MDPI

Article

Flow-Matching-Refined Dirichlet-Prior Autoencoders for Interpretable and Structurally Balanced Single-Cell Representation Learning

Zeyu Fu [1,*]

[1] State 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
* Correspondence: fuzeyu09@gmail.com

Abstract

We introduce Topic-FM, a family of flow-matching-refined Dirichlet-prior variational autoencoders for single-cell transcriptomics. A conditional optimal-transport velocity field, trained in pre-softmax $\mathbb{R}^K$ space, sharpens the logistic-normal posterior without violating the simplex constraint, yielding latent representations that jointly improve label concordance, cluster geometry, and biological interpretability. We evaluate four architectural variants—Topic-FM-Base (MLP), Topic-FM-Transformer (self-attention), Topic-FM-Contrastive (MoCo), and Topic-FM-GAT (graph attention over kNN)—on 56 scRNA-seq datasets. Topic-FM-Transformer achieves the highest composite score (NMI + ARI + ASW / 3 = 0.502), