Shilei Chen¹✉

Single-cell chromatin accessibility profiles are extremely sparse but reflect continuous developmental trajectories. Most existing methods for dimensionality reduction and trajectory analysis optimize reconstruction error or cluster separation, without encoding temporal continuity in the model or providing metrics tailored to this objective. We introduce iAODE, a variational autoencoder that couples a zero-inflated negative binomial likelihood with a latent Neural ODE, low-weight Kullback–Leibler (KL) regularization, and an interpretable reconstruction bottleneck to learn generative, temporally continuous latent spaces. Around iAODE, we build a standardized AnnData benchmark of 248 single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) and 123 single-cell RNA sequencing (scRNA-seq) datasets and a 20-metric evaluation suite that quantifies latent-space continuity, embedding quality, and clustering-coupling structure. Simulations confirm that the metrics respond smoothly to controlled continuity perturbations, and large-scale benchmarks show that the ODE, low- β , and bottleneck components synergistically improve trajectory structure and robustness over established generative and manifold-learning baselines.

Single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) characterizes open chromatin states at single-cell resolution, but the resulting data are typically extremely sparse, high-dimensional, zero-inflated, and strongly affected by batch effects, making it difficult to distinguish technical dropouts from truly inaccessible regions^{1,2}. Deep generative models, especially variational autoencoders (VAEs) and their extensions, provide a unified probabilistic framework for addressing these challenges. SCALE builds on a VAE and models binarized peak accessibility as Bernoulli observations³; PeakVI further introduces more flexible likelihoods and explicit batch-effect modeling to improve the resolution of biological heterogeneity⁴; subsequent work pushes modeling down to fragment counts to exploit count intensity more directly⁵. On this basis, SAILER penalizes information correlated with sequencing depth and batch to learn “biologically invariant” representations⁶; scMVP extends generative modeling to joint scRNA-seq and scATAC-seq in a multimodal setting⁷; GFETM integrates pre-trained genomic embeddings into a topic model to inject prior knowledge⁸.

Beyond VAEs, diffusion-based generative models provide another class of techniques. scDiffusion proposes diffusion models for conditional generation of high-quality single-cell data⁹; scButterfly performs single-cell cross-modality translation via dual-aligned variational autoencoders¹⁰. Complementing algorithmic approaches under extreme scATAC-seq sparsity, large-scale atlas resources provide reference accessibility landscapes across tissues and cell types¹¹. However, as sequencing scales increase and multi-omics measurements become more common, current methods, though successful for static latent representations, batch correction, and multimodal integration, largely treat developmental trajectories as a post-processing task: typically, trajectories are inferred after representation learning using pseudotime or graph abstraction algorithms applied to the learned embeddings^{12,13}. Systematic benchmarking further shows that different models emphasize different aspects, such as embedding quality, clustering performance, or downstream tasks¹⁴, making it difficult to answer the core question: “which type of latent structure is most favorable for recovering temporal continuity and regulatory trajectories?” Overall, the

¹State Key Laboratory of Trauma and Chemical Poisoning, Institute of Combined Injury, Chongqing Engineering Research Center for Nanomedicine, College of Preventive Medicine, Army Medical University, Chongqing, China. ²Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ³These authors contributed equally: Zeyu Fu, Chunlin Chen. ✉e-mail: fuzeyu99@126.com; wangjunping@tmmu.edu.cn; chen.shilei@foxmail.com

current landscape highlights two key gaps: (i) a lack of generative frameworks that encode temporal continuity directly in model structure, and (ii) a lack of unified evaluation and benchmarking specifically targeting continuum modeling.

Continuum modeling aims to recover the ordered evolution of cell states in latent space and support trajectory inference and time-related downstream analyses^{15,16}. For scATAC-seq, this involves smooth transitions of open-chromatin patterns across states and preservation of trajectory topology and temporal consistency under strong noise, motivating a multi-layer evaluation system. At the clustering level, metrics such as average silhouette width (ASW), Calinski-Harabasz index (CAL), and Davies-Bouldin index (DAV) assess separation and organization of cell states^{17–19}. At the embedding level, distance correlation (DC) evaluates preservation of pairwise distances; local and global quality metrics (QL, QG) derived from co-ranking matrices characterize neighborhood preservation at different scales; trustworthiness and continuity quantify local neighborhood distortion from complementary perspectives^{20–22}. At the trajectory level, the dynverse framework proposes geodesic distance correlation, Hamming-Ipsen-Mikhailov (HIM) distance, and branch F1 scores to compare inferred trajectories with references^{23–25}; partition-based graph abstraction (PAGA) abstracts cell graphs at the cluster level to identify backbones and branch structures²⁶; continuity evaluation also considers concordance between pseudotime orderings and known time labels, and whether key marker regions exhibit smooth changes along pseudotime^{15,24}. For scATAC-seq, extreme sparsity makes peak-gene link inference sensitive to noise¹⁹, requiring coherence examination at peak, gene, and multi-modal integration levels simultaneously^{27,28}.

Recent benchmarks emphasize that evaluations should integrate trajectory topology, differential accessibility, and multi-modal integration under a unified framework^{14,18,29}. However, these metrics are mostly used for post-hoc evaluation and do not directly constrain temporal dynamics at the model level, leaving room for explicit dynamical modeling. Unlike earlier metric suites designed primarily around discrete clustering or specific trajectory inference algorithms, this work focuses on continuous trajectories and smooth transitions in latent or embedding spaces rather than “continuity” of raw accessibility counts. Building on previous work, we define a 20-metric evaluation framework focused on latent continuum modeling, organized along three complementary dimensions: continuum (intrinsic latent-space continuity), embedding quality, and clustering/coupling, enabling systematic assessment of whether models can jointly capture generative distributions, trajectory structure, and interpretable regulatory modules.

To simultaneously capture temporal continuity and coordinated changes across multiple features within a deep generative model, we need explicit constraints on both latent dynamics and information structure—this is the central design motivation for iAODE. Neural ordinary differential equations (Neural ODEs) parameterize a vector field with a neural network and combine it with an ODE solver, assuming that latent states are continuously differentiable in time, which fits gradual processes such as development and differentiation. DeepVelo used Neural ODEs to fit smooth RNA kinetics in single-cell transcriptomics³⁰; related approaches have extrapolated future transcriptional states, reconstructed past trajectories^{31,32}, and modeled complex trajectories with stochastic perturbations³³. On the representation side, standard VAE Kullback–Leibler (KL) regularization encourages independence across latent dimensions, which can break the modular coupling structures that naturally arise in biological processes^{34,35}. Reducing the KL weight (low- β configuration) retains correlations between dimensions, allowing the model to capture coordinated directions corresponding to regulatory modules. Information bottleneck (IB) theory formalizes learning compressed yet informative representations³⁶; biolord and scInfoVAE instantiate this principle for single-cell data, introducing constraints that encourage latent variables to be both informative and well-structured^{37,38}.

Combining these three elements yields a continuous and coupled dynamical picture in latent space: the information bottleneck retains feature

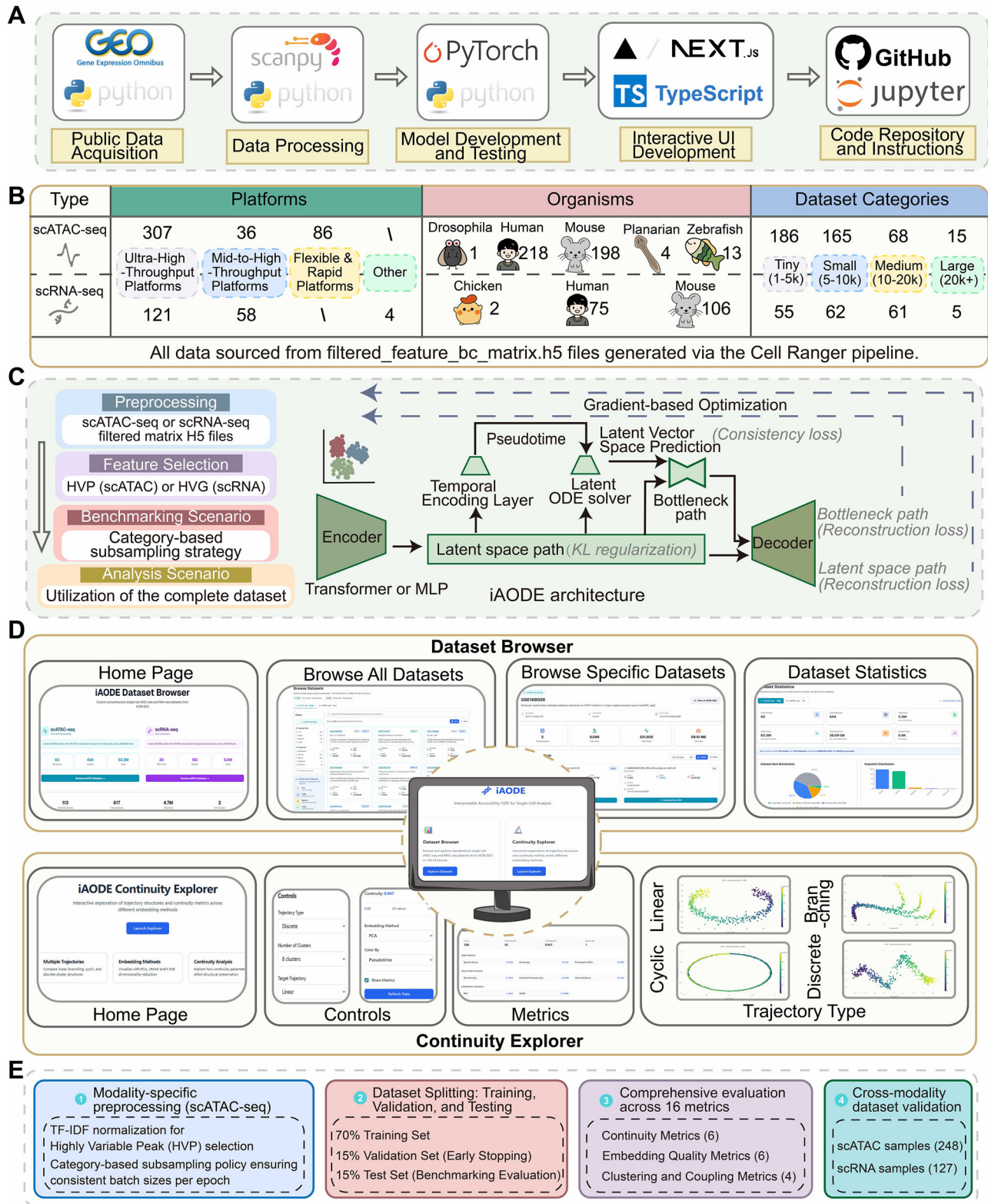
combinations relevant to trajectories; low- β KL allows these features to form correlated clusters; the Neural ODE learns a vector field over this space to model directions of coordinated change and their temporal evolution, making trajectories appear as paths along which multiple regulatory modules change in a coordinated fashion^{30,37–39}. This is consistent with classical results on β -VAE⁴⁰: when $\beta > 1$, stronger KL regularization trades reconstruction accuracy for more disentangled latent factors; when β is near 1 or slightly below, one can maintain good reconstruction while retaining correlation across dimensions. In this work, we mainly explore the low-penalty region $\beta \leq 1$ to study its effects on latent-space continuity and geometry, using high- β configurations only as controls to verify that they can impair both reconstruction and continuity. Our goal is not to seek fully factorized representations but to use low- β configurations to enhance coupling structures related to regulatory modules in latent space.

Existing deep generative models for scATAC-seq excel at static representation, batch correction, and multimodal integration but do not explicitly model continuous developmental dynamics; continuity relies on post-hoc metrics. We ask: *how to learn a latent space that is both generative and endowed with intrinsic continuous trajectory structure under extreme scATAC-seq sparsity?* We propose iAODE (interpretable Accessibility ODE-based variational autoencoder), built on a variational autoencoder (VAE) with zero-inflated negative binomial (ZINB) likelihood for discrete counts and zero inflation. The encoder predicts normalized pseudotime and introduces a Neural ODE vector field in latent space describing continuous chromatin-state evolution; low KL weights ($\beta \leq 0.1$) preserve coupling across latent dimensions associated with regulatory modules; an interpretable reconstruction (irecon) bottleneck acts as an information-bottleneck constraint encouraging interpretable accessibility patterns. We build a standardized evaluation environment with simulated topologies and multi-scale real datasets to validate continuum metrics, perform component/ablation experiments, and benchmark iAODE against deep generative models (PoissonVI, scVI, PeakVI, scTour) and traditional methods (principal component analysis (PCA), independent component analysis (ICA), Diffusion Maps). Our contributions are: (1) an ODE-VAE framework jointly modeling generative distributions and continuous trajectories for scATAC-seq; (2) standardized benchmark resources and evaluation suite spanning simulated and multi-scale real data; (3) multi-dimensional benchmarks demonstrating ODE-based frameworks’ potential for trajectory continuity, embedding quality, and biological interpretability. We focus on VAEs and variants—including high- β variant (HB, $\beta = 5$) and mutual-information/total-correlation regularizers (InfoVAE, β -TCVAE, DIP-VAE) that strengthen decoupling—to enable fair comparisons of latent-structure constraints under unified probabilistic objectives, contrasting iAODE’s low- β , coupling-emphasizing, dynamics-centric design. Systematic benchmarking of diffusion and non-VAE paradigms awaits future work building on this evaluation framework.

Results

iAODE framework and standardized multi-modal benchmark resources

We summarize the iAODE framework together with the standardized benchmark resources and evaluation protocol that support all experiments. To address the high-dimensional sparsity and heterogeneous origins of scATAC-seq data, we couple local model training with an online resource platform (Fig. 1A): core algorithms are implemented in Python/PyTorch, while a static Next.js/TypeScript site serves curated data samples, metrics, and documentation. scATAC-seq and scRNA-seq data from multiple species and platforms are obtained from public repositories and uniformly processed from Cell Ranger filtered matrices (`@ltered_feature_bc_matrix.h5`; Fig. 1B). The initial collection contains 434 scATAC-seq samples from 93 studies and 183 scRNA-seq samples from 20 studies; after stratifying by cell count into Tiny (<5k), Small (5k–10k), Medium (10k–20k), and Large (>20k) tiers and discarding Tiny data samples, the final benchmark comprises 248 scATAC-seq and 123 scRNA-seq datasets



that span diverse real biological topologies. The iAODE architecture (Fig. 1C) unifies static feature extraction and dynamic trajectory inference in a single probabilistic model: term frequency–inverse document frequency (TF-IDF)-normalized, highly variable peak/gene-filtered matrices enter an encoder that produces low-dimensional latents and a normalized pseudotime, which parametrizes a latent Neural ODE; a dual-path decoder reconstructs counts and aligns static and ODE-propagated latents via

reconstruction and consistency losses under low- β KL regularization. For each dataset, we apply a fixed protocol (Fig. 1E) with a 70%/15%/15% train/validation/test split and a 20-metric suite covering continuum quality, embedding fidelity, and clustering/coupling. An interactive “Dataset Browser” and “Continuity Explorer” (Fig. 1D) expose these standardized resources and metrics, providing a reusable platform for single-cell continuum modeling.

Fig. 1 | Overall architecture of the iAODE platform, benchmark resources, and evaluation protocol. **A** End-to-end stack and workflow: raw scATAC-seq/scRNA-seq data are downloaded from public repositories (e.g., GEO), preprocessed with Scanpy, and modeled with PyTorch-based deep generative models, while results and curated data are displayed via a Next.js/TypeScript static site; code and tutorials are hosted on GitHub. **B** Standardized multi-modal data resources: 434 scATAC-seq and 183 scRNA-seq datasets are collected and stratified by platform, species, and scale (Tiny/Small/Medium/Large). All datasets originate from Cell Ranger filtered matrices for comparability. **C** iAODE architecture: an overview of the “preprocessing-feature selection-encoding-latent ODE-dual-path decoding” pipeline, highlighting the roles of the time-encoding layer, latent ODE vector field, and irecon

bottleneck within the unified probabilistic model. **D** Interactive frontend modules: the Dataset Browser for browsing and filtering standardized AnnData resources; the Continuity Explorer for interactive visualization of latent embeddings, pseudotime, and continuum metrics across different simulated topologies. **E** Unified evaluation protocol: an overview of modality-specific preprocessing, the 70%/15%/15% train/validation/test split, and the 20-metric evaluation suite covering continuum, embedding quality, and clustering/coupling, which underpins all subsequent experiments. scATAC-seq single-cell Assay for Transposase-Accessible Chromatin using sequencing, scRNA-seq single-cell RNA sequencing, GEO Gene Expression Omnibus, ODE ordinary differential equation, AnnData annotated data matrix.

Topological simulations and continuum metrics on real scATAC data, behavior and hyperparameter priors

To validate metric behavior under controlled conditions, we evaluate continuum metrics on simulated datasets with cyclic, linear, and branching topologies in two scenarios: a *Continuum* scenario with fine-tuned continuity (0.85–0.95; Fig. 2A,B) and a *Clustering to Continuum* scenario transitioning from discrete clusters to continuous manifolds (0.00–1.00; Fig. 2C,D). In the Continuum scenario, six core metrics—Spectral Decay, Anisotropy, Participation Ratio, Trajectory Directionality, Manifold Dimensionality, and Noise Resilience—exhibit nearly linear, monotonic relationships with continuity (Fig. 2B; Supplementary Table 1). Across topologies, goodness-of-fit is consistently high (typically $R^2 > 0.8$), supporting smooth, approximately linear responses to small continuity perturbations. Clustering (ASW, CAL, DAV), embedding (DC, QL, QG), and latent coupling (COR) metrics show consistent patterns (Supplementary Fig. 1A,B). In the Clustering to Continuum scenario, linearity remains strong across the full 0–1 interval (Fig. 2D; Supplementary Fig. 1C; Supplementary Table 1), and most correlations are highly significant ($p < 0.001$), justifying metric use in real-data comparisons.

We then perform hyperparameter prescreening on 248 real scATAC-seq datasets using a baseline VAE without an ODE module to isolate regularizer effects. Friedman tests across KL weight β and irecon weight I_{rec} show significant differences (global $p < 0.001$; Fig. 2E; Supplementary Fig. 1D; Supplementary Table 2). Lower β (≤ 0.1) consistently outperforms higher- β configurations: in Medium datasets, $\beta = 0.1$ achieves advantage scores of +0.15 and +0.09 for Manifold Dimensionality and Spectral Decay versus $\beta = 10$. Increasing I_{rec} improves geometric metrics and especially Noise Resilience; in Small datasets, $\beta = 0.01$, $I_{\text{rec}} = 10$ versus $\beta = 0.01$, $I_{\text{rec}} = 0.1$ yields +0.87 for Noise Resilience (most $p < 0.001$). Notably, metric variance across β and I_{rec} is substantial in this no-ODE setting, indicating high sensitivity when latent dynamics are absent. These results motivate focusing on low- β and moderate-to-high I_{rec} ranges for iAODE, while final defaults are determined by the robustness analyses below.

Component synergy and ablations in multi-scale scATAC data

After validating metric behavior and hyperparameter ranges, we investigate the roles of low- β KL (LB, $\beta = 0.01$), irecon bottleneck (IR), and ODE module via additive and pairwise ablation experiments across 165 Small, 68 Medium, and 15 Large scATAC-seq datasets (Fig. 3A,B; Supplementary Table 3). In additive experiments (Fig. 3A), full iAODE outperforms baseline VAE (Base) in Small by +0.57 in Overall intrinsic quality, +0.86 in Noise Resilience, and +0.60 in Trajectory Directionality, with gains of +0.30–0.52 in Manifold Dimensionality, Spectral Decay, Participation Ratio, and Anisotropy (all $p < 0.001$). Compared to single-component variants (LB, IR, or ODE), Full gains +0.44–0.47 in Overall quality over LB and IR but only +0.13 over ODE; for Trajectory Directionality, gains are +0.48–0.50 relative to LB/IR and +0.08 relative to ODE, suggesting ODE substantially improves trajectory directionality while LB and IR refine geometry and denoising. For embedding quality under UMAP, Full versus Base yields +0.30, +0.26, +0.15, and +0.24 in DC, QL, QG, and OV; versus LB/IR gains are +0.17–0.25 for DC, QL, OV and +0.11–0.12 for QG, while versus ODE differences shrink to +0.02–0.05 for DC, QG, OV and +0.11 for QL (Supplementary Fig. 2A; Supplementary Table 4). Clustering metrics

show Full versus Base advantages of +0.15 in ASW, +1,028 in CAL, and +4.52 in COR; Full versus ODE reduces to +0.11, +659, and +1.39 (Supplementary Fig. 3A; Supplementary Table 5). Medium-group differences are slightly larger, and Large-group trends remain consistent.

Pairwise ablations clarify component synergy (Fig. 3B; Supplementary Fig. 3B). In Small, LB+IR (no ODE) degrades most strongly: relative to Full, Overall quality drops by 0.35, Noise Resilience by 0.65, Trajectory Directionality by 0.36, while CAL and COR decrease by 889 and 2.94. In contrast, LB+ODE and IR+ODE differ minimally from Full: Overall quality is only 0.03–0.04 lower, Trajectory Directionality 0.01–0.04 lower, and Noise Resilience/COR differ by only 0.04–0.09 and 0.55–0.60, supporting that ODE provides the global backbone for trajectories while LB and IR refine geometry and denoising. Embedding metrics mirror these patterns (Supplementary Fig. 2B): Full versus LB+IR shows +0.17, +0.15, +0.09, and +0.14 for DC, QL, QG, and OV under UMAP, whereas Full versus LB+ODE/IR+ODE yields near-zero differences in DC, QG, OV (within 0–0.01) and only +0.03–0.06 in QL. Cross-scale results show component synergy persists: Full versus Base gains in Overall quality increase from +0.57 (Small) to +0.58 (Medium) and +0.63 (Large); CAL and COR gains grow from +1,028/+4.52 (Small) to +1,394/+4.67 (Medium) and +4,958/+5.34 (Large). Overall, the three components form a cooperative system where low- β KL frees correlated directions, the irecon bottleneck filters noise, and the ODE encodes trajectories, with the full configuration consistently delivering balanced performance across continuum, embedding, and clustering metrics.

Continuum and embedding benchmarks against deep generative models across multi-scale scATAC data

We benchmark iAODE against VAE-family and deep generative baselines across Small (165), Medium (68), and Large (15) scATAC-seq datasets (Fig. 4A–C; Supplementary Table 6). Baselines include high- β variant (HB, $\beta = 5$), disentangling VAEs (DIP, TC, INFO), PoissonVI, scVI, PeakVI, and ODE-based scTour; Friedman tests and repeated-measures analysis of variance (RM-ANOVA) indicate significant overall differences (global $p < 0.001$). On continuum metrics, iAODE consistently outperforms other baselines. In Small, Full versus scVI yields +0.29 in Overall quality and +0.54 in Noise Resilience; versus PeakVI +0.21 and +0.38; versus scTour +0.13 and +0.27 (all $p < 0.001$); Manifold Dimensionality, Spectral Decay, Participation Ratio, Anisotropy, and Trajectory Directionality improve by +0.15–0.33 versus scVI and +0.05–0.13 versus scTour. Medium and Large show similar or amplified trends: in Medium, iAODE versus scVI yields +0.35 and +0.74 in Overall quality and Noise Resilience; in Large, differences grow to +0.53 and +0.81. HB, DIP, INFO, and TC sometimes achieve slightly higher Participation Ratios due to stronger decoupling but fall behind on Noise Resilience and Trajectory Directionality, consistent with strongly disentangled configurations favoring isolated factors over continuous trajectories.

Embedding quality under Uniform Manifold Approximation and Projection (UMAP) and t-distributed stochastic neighbor embedding (t-SNE) also favors iAODE or places it among top models (Fig. 4A–C; Supplementary Table 6). In Small (UMAP), iAODE versus scVI yields +0.07, +0.05, +0.05, and +0.06 in DC, QL, QG, and OV; versus scTour +0.02, +0.10, +0.02, and +0.05. In Medium, DC(UMAP), QL(UMAP),

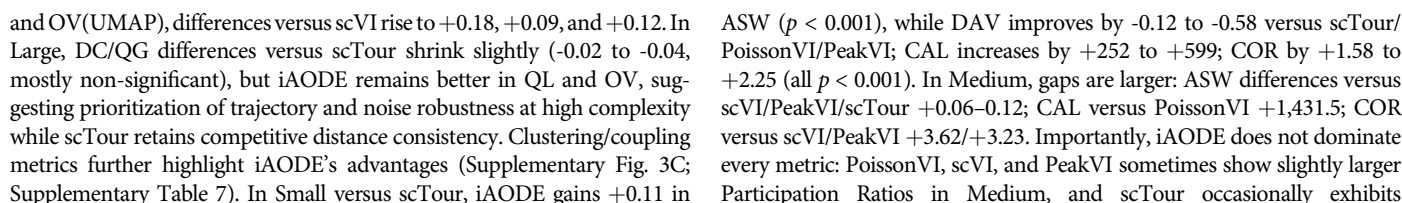


Fig. 2 | Validation of continuum metrics on simulated topologies and hyperparameter prescreening on real scATAC-seq data. **A** “Continuum” scenario visualizations: in cyclic, linear, and branching topologies, continuity is tuned from 0.85 to 0.95 (colored by pseudotime), while topology is fixed. **B** Trends of continuum metrics in the Continuum scenario: six core metrics (Spectral Decay, Anisotropy, Participation Ratio, Trajectory Directionality, Manifold Dimensionality, Noise Resilience) are plotted against the continuity setting, together with linear fits and Pearson R , slope, and R^2 , illustrating monotonic, approximately linear responses to small continuity changes. **C** “Clustering to Continuum” scenario: visualization of transitions from discrete clusters (Continuity = 0) to a continuous manifold (Continuity = 1) under the same three topologies. **D** Metric responses in the

Clustering to Continuum scenario: the same six metrics as in (B) evaluated across the full 0–1 interval, summarizing cluster-to-manifold transitions for different topologies. **E** Real scATAC-seq hyperparameter prescreening without ODE: baseline VAE without an ODE module with varying KL weights $\beta \in \{0.01, 0.1, 1, 10\}$ and irecon weights $I_{\text{rec}} \in \{0.1, 1, 10\}$ in the Medium dataset category. Friedman tests (all $p < 0.001$) indicate that low β combined with moderate I_{rec} is associated with higher manifold-geometry and noise-resilience scores, and that, in the absence of ODE constraints, these metrics are comparatively sensitive to regularization hyperparameters. VAE variational autoencoder, ODE ordinary differential equation, β , KL-divergence weight, I_{rec} irecon loss weight. Where shown, significance stars denote * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

comparable DC/QG in Large, reflecting trade-offs between local focus and compact continuity-oriented structures. Yet, combining continuum, embedding, and clustering dimensions, iAODE offers a more balanced and stable deep generative solution.

Continuum and clustering benchmarks against linear dimensionality reduction and manifold-learning methods

We compare iAODE with six traditional methods—linear reductions (principal component analysis (PCA), independent component analysis (ICA), factor analysis (FA), and non-negative matrix factorization (NMF)) and non-linear manifold tools (Diffusion Maps, Palantir)—across Small (165), Medium (68), and Large (15) scATAC-seq datasets (Fig. 5A–C; Supplementary Table 8). Friedman and RM-ANOVA tests show significant differences for most metrics (overall $p < 0.001$). On continuum metrics, iAODE outperforms linear methods consistently: in Small, Overall quality versus ICA/FA/NMF/PCA differs by +0.74, +0.74, +0.41, and +0.14; Noise Resilience versus PCA by +0.39 (all $p < 0.001$); Manifold Dimensionality and Participation Ratio versus ICA/FA/NMF/PCA differ by +0.55, +0.55, +0.28, +0.09 and +0.87, +0.87, +0.27, +0.06, respectively. Similar patterns hold in Medium and Large, reflecting limitations of linear projections in capturing complex non-linear manifolds. Comparisons with Diffusion Maps and Palantir highlight complementary strengths: for Trajectory Directionality, iAODE shows substantially higher values, with differences of +0.77, +0.73, +0.77 versus Diffusion Maps and +0.67, +0.72, +0.76 versus Palantir (Small, Medium, Large; all $p < 0.001$ or $p < 0.01$), consistent with graph-based methods tending to fragment trajectories whereas iAODE’s latent ODE provides smoother interpolation. Clustering/coupling metrics amplify these differences (Supplementary Fig. 3D; Supplementary Table 9): in Small, CAL versus Palantir and Diffusion Maps differ by +8,294 and +8,629; COR by +6.12 and +6.23; in Medium, these increase to CAL +9,088/+9,317 and COR +5.34/+5.63.

Conversely, Palantir and Diffusion Maps retain advantages on local compactness. In Small, ASW differences versus Palantir and Diffusion Maps are about −0.17 and −0.15; in Medium and Large, ASW declines by −0.18 to −0.28; UMAP/t-SNE QL values are also sometimes slightly lower for iAODE versus Palantir, while QG and OV generally favor iAODE (Supplementary Table 8). This pattern suggests graph-based methods prioritize local neighborhood preservation, whereas iAODE allows slightly looser local packing to achieve stronger global topological continuity and coupling—particularly beneficial for reconstructing continuous trajectories spanning multiple lineages. Overall, iAODE shows substantial, scale-robust improvements over linear and classical manifold-learning methods in continuum and coupling metrics, while Palantir and Diffusion Maps retain strengths for local compactness. In practice, graph-based methods can serve as complementary tools for fine-grained subpopulation structure, whereas ODE-based deep generative models are more suitable when global trajectory topology and regulatory interpretability are primary objectives.

Cross-modal transferability of iAODE components in the scRNA-seq modality and clustering-coupling performance

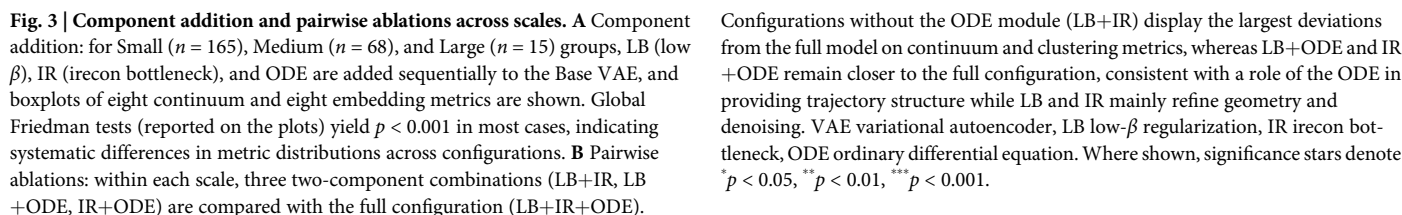
Although iAODE is designed for sparse scATAC-seq data, its latent ODE-based continuous dynamical modeling is applicable to scRNA-seq. We systematically evaluated baseline VAE (Base), single-component variants

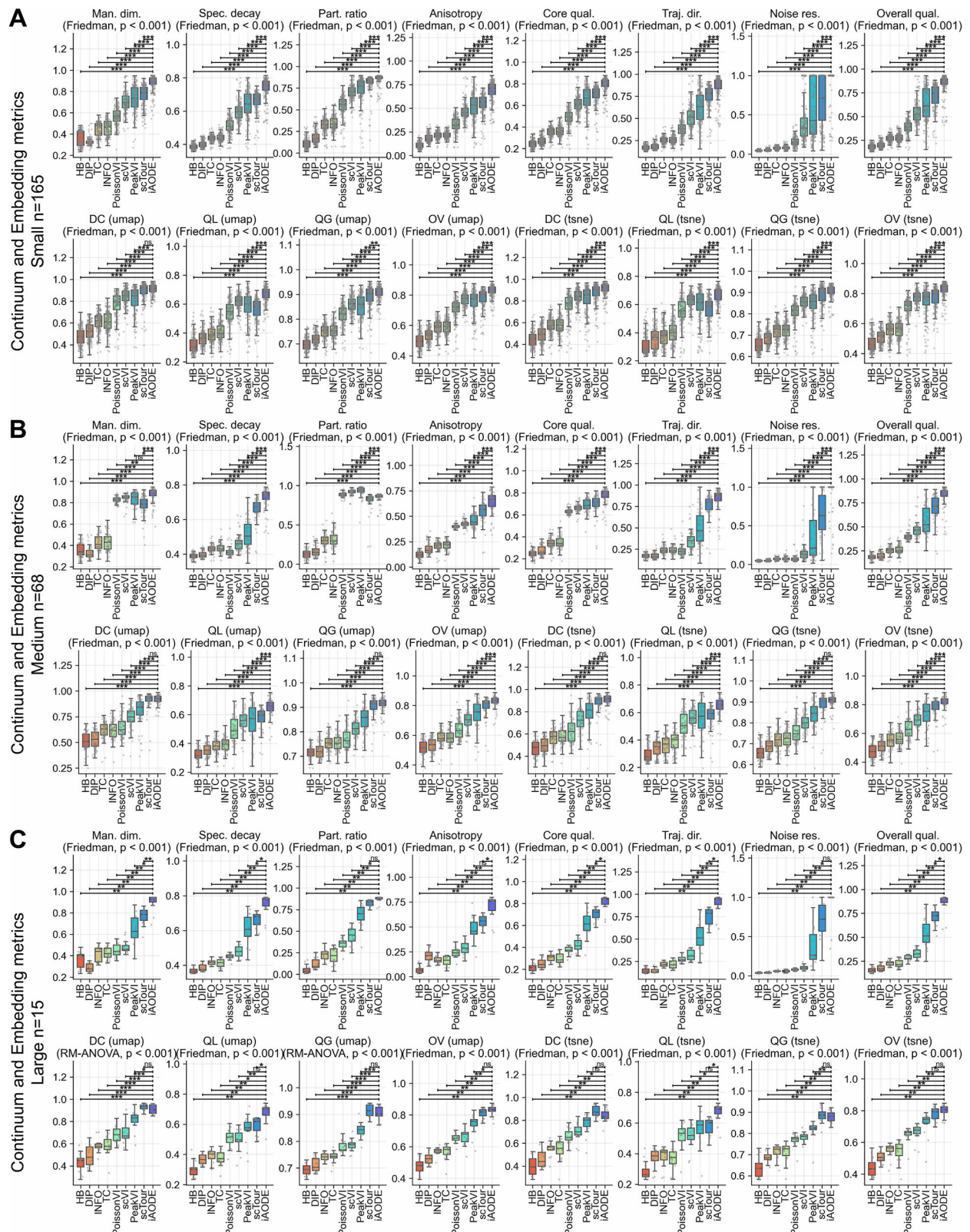
(LB, IR, ODE), and full configuration (Full: LB+IR+ODE) on 62 Small and 61 Medium scRNA-seq datasets (Figure 6A,B; Supplementary Table 10); Friedman tests indicate significant overall differences (global $p < 0.001$). For continuum metrics, full iAODE achieves performance patterns in RNA closely mirroring those in ATAC: compared with Base, Small shows median gains of approximately +0.73 and +0.95 in Overall intrinsic quality and Noise Resilience, and +0.79 in Trajectory Directionality; in Medium these gains are +0.75, +0.95, and +0.80. Manifold Dimensionality, Spectral Decay, Participation Ratio, Anisotropy, and Core intrinsic quality improve monotonically as components are added: in Medium, Full versus Base yields increases of roughly +0.59, +0.41, +0.81, +0.71, and +0.63. Comparisons with single-component variants show ODE contributes most to continuity: relative to Full, additional gains over ODE-only are only +0.37 (Small) and +0.50 (Medium) for Noise Resilience, much smaller than gains over Base. For embedding quality, Full outperforms Base on UMAP/t-SNE metrics (DC, QL, QG, OV) in both groups: in Medium, UMAP-based DC, QL, QG, and OV improve by +0.46, +0.36, +0.21, and +0.34 versus Base. Relative to ODE-only, QL and OV increase by +0.17 and +0.07, whereas DC is nearly unchanged, indicating ODEs capture most global distance structure while LB and IR improve local neighborhood geometry.

For clustering and coupling metrics, iAODE in RNA shows the same pattern of simultaneously improved cluster separation and continuity as in ATAC (Fig. 6C). Using Base as reference, Small yields gains of roughly +0.21, −0.92, and +2,714 in ASW, DAV, and CAL; for Medium these reach +0.24, −1.09, and +4,684 (all $p < 0.001$). Importantly, latent coupling (COR) rises markedly: compared with Base, COR increases by +6.38 and +6.48 in Small and Medium; even relative to ODE-only, gains of +2.90 and +3.03 remain, showing iAODE enhances cluster-level separation while strengthening coordinated variation. Pairwise ablations clarify cooperative roles (Supplementary Fig. 4A–D; Supplementary Table 11): LB+IR (without ODE) degrades most strongly in continuity and clustering; in Medium, Full versus LB+IR yields gains of +0.42, +0.47, and +0.70 in Overall quality, Trajectory Directionality, and Noise Resilience, and differences of approximately +4,291, −0.40, +0.11, and +3.99 in CAL, DAV, ASW, and COR. Overall, in scRNA-seq the three-way combination yields performance patterns highly consistent with ATAC: Full shows stable, significant gains over Base and single-component variants while maintaining or improving clustering geometry, supporting iAODE as a continuous modeling framework that transfers across single-cell modalities.

Robustness and deployability of iAODE across hyperparameters, encoder architectures, and computational cost

Having established iAODE’s relative performance across scales, baselines, and modalities, we assess its sensitivity to hyperparameters, encoder architectures, and computational requirements. iAODE is most sensitive to KL divergence weight β , while showing relaxed tolerance around reconstruction weight I_{rec} and mixing coefficient α (Fig. 7A–C; Supplementary Table 12; Supplementary Fig. 5A–C; Supplementary Table 13). In Medium ($n = 68$), increasing β from 0.1 to 10.0 causes monotonic decreases in Overall intrinsic quality, Noise Resilience, Manifold Dimensionality, Anisotropy, and Trajectory Directionality; at $\beta = 10.0$, Overall quality and Noise Resilience drop by 0.29 and 0.61, CAL is markedly reduced, DAV increases, and COR weakens (all $p < 0.001$), indicating overly strong KL regularization





substantially impairs continuity. In contrast, varying I_{rec} has milder effects: reducing I_{rec} from 10.0 to 0.1 mainly lowers Noise Resilience and Overall quality (Medium declines of 0.42 and 0.15), while core geometric metrics remain stable. Mixing coefficient α is relatively robust for continuum metrics but impacts clustering: when α decreases from 0.75 to 0.25, Medium shows modest drops of 0.14 in Overall quality, but CAL decreases by 3,200

($p < 0.001$), much larger than maximum CAL changes induced by I_{rec} tuning (800). Small ($n = 165$) exhibits consistent patterns. Based on these analyses, we use $\beta = 0.1$, $I_{rec} = 10.0$, and $\alpha = 0.75$ in main benchmarks.

At the network level, Transformer-based architectures outperform multilayer perceptrons (MLPs) for most continuity and clustering metrics, at the cost of mildly lower distance correlation (DC) (Fig. 7D;

Fig. 4 | Continuum and embedding benchmarks across multi-scale scATAC-seq datasets for deep generative models. A–C For Small ($n = 165$), Medium ($n = 68$), and Large ($n = 15$) scATAC-seq dataset groups, iAODE is compared with VAE variants and deep generative models (HB, DIP, TC, INFO, PoissonVI, scVI, PeakVI, scTour). Each panel summarizes eight continuum/manifold-geometry metrics (Manifold dimensionality, Spectral decay, Participation ratio, Anisotropy, Core intrinsic quality, Trajectory directionality, Noise resilience, Overall intrinsic quality) together with embedding-quality metrics (DC, QL, QG, OV) computed from UMAP and t-SNE embeddings. Global Friedman or RM-ANOVA p -values (mostly

$p < 0.001$) are reported above the plots, and stars mark significant pairwise differences after multiple-testing correction. Across scales, iAODE typically attains higher or comparable continuum and coupling scores and maintains competitive embedding quality relative to the other deep generative models. HB Higgins β -VAE, DIP DIP-VAE, TC β -TCVAE, INFO InfoVAE, UMAP Uniform Manifold Approximation and Projection, t-SNE t-distributed stochastic neighbor embedding, DC distance correlation, QL local quality, QG global quality, OV overall embedding quality. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (multiple-testing corrected).

Supplementary Fig. 5D; Supplementary Tables 12 and 13). In Medium, Transformers yield gains of 0.073, 0.060, 0.171, 0.129, 0.102, and 0.102 for Manifold Dimensionality, Spectral Decay, Anisotropy, Trajectory Directionality, Noise Resilience, and Overall quality (all $p < 0.001$); for clustering, CAL increases by 4,979.3, DAV decreases by 0.336, ASW increases by 0.107, and COR rises by 1.274 (all $p < 0.001$). However, DC(UMAP) and DC(t-SNE) for Transformers are lower than MLPs by 0.031 and 0.074 ($p < 0.001$), while QL improves by 0.062/0.065; QG and OV differences are negligible.

Regarding computational cost (Fig. 7E), we measured per-epoch wall-clock time and peak GPU memory as functions of cell number and fitted linear models. Because Large datasets are few and unevenly distributed, the fits use only Small and Medium datasets (up to $\sim 10^5$ cells), where iAODE, scVI, PeakVI, and PoissonVI all show approximately linear scaling and iAODE incurs a modest constant-factor overhead that remains compatible with a single 24 GB GPU. Extrapolating these Small+Medium fits to a hypothetical 10^6 -cell dataset yields predicted per-epoch times of 250.5 s (iAODE), 16.4 s (scVI), 16.0 s (PeakVI), and 17.0 s (PoissonVI), and peak GPU memories of 64.4 GB, 27.3 GB, 17.1 GB, and 14.9 GB, respectively. These extrapolated values should be viewed as upper bounds under our current architecture and fixed-batch protocol—particularly the 64.4 GB estimate for iAODE, which would require memory-saving strategies such as gradient checkpointing or multi-GPU training—whereas in the empirically evaluated regime ($\lesssim 10^5$ cells) iAODE is readily trainable on a single workstation-class GPU.

Multi-scale trajectory reconstruction and biological interpretability of latent dynamics

While the quantitative metrics above systematically compare models on continuum modeling, intuitive trajectory visualizations and biological interpretability remain crucial for assessing the practical utility of deep generative models. We therefore first visually compare iAODE and several representative baselines across three cell scales (approximately 5k, 10k, and 20k cells) on benchmark datasets (Fig. 8A). In contrast to the fragmented branches, over-clustered structures, or poorly oriented trajectories that sometimes appear in scVI, PoissonVI, and PeakVI, the latent embeddings produced by iAODE at all three scales form smooth, continuous “cell flows” with consistent directions, closely recovering the expected developmental backbone and branches.

To investigate biological meaning in latent space, we next align high-activation regions of specific latent dimensions with the spatial distributions of key marker genes and regulatory elements in a mouse brain dataset and a human peripheral blood mononuclear cell (PBMC) dataset (Fig. 8B). In the mouse brain data, high-valued regions in certain latent dimensions strongly co-localize with the distributions of canonical neurogenesis markers such as the *Notch1* gene body and the *Neurod2* promoter. In the human PBMC data, promoter regions of immunologically relevant genes such as *BICDL1* are enriched along corresponding latent directions. This spatial concordance between “latent axes” and marker features suggests that iAODE’s latent dimensions are not merely abstract statistical factors but align with concrete transcriptional regulatory modules and epigenetic states.

Finally, we perform gene ontology (GO) enrichment analysis on genes associated with high-weight features from iAODE to assess higher-level functional relevance (Fig. 8C). In the mouse brain dataset, genes linked to prominent latent features are significantly enriched for terms such as

“central nervous system development” and “neurogenesis”, consistent with tissue origin; in human PBMC, enriched terms include “lymphocyte activation” and “hematopoietic process”, matching immune and hematopoietic context. Overall, these analyses indicate that, in addition to recovering global trajectory topology and continuity, iAODE’s latent representations are biologically interpretable across marker genes, regulatory regions, and functional pathways, providing a foundation for integrating regulatory network inference and causal intervention design. We stress that the mouse brain and human PBMC analyses are computational case studies on well-characterized differentiation systems used to check whether iAODE produces plausible results under established marker and trajectory frameworks. Stronger biological validation—such as systematic comparisons with lineage tracing, time-course, and perturbation experiments—remains an important direction for future work.

Discussion

iAODE is a deep generative framework for single-cell chromatin accessibility that models continuity as a structural prior in latent space. It combines a Neural ODE vector field with a low- β KL setting and an interpretable reconstruction (irecon) bottleneck. Across simulated topologies and a standardized library of real scATAC-seq and scRNA-seq datasets, iAODE improves continuum geometry metrics, embedding quality, and clustering/coupling scores relative to VAE-family baselines and time-regularized models.

A key implication is that continuity can be enforced during representation learning rather than recovered only by downstream pseudotime or graph-based procedures^{41,42}. In ablations, the ODE component primarily supports global trajectory directionality, while low- β and irecon improve denoising and local geometry, and the full combination yields the best balance.

Chromatin accessibility is often shaped by gradual regulatory remodeling^{43,44}, making smooth latent dynamics a natural inductive bias for scATAC-seq. iAODE also transfers to scRNA-seq as a model of dominant trends, although RNA expression can include sharper transient changes that motivate alternative kinetic priors in some settings⁴⁵.

iAODE is not intended to replace all existing approaches. Topic and graph-based methods can be strong for discrete state discovery and local compactness, and diffusion-based generative models are increasingly effective for generation and augmentation, but they are not yet standardized for continuity-focused evaluation under a unified protocol. Our results suggest that when the primary objective is coherent long-range trajectory structure under strong sparsity, explicitly parameterizing latent dynamics (via an ODE) and relaxing overly factorizing regularization can provide a robust inductive bias.

Beyond the model, we provide a multi-metric evaluation suite spanning intrinsic continuum geometry, embedding fidelity, and clustering/coupling, together with a standardized benchmark collection to promote reproducible comparisons. These experiments clarify trade-offs between local compactness and global topology, and highlight iAODE as a practical option when coherent long-range structure is a priority.

The benchmarking component is an equally important contribution. By releasing a standardized multi-scale dataset collection and a multi-metric evaluation suite that separates intrinsic continuum geometry, embedding fidelity, and clustering/coupling behavior, we aim to reduce ambiguity in

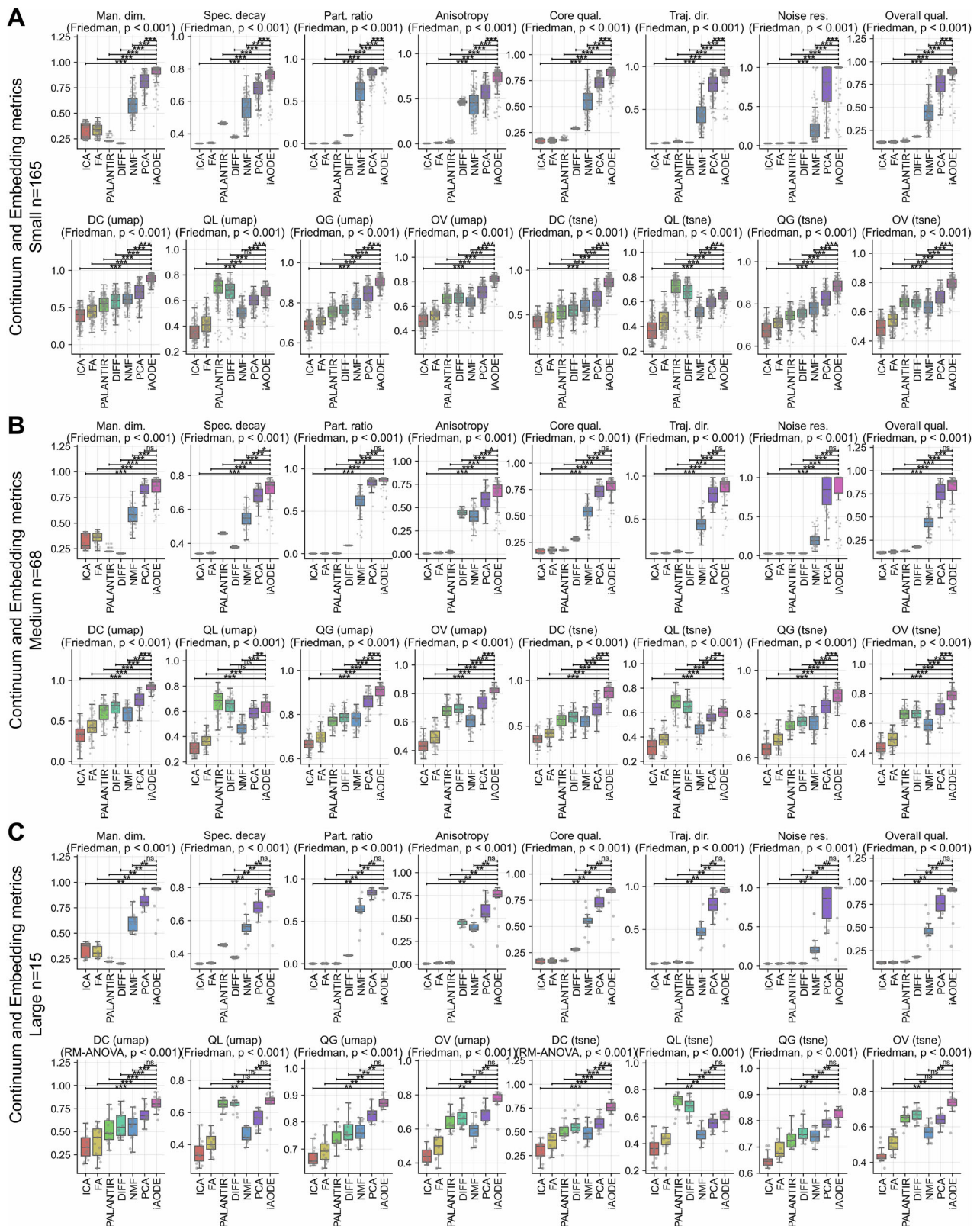


Fig. 5 | Continuum and embedding benchmarks against linear dimensionality reduction and manifold-learning methods. A–C For Small ($n = 165$), Medium ($n = 68$), and Large ($n = 15$) scATAC-seq datasets, iAODE is compared with ICA, FA, NMF, PCA, Diffusion Maps (DIFF), and Palantir. Each panel summarizes continuum/manifold-geometry metrics together with embedding-quality metrics computed from UMAP and t-SNE embeddings. Global Friedman or RM-ANOVA tests (all $p < 0.001$) indicate overall differences across methods. iAODE attains higher scores on Manifold dimensionality, Participation ratio, Noise resilience, Trajectory directionality, CAL, and COR than the linear and graph-based methods examined

here, whereas Palantir and Diffusion Maps tend to yield higher ASW and QL, indicating relatively stronger emphasis on local compactness. DIFF Diffusion Maps, UMAP Uniform Manifold Approximation and Projection, t-SNE t-distributed stochastic neighbor embedding, ASW average silhouette width, DAV Davies–Bouldin index, CAL Calinski–Harabasz index, COR correlation-based coupling metric, DC distance correlation, QL local quality, QG global quality, OV overall embedding quality. Where shown, significance stars denote * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

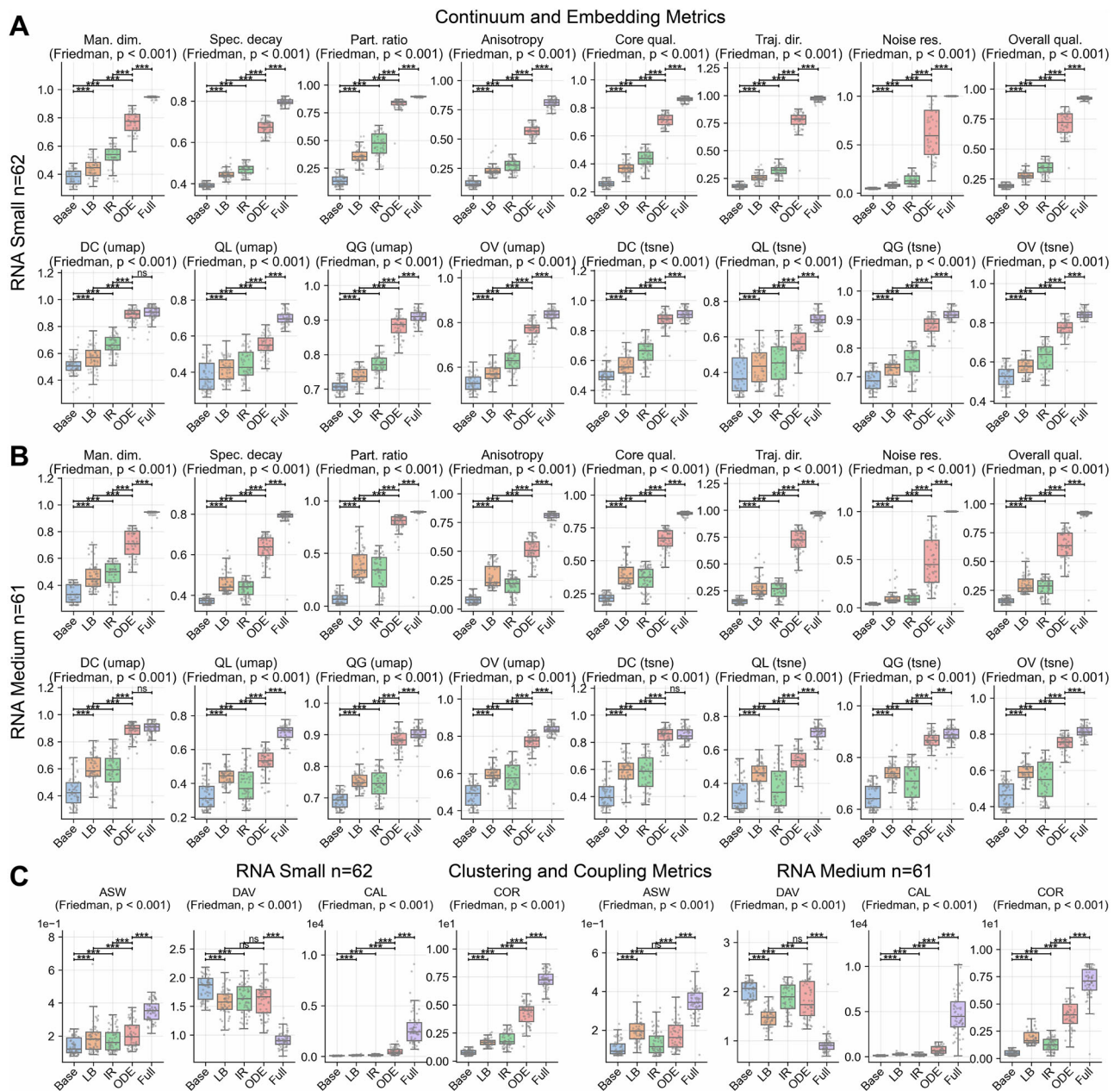


Fig. 6 | Component addition and ablations in the scRNA-seq modality: continuity, embedding quality, and clustering-coupling. **A, B** For RNA Small ($n = 62$) and RNA Medium ($n = 61$) datasets, boxplots compare the baseline VAE (Base), single-component variants (LB, IR, ODE), and the full iAODE configuration (Full: LB+IR+ODE) on eight intrinsic continuum metrics and UMAP/t-SNE embedding metrics. Global Friedman tests ($p < 0.001$) indicate systematic differences among configurations in the RNA modality, similar to the patterns observed for scATAC-seq. **C** Clustering and coupling metrics (ASW, DAV, CAL, COR) for Base and all component variants in both size groups. The full configuration yields higher cluster

separation and latent coupling than Base and single-component variants, supporting that the combination of low- β regularization, IR, and ODE is also effective for transcriptomic data. scRNA-seq single-cell RNA sequencing, VAE variational autoencoder, LB low- β regularization, IR irecon bottleneck, ODE ordinary differential equation, UMAP Uniform Manifold Approximation and Projection, t-SNE t-distributed stochastic neighbor embedding, ASW average silhouette width, DAV Davies–Bouldin index, CAL Calinski–Harabasz index, COR correlation-based coupling metric. Where shown, significance stars denote * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

future comparisons. In practice, the results show clear trade-offs (e.g., local compactness versus global directionality) that can be obscured when a single metric or a single downstream task is used as a proxy.

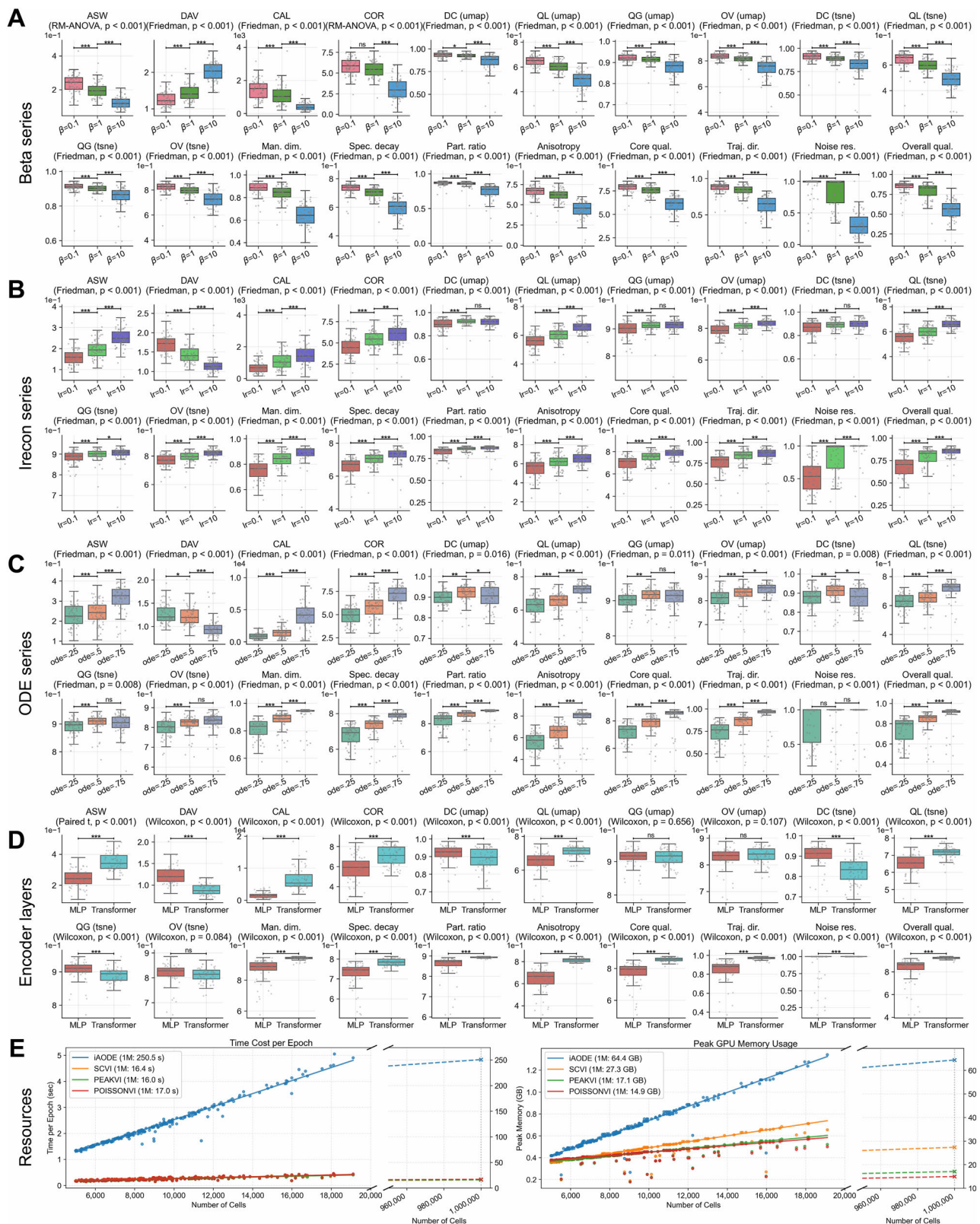
Limitations remain. Performance depends on hyperparameters (notably β , I_{rec} , and the ODE mixing weight α), and a single smooth vector field may over-smooth abrupt or multi-phase processes. Some application settings may require explicit covariate modeling beyond our default protocol, and biological validation remains indirect when only snapshot data are available. Future work could incorporate stochastic or piecewise dynamics, extend the framework to joint multi-modal

(RNA+ATAC) trajectory inference, and use time-course or lineage-tracing data to directly assess dynamical fidelity beyond marker-based interpretation⁴⁶.

Methods

iAODE latent ODE-VAE architecture and irecon bottleneck design

The encoder network E_ϕ transforms preprocessed scATAC-seq input vectors $\mathbf{x} \in \mathbb{R}^d$ (e.g., TF-IDF features) derived from raw count vectors $\mathbf{x}_{\text{raw}}$ to latent distribution parameters through a hierarchical neural network, where



d is the number of input features (peaks for scATAC-seq):

$$\mathbf{h} = \text{ReLU}(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2). \quad (1)$$

The latent distribution parameters are computed as

$$[\mu_z, \log \sigma_z^2] = \mathbf{W}_3 \mathbf{h} + \mathbf{b}_3. \quad (2)$$

Fig. 7 | Hyperparameter sensitivity, encoder architectures, and computational scalability of iAODE. **A–C** Hyperparameter sensitivity: for Small ($n = 165$) and Medium ($n = 68$) scATAC-seq datasets, KL weight $\beta \in \{0.1, 1, 10\}$, irecon weight $I_{\text{rec}} \in \{0.1, 1, 10\}$, and ODE mixing coefficient $\alpha \in \{0.25, 0.5, 0.75\}$ are varied, and distributions of continuum, embedding, and clustering/coupling metrics are compared. Global Friedman or RM-ANOVA tests ($p < 0.001$) indicate that increasing β has the strongest negative impact on continuum metrics, whereas I_{rec} and α have more moderate effects around their baseline values; small α is associated with reduced CAL. **D** Encoder architecture comparison: under matched training settings, metric differences between Transformer and MLP encoders are shown. Wilcoxon

tests indicate that Transformers tend to increase most continuity and clustering/coupling metrics, with smaller and partly mixed effects on distance-correlation metrics (DC). **E** Computational resources and scalability: per-epoch training time and peak GPU memory usage for iAODE, scVI, PeakVI, and PoissonVI as a function of cell number. All methods show approximately linear scaling with dataset size and comparable resource requirements, indicating that iAODE remains computationally tractable on a single GPU. β KL-divergence weight, I_{rec} irecon loss weight, α ODE mixing coefficient, CAL Calinski–Harabasz index, DC distance correlation. Where shown, significance stars denote $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Latent variables are sampled using the reparameterization trick:

$$\mathbf{z} = \boldsymbol{\mu}_z + \boldsymbol{\sigma}_z \odot \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (3)$$

To capture temporal dynamics in chromatin accessibility, the encoder additionally predicts a time parameter

$$t = \sigma(\mathbf{W}_t \mathbf{h} + \mathbf{b}_t), \quad (4)$$

where $\sigma(\cdot)$ ensures $t \in [0, 1]$ and represents the relative position along the accessibility trajectory.

The temporal evolution of chromatin accessibility states is modeled with a Neural ODE that learns a velocity field in latent space:

$$\frac{d\mathbf{z}}{dt} = f_\varphi(t, \mathbf{z}) = \mathbf{W}_f^{(2)} \text{ELU}(\mathbf{W}_f^{(1)} \mathbf{z} + \mathbf{b}_f^{(1)}) + \mathbf{b}_f^{(2)}. \quad (5)$$

This enables continuous modeling of chromatin-state transitions and regulatory dynamics inherent in scATAC-seq data.

To enforce biologically meaningful and structured representations, we implement an interpretable reconstruction (irecon) module that creates a compressed bottleneck in latent space:

$$\mathbf{z}_I = \text{Linear}(\mathbf{z}; \mathbf{W}_I, \mathbf{b}_I), \hat{\mathbf{z}} = \text{Linear}(\mathbf{z}_I; \mathbf{W}_O, \mathbf{b}_O), \quad (6)$$

where $\mathbf{z}_I \in \mathbb{R}^{d_I}$ is a compressed irecon representation with $d_I < d_z$, d_z is the latent space dimension, capturing essential chromatin-accessibility patterns while reducing redundancy.

The decoder network D_ψ reconstructs scATAC-seq counts from latent representations. Given the discrete and sparse nature of accessibility measurements, we parameterize the reconstruction with a softmax-normalized output:

$$\hat{\mathbf{x}} = \text{Softmax} \left(\mathbf{W}_d^{(2)} \text{ReLU}(\mathbf{W}_d^{(1)} \mathbf{z} + \mathbf{b}_d^{(1)}) + \mathbf{b}_d^{(2)} \right). \quad (7)$$

For zero-inflated scenarios, the decoder additionally predicts dropout probabilities

$$\boldsymbol{\pi} = \mathbf{W}_\pi \mathbf{h}_d + \mathbf{b}_\pi, \quad (8)$$

where $\mathbf{h}_d = \text{ReLU}(\mathbf{W}_d^{(1)} \mathbf{z} + \mathbf{b}_d^{(1)})$ is the decoder hidden state.

The overall iAODE-VAE objective combines multiple loss components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{irecon}} + \mathcal{L}_{\text{ODE}} + \mathcal{L}_{\text{KL}}. \quad (9)$$

For scATAC-seq counts we use a ZINB reconstruction loss:

$$\mathcal{L}_{\text{recon}} = -\mathbb{E}_{q(\mathbf{z}|\mathbf{x})} [\log p_{\text{ZINB}}(\mathbf{x}|\hat{\mathbf{x}}, \boldsymbol{\theta}, \boldsymbol{\pi})], \quad (10)$$

where $\hat{\mathbf{x}} = \text{Softmax}(\mathbf{D}_\psi(\mathbf{z})) \cdot \|\mathbf{x}_{\text{raw}}\|_1$ ensures library-size normalization with $\|\mathbf{x}_{\text{raw}}\|_1 = \sum_j x_{\text{raw},j}$ being the cell library size, $\boldsymbol{\theta}$ denotes peak-specific dispersion parameters, and $\boldsymbol{\pi}$ models technical dropouts.

The irecon module defines a compressed reconstruction loss

$$\mathcal{L}_{\text{irecon}} = -\mathbb{E}_{q(\mathbf{z}_I|\mathbf{z})} [\log p_{\text{ZINB}}(\mathbf{x}|\hat{\mathbf{x}}_I, \boldsymbol{\theta}_I, \boldsymbol{\pi}_I)], \quad (11)$$

where $\mathbf{z}_I \in \mathbb{R}^{d_I}$ with $d_I \ll d_z$ enforces an information-constrained latent space that encourages learning of essential accessibility patterns. Here, $\boldsymbol{\theta}_I$ and $\boldsymbol{\pi}_I$ play the same roles as $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ in the main decoder, but are predicted from the compressed representation $\mathbf{z}_I$ and thus capture bottleneck-specific dispersion and dropout patterns. Analogously, $\hat{\mathbf{x}}_I$ denotes the ZINB mean reconstructed from the compressed representation $\mathbf{z}_I$ and is rescaled by the same library size $\|\mathbf{x}_{\text{raw}}\|_1$ as $\hat{\mathbf{x}}$, so that both reconstruction terms operate on a comparable count scale.

Temporal consistency is encouraged via an ODE loss:

$$\mathcal{L}_{\text{ODE}} = \|\mathbf{z} - \mathbf{z}_{\text{ODE}}\|_2^2, \quad (12)$$

where $\mathbf{z}_{\text{ODE}}$ is the evolved latent state produced by integrating $\frac{d\mathbf{z}}{dt} = f_\varphi(t, \mathbf{z})$.

The standard KL divergence matches the posterior to the prior:

$$\mathcal{L}_{\text{KL}} = \text{KL}(q(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})). \quad (13)$$

For trajectory inference in scATAC-seq data, we construct transition matrices from ODE-derived velocity fields:

$$\mathbf{v} = f_\varphi(t, \mathbf{z}), \mathbf{z}_{\text{future}} = \mathbf{z} + \epsilon \cdot \mathbf{v}. \quad (14)$$

Transition probabilities use adaptive Gaussian kernels with median-based bandwidth:

$$P_{ij} = \frac{\exp(-\|\mathbf{z}_i - \mathbf{z}_{\text{future},j}\|^2 / 2\sigma^2)}{\sum_k \exp(-\|\mathbf{z}_i - \mathbf{z}_{\text{future},k}\|^2 / 2\sigma^2)}. \quad (15)$$

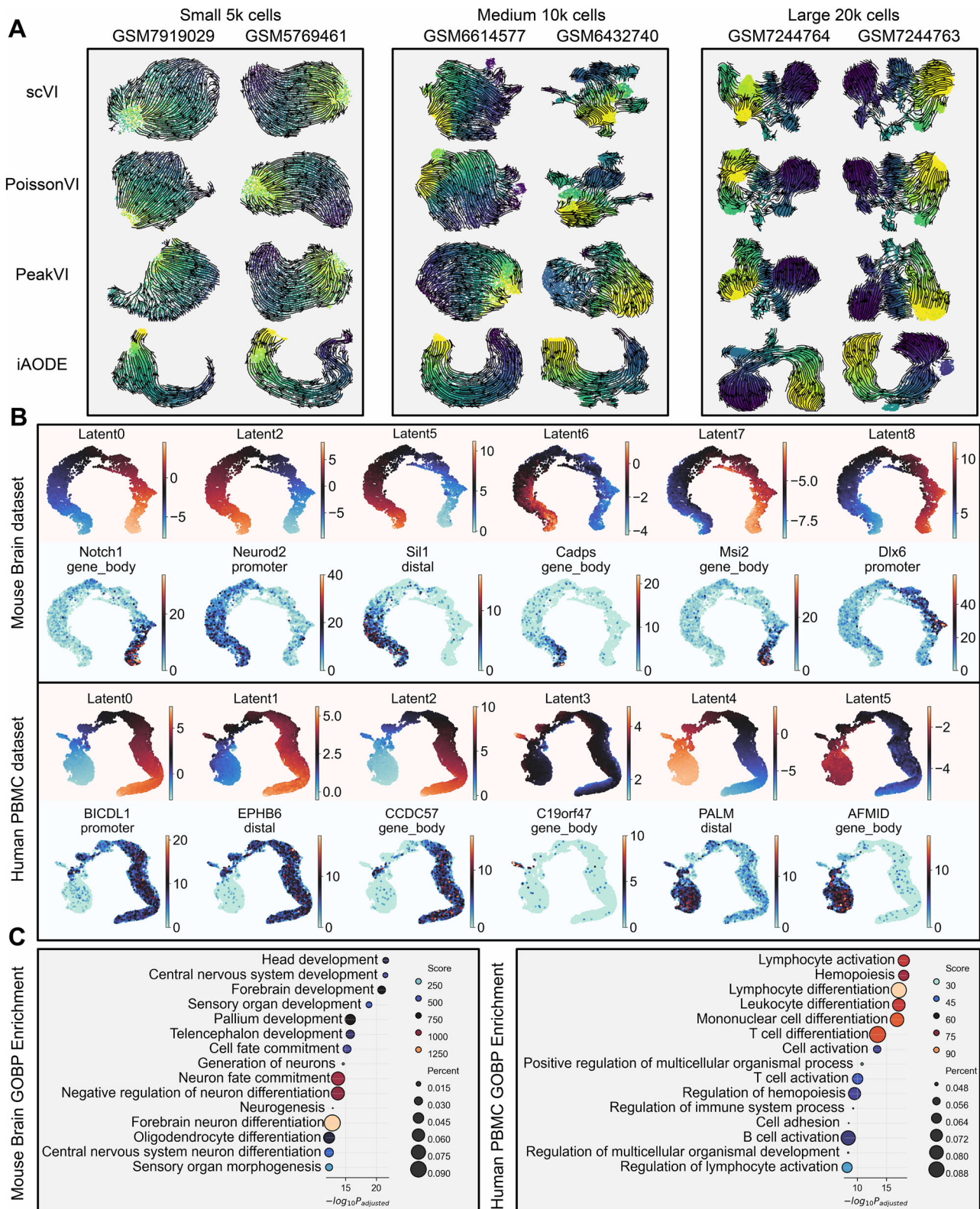
The final representation linearly combines static and dynamic latents:

$$\mathbf{z}_{\text{final}} = (1 - \alpha) \mathbf{z} + \alpha \mathbf{z}_{\text{ODE}}, \quad (16)$$

where $\alpha \in [0, 1]$ controls the strength of ODE dynamics, enabling joint modeling of chromatin landscapes and regulatory dynamics.

Standardized benchmark library construction and quality control

To ensure reproducibility and facilitate systematic benchmarking, we constructed a standardized library of scATAC-seq and scRNA-seq datasets from publicly available sources in the Gene Expression Omnibus (GEO). All datasets were derived from Cell Ranger standard outputs: for scATAC-seq we collected `filtered_peak_bc_matrix.h5` files containing fragment counts per cell for consensus peak regions, and for scRNA-seq we collected `filtered_feature_bc_matrix.h5` files containing unique molecular identifier (UMI) counts per gene. Because Cell Ranger applies



standard quality-control (QC) filters during matrix generation—removing low-quality cells based on total counts, fraction of reads in peaks or genes, and other technical metrics—these filtered matrices represent data that have already passed instrument- and protocol-specific QC thresholds. We therefore use these outputs directly without imposing additional uniform filtering criteria, recognizing that optimal QC varies across experimental protocols, tissue types, and sequencing depths.

After loading each filtered matrix, we categorized datasets by cell count N into four size groups: Tiny ($N < 5,000$), Small ($5,000 \leq N < 10,000$), Medium ($10,000 \leq N < 20,000$), and Large ($N \geq 20,000$). Our initial collection comprised 434 scATAC-seq datasets distributed as 186 Tiny, 165 Small, 68 Medium, and 15 Large, and 183 scRNA-seq datasets. To ensure statistical robustness for manifold geometry estimation and continuum evaluation, we excluded Tiny-scale datasets and, for scRNA-seq, focused on Small and

Fig. 8 | Multi-scale trajectory visualization and biological interpretability of iAODE latent dynamics. **A** Latent-space embeddings and flow-field trajectories across scales. For three benchmark datasets (Small, ~5k cells; Medium, ~10k cells; Large, ~20k cells, labeled by GSM IDs), low-dimensional embeddings and inferred streamlines from iAODE are described alongside embeddings from scVI, PoissonVI, and PeakVI. In these examples, iAODE forms smooth continuous trajectories at all scales, whereas baseline methods sometimes display more fragmented or cluster-like structures. **B** Alignment between latent features and biological markers. In mouse brain and human PBMC datasets, activation patterns of selected latent dimensions (Latent0-8) are compared with the embedded distributions of marker genes or

regulatory regions (for example, *Notch1* gene body, *Neurod2* promoter, *BICDL1* promoter). Spatial co-localization suggests that certain latent directions correspond to specific regulatory or marker-associated patterns. **C** Gene Ontology enrichment analysis. Bubble plots show significantly enriched biological processes for genes associated with high-weight iAODE features in the mouse brain and human PBMC (shown in separate subplots). Enriched terms include central nervous system development and neurogenesis in the mouse brain and lymphocyte activation and hematopoietic processes in PBMC, indicating that the learned representations capture biologically meaningful functional programs in these examples. GSM GEO sample accession, PBMC peripheral blood mononuclear cell, GO Gene Ontology.

Medium scales due to limited availability of large-scale RNA data. After these filters, our final standardized benchmark library comprises 248 scATAC-seq datasets (165 Small, 68 Medium, 15 Large) and 123 scRNA-seq datasets (62 Small, 61 Medium). All datasets were converted to AnnData objects with consistent structure.

scATAC-seq preprocessing and selection of highly variable peaks

Across all benchmark datasets we use a unified scATAC-seq preprocessing and feature selection pipeline. Let $\mathbf{X} \in \mathbb{R}^{N \times d}$ be the raw fragment count matrix, where N is the number of cells and d the number of peaks. For each cell i we compute its library size

$$L_i = \sum_{j=1}^d X_{ij}, \quad (17)$$

and obtain per-cell normalized term frequency (TF):

$$\text{TF}_{ij} = \frac{X_{ij}}{L_i}. \quad (18)$$

For each peak j , we compute the number of accessible cells n_j and its accessible fraction n_j/N , and then define the inverse document frequency (IDF) following standard Signac/SnapATAC2 practice:

$$\text{idf}_j = \log \left(1 + \frac{N}{n_j} \right). \quad (19)$$

The resulting TF-IDF matrix is

$$\tilde{X}_{ij} = \text{TF}_{ij} \cdot \text{idf}_j \cdot s, \quad (20)$$

where s is a global scaling factor (default $s = 10^4$). We store $\tilde{\mathbf{X}}$ as a sparse matrix to maximize efficiency for large datasets.

For feature selection, we first pre-filter peaks by accessibility fraction: we retain only peaks whose accessibility rate lies in $[p_{\min}, p_{\max}]$ (with default thresholds $p_{\min} = 0.01$ and $p_{\max} = 0.99$) to remove extremely rare and nearly ubiquitously accessible peaks. Denote the filtered peak set by $\mathcal{P}$. We then compute variance- or variance-to-mean-based variability measures for each peak in $\mathcal{P}$, rank them, and select the top n_{top} highly variable peaks as the final feature subset, used both for training iAODE and as inputs to comparison methods. By default we keep 20,000 highly variable peaks for scATAC-seq and 5000 highly variable genes for scRNA-seq (using Scanpy's default highly variable gene (HVG) selection pipeline).

iAODE training strategy and dataset splitting

After preprocessing and feature selection, each dataset is split into training, validation, and test sets with a default ratio of 0.7/0.15/0.15, using a fixed

random seed for reproducibility. For each mini-batch, we form pairs $(\mathbf{x}, \mathbf{x}_{\text{raw}})$, where $\mathbf{x}$ is the TF-IDF-transformed and peak-filtered input and $\mathbf{x}_{\text{raw}}$ is the corresponding raw count vector. The encoder estimates the latent posterior $q(\mathbf{z}|\mathbf{x})$ and time variable t in log-transformed space; the decoder predicts library-normalized means which are rescaled by the cell-specific library size $L = \|\mathbf{x}_{\text{raw}}\|_1$ to reconstruct counts. These predictions are used to compute the reconstruction loss $\mathcal{L}_{\text{recon}}$ and irecon loss $\mathcal{L}_{\text{irecon}}$ under a negative binomial (NB) or ZINB likelihood.

We then add the ODE consistency term $\mathcal{L}_{\text{ODE}} = \|\mathbf{z} - \mathbf{z}_{\text{ODE}}\|_2^2$ and KL regularization $\mathcal{L}_{\text{KL}} = \text{KL}(q(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}))$. In practice, we weight each loss component as

$$\mathcal{L}_{\text{total}} = \lambda_{\text{recon}} \mathcal{L}_{\text{recon}} + \lambda_{\text{irecon}} \mathcal{L}_{\text{irecon}} + \lambda_{\text{ODE}} \mathcal{L}_{\text{ODE}} + \beta \mathcal{L}_{\text{KL}}, \quad (21)$$

where λ_{recon} , λ_{irecon} (denoted I_{rec} in hyperparameter analysis), λ_{ODE} , and β control the relative contributions of reconstruction, irecon bottleneck, ODE consistency, and KL regularization, respectively.

We optimize with Adam in a mini-batch setting. After every fixed number of epochs we compute validation loss and internal clustering/coupling metrics (ASW, CAL, DAV, COR) and use early stopping with patience: if validation loss does not improve for a predefined number of evaluations, training is terminated, and parameters are rolled back to the best epoch. This ensures convergence while mitigating overfitting and reducing compute cost.

For the iAODE objective, we set $\lambda_{\text{recon}} = 1.0$, $\lambda_{\text{irecon}} = 1.0$, and $\lambda_{\text{ODE}} = 1.0$. When illustrating the “LB” configuration we use a low- β KL weight $\beta = 0.01$ to enhance latent correlation. The baseline VAE and other VAE variants use $\beta = 1$ as standard, while a High-Beta (HB) variant uses a stronger KL weight $\beta = 5$ to represent more disentangled, cluster-oriented configurations. In the hyperparameter sensitivity analysis, we systematically vary $\beta \in \{0.1, 1, 10\}$, $I_{\text{rec}} \in \{0.1, 1, 10\}$, and the mixing coefficient $\alpha \in \{0.25, 0.5, 0.75\}$. Based on these sensitivity results, the main benchmarks in this manuscript use $\beta = 0.1$, $I_{\text{rec}} = 10.0$, and $\alpha = 0.75$. Latent ODE trajectories are solved using an adaptive-step numerical integrator from the PyTorch ODE stack, which automatically adjusts internal step sizes based on error control; explicit time-step size is thus not exposed as a separate hyperparameter.

Baseline models and unified implementation of regularization terms

For baseline and ablation comparisons we include common regularization strategies used in deep generative modeling to assess how different latent-structure constraints affect continuum modeling and representation quality in a unified framework. These regularizers are only enabled when reproducing classical β -VAE, DIP-VAE, β -TCVAE, and InfoVAE baselines and are not essential components of iAODE.

The β -VAE framework augments the reconstruction loss with a weighted KL penalty:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{KL}}, \quad \mathcal{L}_{\text{KL}} = \text{KL}(q(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})), \quad (22)$$

where higher $\beta > 1$ encourages more disentangled latent factors while $\beta < 1$ retains more correlations⁴⁰.

DIP-VAE further constrains the covariance of the latent means by penalizing deviations of $\text{Cov}_{q(z)}[\mu_z]$ from the identity:

$$\mathcal{L}_{\text{DIP}} = \left\| \text{offdiag} \left(\text{Cov}_{q(z)}[\mu_z] \right) \right\|_F^2 + \lambda_{\text{diag}} \left\| \text{diag} \left(\text{Cov}_{q(z)}[\mu_z] \right) - \mathbf{1} \right\|_2^2, \quad (23)$$

where offdiag and diag select off-diagonal and diagonal entries respectively, and λ_{diag} balances the two penalties⁴⁷.

β -TCVAE isolates total correlation (TC) in the KL decomposition and penalizes it explicitly:

$$\text{TC}(\mathbf{z}) = \text{KL} \left(q(\mathbf{z}) \parallel \prod_j q(z_j) \right), \quad (24)$$

with regularizer $\mathcal{L}_{\text{TC}} = \gamma \text{TC}(\mathbf{z})$, where γ controls the strength of disentanglement⁴⁸.

InfoVAE and related mutual-information extensions add mutual information and maximum mean discrepancy (MMD) terms to balance reconstruction fidelity and prior matching:

$$\mathcal{L}_{\text{Info}} = \mathcal{L}_{\text{recon}} + \alpha_{\text{Info}} \mathcal{L}_{\text{KL}} + \lambda_{\text{MMD}} \text{MMD}(q(\mathbf{z}), p(\mathbf{z})), \quad (25)$$

where MMD measures the discrepancy between $q(\mathbf{z})$ and $p(\mathbf{z})$ in kernel mean embedding space, and α_{Info} and λ_{MMD} weight KL and MMD respectively. In our baselines, we use the same encoder/decoder architecture and switch regularization terms and weights to reproduce or approximate DIP-VAE, β -TCVAE, InfoVAE, and related methods, enabling quantitative comparisons of different latent-structure constraints under the same implementation and metrics⁴⁹.

Comprehensive evaluation of continuity, embedding quality, clustering, and coupling

To assess the quality of learned latent spaces—especially the ability to *preserve biological coupling and model continuous trajectories of cell states*—we design a multi-dimensional evaluation system. It covers standard clustering performance as well as co-ranking analysis and spectral-geometric features to quantify intrinsic manifold properties independently of 2D visualization. Our final framework uses 20 metrics across three complementary dimensions: 8 intrinsic metrics describing latent-space continuity and manifold geometry, 8 structure-fidelity metrics based on 2D UMAP/t-SNE embeddings, and 4 clustering/coupling metrics.

Let $\mathbf{X} \in \mathbb{R}^{N \times d}$ be the original high-dimensional data and $\mathbf{Z} \in \mathbb{R}^{N \times d_z}$ the learned latent representation, where N is the number of cells and d_z is the latent dimension.

Intrinsic continuity and manifold-geometry metrics.

We consider a set of metrics based on the eigenvalue spectrum $\lambda = [\lambda_1, \dots, \lambda_{d_z}]$ of the covariance matrix in latent space to characterize intrinsic geometry, independent of explicit 2D projections.

Manifold dimensionality score (M_{dim}). This measures representational compactness. Higher values indicate that essential biological variation is captured with fewer effective dimensions:

$$M_{\text{dim}} = 1 - \frac{d_{\text{eff}} - 1}{d_z - 1}, \quad (26)$$

where d_{eff} is the number of dimensions required to explain 95% of the variance.

Spectral decay (S_{decay}). This quantifies how steeply eigenvalues decline. A stronger decay suggests a clear hierarchical structure, with dominant

developmental axes and secondary variation patterns:

$$S_{\text{decay}} = \frac{1}{1 + \exp(\kappa)} \cdot \frac{\lambda_1}{\sum_{i=1}^{d_z} \lambda_i}, \quad (27)$$

where κ is the slope of a linear regression on log-eigenvalues.

Participation ratio (P_{ratio}). This measures how evenly variance is distributed across dimensions. Higher values indicate balanced representation of multiple biological processes rather than reliance on a few dominant factors:

$$P_{\text{ratio}} = \frac{(\sum_i \lambda_i)^2}{d_z \sum_i \lambda_i^2}. \quad (28)$$

Anisotropy score (A_{score}). Using a hyperbolic tangent transform of log-eigenvalues, this quantifies the directional strength of the manifold. Large values indicate strong anisotropy, often corresponding to clear differentiation pathways:

$$A_{\text{score}} = \tanh \left(\frac{\log(\lambda_1) - \log(\lambda_{d_z} + \epsilon)}{4} \right), \quad (29)$$

where $\epsilon = 10^{-8}$ is a small constant to prevent division by zero.

Trajectory directionality (T_{dir}) and noise resilience (N_{res}). The former measures dominance of the first principal component as a putative developmental axis; the latter approximates a signal-to-noise ratio assessing separation of biological signal from technical noise:

$$T_{\text{dir}} = \frac{\lambda_1}{\sum_{i=2}^{d_z} \lambda_i + \epsilon}, \quad N_{\text{res}} = \min \left(\frac{\lambda_1 + \lambda_2}{\sum_{i=3}^{d_z} \lambda_i + \epsilon} \cdot \frac{1}{10}, 1 \right). \quad (30)$$

Composite manifold-quality scores.

We define two composite scores to summarize manifold quality, distinguishing core geometric integrity from trajectory-oriented performance.

Core intrinsic quality (Q_{core}). This aggregates M_{dim} , S_{decay} , P_{ratio} , and A_{score} to capture basic geometric properties relevant for generic representation learning:

$$Q_{\text{core}} = \frac{1}{4} (M_{\text{dim}} + S_{\text{decay}} + P_{\text{ratio}} + A_{\text{score}}). \quad (31)$$

Overall intrinsic quality (Q_{overall}). This task-oriented score combines core geometry with trajectory-specific metrics via a weighted sum:

$$Q_{\text{overall}} = w_{\text{core}} \cdot Q_{\text{core}} + w_{\text{dir}} \cdot T_{\text{dir}} + w_{\text{noise}} \cdot N_{\text{res}}, \quad (32)$$

with $(w_{\text{core}}, w_{\text{dir}}, w_{\text{noise}}) = (0.5, 0.3, 0.2)$. This gives highest weight to core structure ($w_{\text{core}} = 0.5$), substantial weight to trajectory directionality ($w_{\text{dir}} = 0.3$) to reflect developmental analysis needs, and moderate weight to noise resilience ($w_{\text{noise}} = 0.2$) to handle sparse, noisy single-cell data.

Embedding-quality metrics.

To assess structural fidelity of high-dimensional latent representations under commonly used 2D visualizations, we fix UMAP/t-SNE hyperparameters and project each model's latent space to 2D under this common protocol, then compute distance-correlation and co-ranking-based metrics. Crucially, we are evaluating how well different latent spaces preserve structure *under the same 2D operator*, not comparing visualization algorithms themselves, which here serve only as standard embedding operators for downstream visualization scenarios.

Distance correlation (DC). We compute Spearman correlation between vectorized pairwise distance matrices in high and low dimensions, capturing global structural preservation:

$$\rho_{\text{dist}} = \rho_s(\text{vec}(\mathbf{D}_{\text{high}}), \text{vec}(\mathbf{D}_{\text{low}})), \quad (33)$$

where $\mathbf{D}_{\text{high}}$ and $\mathbf{D}_{\text{low}}$ are pairwise distances in the original and embedded spaces, ρ_s denotes Spearman's rank correlation, and $\text{vec}(\cdot)$ vectorizes a matrix by stacking its columns.

Local and global quality (QL, QG). Based on the co-ranking matrix $\mathbf{Q}$, where Q_{kl} counts neighbor pairs ranked k -th in high dimension and l -th in low dimension, we define

$$Q_{\text{NX}}(K) = \frac{1}{KN} \sum_{k=1}^K \sum_{l=1}^K Q_{kl}. \quad (34)$$

We use the local continuity meta-criterion (LCMC) to find the optimal neighborhood size

$$K_{\text{max}} = \arg \max_K \text{LCMC}(K), \quad (35)$$

and define

$$Q_{\text{local}} = \frac{1}{K_{\text{max}}} \sum_{k=1}^{K_{\text{max}}} Q_{\text{NX}}(k), \quad Q_{\text{global}} = \frac{1}{N - K_{\text{max}}} \sum_{k=K_{\text{max}}+1}^N Q_{\text{NX}}(k). \quad (36)$$

K_{max} is an intermediate quantity, not reported as an independent metric. We define overall embedding quality (OV) as

$$Q_{\text{embed}} = \frac{1}{3}(\rho_{\text{dist}} + Q_{\text{local}} + Q_{\text{global}}). \quad (37)$$

Latent representations, clustering quality, and coupling.

Although our focus is on continuous trajectories and manifold structure rather than purely discrete clustering, we retain intrinsic clustering metrics that depend directly on latent distances to constrain continuity from the perspective of “state segments” (e.g., cell types or developmental stages). Unlike external metrics such as normalized mutual information (NMI) or adjusted Rand index (ARI), ASW, DAV, and CAL here are computed solely from latent-space distances and do not rely on external label assignments; they capture cluster compactness and separation and partly reflect latent geometry. High scores under continuum modeling indicate that the model preserves smooth trajectories and topology while maintaining local separability and group boundaries—important when distinguishing discrete types within continuous pseudotime.

Average silhouette width (ASW). This measures intra-cluster cohesion and inter-cluster separation:

$$\text{ASW} = \frac{1}{N} \sum_{i=1}^N \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (38)$$

where a_i is the mean distance from sample i to points in the same cluster and b_i is the mean distance to the nearest other cluster. Values near 1 indicate well-separated, compact clusters.

Calinski-Harabasz index (CAL). This is the ratio of between-cluster to within-cluster dispersion:

$$\text{CAL} = \frac{\text{SS}_B / (C - 1)}{\text{SS}_W / (N - C)}, \quad (39)$$

where SS_B and SS_W are between- and within-cluster sums of squares and C is the number of clusters. Larger CAL implies tighter and more separated clusters.

Davies-Bouldin index (DAV). This averages similarity between each cluster and its most similar neighbor:

$$\text{DAV} = \frac{1}{C} \sum_{i=1}^C \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right), \quad (40)$$

where S_i is the average intra-cluster distance and M_{ij} is the distance between cluster centroids. Lower values are better.

Latent coupling (COR). We compute the mean absolute Pearson correlation of latent dimensions:

$$\text{COR} = \frac{1}{d_z(d_z - 1)} \sum_{i < j} |\text{corr}(\mathbf{z}_i, \mathbf{z}_j)|, \quad (41)$$

where $\mathbf{z}_i$ is the i -th latent dimension across cells. Higher COR indicates stronger coordinated variation among latent axes, consistent with modular biological programs.

iAODE software and visualization ecosystem for training, data browsing, and continuum exploration

To reduce the barrier to entry and ensure reproducibility, we build an ecosystem around iAODE that integrates model training, data management, and continuum evaluation. The core implementation is in Python 3.9+ and PyTorch, packaged as an open-source library iAODE (<https://github.com/PeterPonyu/iAODE>) and archived on Zenodo (<https://doi.org/10.5281/zenodo.18453104>). We provide a static web frontend hosted on GitHub Pages (<https://peterponyu.github.io/iAODE>) for browsing standardized datasets and interactively inspecting continuity metrics.

In a Python/Scanpy environment, users can pass preprocessed AnnData objects directly to iAODE, configure latent dimension, loss weights, and optimization parameters, launch training from the command line or notebooks, and access latent embeddings, pseudotime, and continuity metrics via a unified API. Alternatively, they can run the local FastAPI server and use a browser GUI to upload data, set hyperparameters, and export results.

The frontend has two main pages. The *Dataset Browser* shows metadata of standardized AnnData resources (cell/feature counts, species, platform, batches, etc.) and supports filtering and subsampling. The *Continuity Explorer* focuses on representative simulated datasets and interactively visualizes latent embeddings along with continuity metrics such as manifold dimensionality, spectral decay, trajectory directionality, and noise resilience, providing geometric intuition for numerical results. To facilitate reproduction of all experiments, the GitHub repository includes example scripts in `notebooks/` and `examples/` for both ATAC and RNA modalities, covering raw-matrix loading, TF-IDF normalization, highly variable (HV) peak/gene selection, iAODE training, trajectory inference, metric computation, and visualization.

Statistics and reproducibility

Unless noted otherwise, cross-dataset comparisons use dataset-level aggregation: each dataset yields one summary metric and is treated as an independent observation. Within a given size category (Small, Medium, Large) or modality (scATAC-seq, scRNA-seq), we use Friedman tests or repeated-measures ANOVA (RM-ANOVA) to assess global differences across multiple models, components, or hyperparameter configurations. When comparing three or more settings and normality is doubtful, we use Friedman tests; for smaller sample sizes with approximate homoscedasticity we use RM-ANOVA. If the global test is significant (typically $p < 0.001$), we perform pairwise comparisons: for each dataset, we compare

metrics between two configurations using paired Wilcoxon signed-rank tests or paired *t*-tests for continuous variables; we apply Bonferroni correction to all pairwise *p*-values in multi-group comparisons.

For interpretability, we report “advantage scores” in summary tables, e.g., $\Delta = \text{Full} - \text{Variant}$ for component analysis, representing the difference between full iAODE and a given variant on a given metric. In hyperparameter analyses, we use “reference configuration - alternative configuration” to summarize relative gains. For metrics where lower is better (e.g., DAV), signs are interpreted accordingly; for others, positive Δ indicates the first configuration is superior. For simulated data, we assess linear relationships between the user-specified continuum setting and metric scores via simple linear regression, reporting Pearson *R*, slope, R^2 , and two-sided *p*-values. All statistical analyses are conducted in Python using `scipy` and `numpy`, with visualizations produced via `matplotlib` and `seaborn`.

Hardware and computational-resource evaluation

All experiments are run on a single-GPU workstation equipped with an NVIDIA GeForce RTX 5090 Laptop GPU (24 GB VRAM), a 24-core CPU, and 64 GB RAM. To ensure fair comparisons across models and settings, we fix the number of mini-batches per epoch across data scales by adjusting batch sizes: 128, 256, and 512 for Small, Medium, and Large datasets, respectively, yielding approximately 27 mini-batches per epoch per dataset. The maximum number of epochs is 400, with validation every 5 epochs. Early stopping with a patience of 25 validation checks is used: if validation loss does not decrease across 25 consecutive evaluations, training stops and parameters are rolled back to the epoch with the minimum validation loss.

Data availability

All datasets analyzed in this study are publicly available from existing repositories as described in the Methods and in the Supplementary Information. Numerical results and summary statistics supporting the figures and plots are provided in the Supplementary Information (Supplementary Tables 1–13). The Supplementary Information is also available on Figshare (<https://doi.org/10.6084/m9.figshare.31225099>)⁵⁰.

Code availability

The iAODE source code is available at <https://github.com/PeterPonyu/iAODE>. The exact version used in this study is archived on Zenodo (<https://doi.org/10.5281/zenodo.18453104>)⁵¹.

Received: 11 July 2025; Accepted: 18 February 2026;

Published online: 03 March 2026

References

- Li, Z. et al. Chromatin-accessibility estimation from single-cell ATAC-seq data with scOpen. *Nat. Commun.* **12**, 6386 (2021).
- Rachid, Z. S. MOCHA’s advanced statistical modeling of scATAC-seq data enables functional genomic inference in large human cohorts. *Nat. Commun.* **15**, 528 (2024).
- Xiong, L. et al. SCALE method for single-cell ATAC-seq analysis via latent feature extraction. *Nat. Commun.* **10**, 4576 (2019).
- Ashuach, T., Reidenbach, D. A., Gayoso, A. & Yosef, N. PeakVI: A deep generative model for single-cell chromatin accessibility analysis. *Cell Rep. Methods* **2**, 100182 (2022).
- Martens, L. D., Fischer, D. S., Yépez, V. A., Theis, F. J. & Gagneur, J. Modeling fragment counts improves single-cell ATAC-seq analysis. *Nat. Methods* **21**, 28–31 (2024).
- Cao, Y., Jia, L., Wang, L. & Zhang, J. SAILER: Scalable and accurate invariant representation learning for single-cell ATAC-seq processing and integration. *Bioinformatics* **37**, i317–i327 (2021).
- Li, G. et al. A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. *Genome Biol.* **23**, 20 (2022).
- Fan, Y., Li, Y., Ding, J. & Li, Y. GFETM: Genome Foundation-Based Embedded Topic Model for scATAC-seq Modeling. In: *Research in Computational Molecular Biology*. RECOMB 2024. Lecture Notes in Computer Science. (ed. Ma, J.) vol 14758, pp. 314–319 (Springer, Cham, 2024).
- Zheng, S. C. et al. scDiffusion: conditional generation of high-quality single-cell data using diffusion model. *Bioinformatics* **40**, btac518 (2024).
- Cao, Z. J. et al. scButterfly: a versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. *Nat. Commun.* **15**, 2885 (2024).
- Zhang, K. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021).
- Baek, S., Song, K. & Lee, I. Single-cell foundation models: bringing artificial intelligence into cell biology. *Exp. Mol. Med.* **56**, 2169–2181 (2024).
- Gayoso, A. et al. A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022).
- Brombacher, E., Hackenberg, M. & Kreutz, C. The performance of deep generative models for learning joint embeddings of single-cell multi-omics data. *Front. Mol. Biosci.* **9**, 962644 (2022).
- Buenrostro, J. D. et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell* **173**, 1535–1548.e16 (2018).
- Trevino, A. E. et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053–5069.e23 (2021).
- Song, Q. & Su, J. SMGR: a joint statistical method for integrative analysis of single-cell multi-omics data. *NAR Genom. Bioinform.* **4**, lqac056 (2022).
- Luo, S., Germain, P. L., Robinson, M. D. & von Meyenn, F. Benchmarking computational methods for single-cell chromatin data analysis. *Genome Biol.* **25**, 225 (2024).
- Chen, H. et al. Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol.* **20**, 241 (2019).
- Tian, T., Wan, J., Song, Q. & Wei, Z. Complex hierarchical structures in single-cell genomics data unveiled by deep hyperbolic manifold learning. *Genome Res.* **33**, 821–835 (2023).
- Ahlmann-Eltze, C. & Huber, W. Analysis of multi-condition single-cell data with latent embedding multivariate regression. *Nat. Methods* **21**, 659–667 (2024).
- Sidarta-Oliveira, D., Jara, C. P., Ferruzzi, A. J., Skaf, M. S. & Velloso, L. A. TopOMetry systematically learns and evaluates the latent space of high-dimensional data using topology and deep learning. *eLife* **13**, RP100361 (2024).
- Smolander, J., Junttila, S., Venäläinen, M. S. & Elo, L. L. Cell-connectivity-guided trajectory inference from single-cell data. *Bioinformatics* **39**, btad515 (2023).
- Van den Berge, K. et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
- Shi, Y. et al. scCRT: a contrastive-based dimensionality reduction model for single-cell RNA-seq data clustering and pseudo-time trajectory inference. *Brief. Bioinform.* **25**, bbae204 (2024).
- Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
- Du, J. H., Chen, T., Gao, M. & Wang, J. Joint trajectory inference for single-cell genomics using deep learning with a mixture prior. *Proc. Natl. Acad. Sci. USA* **121**, e2316256121 (2024).
- Xiao, C. et al. Benchmarking multi-omics integration algorithms across cell-type identification, trajectory inference and gene regulatory network reconstruction. *Brief. Bioinform.* **25**, bbae095 (2024).
- Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
- Chen, Z., King, W. C., Hwang, A., Gerstein, M. & Zhang, J. DeepVelo: Single-cell transcriptomic deep velocity field learning

- with neural ordinary differential equations. *Sci. Adv.* **8**, eabq3745 (2022).
31. Erbe, R., Stein-O'Brien, G. & Fertig, E. J. Transcriptomic forecasting with neural ordinary differential equations. *Patterns* **4**, 100793 (2023).
 32. Zhang, J., Xu, C. & Chen, Z. scNODE: generative model for temporal single cell transcriptomic data prediction. *Bioinformatics* **40**, ii146–ii154 (2024).
 33. Zhang, K., Zhu, J., Kong, D. & Zhang, Z. Modeling single cell trajectory using forward-backward stochastic differential equations. *PLoS Comput. Biol.* **20**, e1012015 (2024).
 34. Gao, Y., Huang, X., Chen, A., Sharma, A. & Zhang, L. Causal disentanglement for single-cell representations and controllable counterfactual generation. *Nat. Commun.* **16**, 6775 (2025).
 35. Yu, H. & Welch, J. D. MichiGAN: sampling from disentangled representations of single-cell data using generative adversarial networks. *Genome Biol.* **22**, 158 (2021).
 36. Alemi, A.A., Fischer, I., Dillon, J.V., Murphy, K. Deep variational information bottleneck. In: International Conference on Learning Representations (ICLR); 2017 Apr 24–26; Toulon, France.
 37. Piran, Z., Cohen, N., Hoshen, Y. & Nitzan, M. Disentanglement of single-cell data with biolord. *Nat. Biotechnol.* **42**, 1678–1683 (2024).
 38. Pan, W., Long, F. & Pan, J. ScInfoVAE: interpretable dimensional reduction of single cell transcription data with variational autoencoders and extended mutual information regularization. *BioData Min.* **16**, 20 (2023).
 39. Majima, K. et al. LineageVAE: reconstructing historical cell states and transcriptomes toward unobserved progenitors. *Bioinformatics* **40**, btae520 (2024).
 40. Higgins, I. et al. beta-VAE: Learning basic visual concepts with a constrained variational framework. In: 5th International Conference on Learning Representations (ICLR); 2017 Apr 24–26; Toulon, France.
 41. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. **19**, 477 (2018).
 42. Chen, RTQ, Rubanova, Y., Bettencourt, J., Duvenaud, D. Neural ordinary differential equations. In: Advances in Neural Information Processing Systems (NeurIPS); 2018 Dec 3–8; Montréal, Canada. p. 6571–6583.
 43. Meeussen, J. V. W. & Lenstra, T. L. Time will tell: comparing timescales to gain insight into transcriptional bursting. *Trends Genet.* **40**, 160–174 (2024).
 44. Kartha, V. K. et al. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom.* **2**, 100166 (2022).
 45. Otto, D. J., Jordan, C., Dury, B., Dien, V. & Setty, M. Quantifying cell-state densities in single-cell phenotypic landscapes using Mellon. *Nat. Methods* **21**, 1185–1195 (2024).
 46. Li, C., Virgilio, M. C., Collins, K. L. & Welch, J. D. Multi-modal single-cell velocity inference with MultiVelo. *Nature* **618**, 377–385 (2023).
 47. Kumar, A., Sattigeri, P., Balakrishnan, A. Variational inference of disentangled latent concepts from unlabeled observations. In: International Conference on Learning Representations (ICLR); 2018.
 48. Chen, RTQ, Li, X., Grosse, R.B., Duvenaud, D.K. Isolating sources of disentanglement in variational autoencoders. In: Advances in Neural Information Processing Systems (NeurIPS); 2018. p. 2610–2620.
 49. Zhao, S., Song, J., Ermon, S. InfoVAE: Balancing learning and inference in variational autoencoders. In: AAAI Conference on Artificial Intelligence; 2019. p. 5885–5892.
 50. Fu, Z. Supplementary information for iAODE manuscript. Figshare <https://doi.org/10.6084/m9.figshare.31225099> (2026).
 51. Fu, Z. iAODE: Interpretable Accessibility ODE VAE for scATAC-seq (v0.3.0). Zenodo. <https://doi.org/10.5281/zenodo.18453104> (2026).

Acknowledgements

This research was funded by the National Natural Science Foundation of China (grant numbers 82222060, 82430103, 82473572, 81930090, 81725019, 82073487, 81602790).

Author contributions

Z.F.: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, resources, writing—original draft, writing—review and editing, visualization. C.C.: Formal analysis, data curation, visualization, writing—original draft, writing—review and editing. S.W.: Resources, supervision, project administration. J.W.: Resources, supervision, project administration, funding acquisition. S.C.: Writing—review and editing, resources, supervision, project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-026-09768-8>.

Correspondence and requests for materials should be addressed to Zeyu Fu, Junping Wang or Shilei Chen.

Peer review information *Communications Biology* thanks Yuge Wang and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Xiangtao Li and Mengtan Xing. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026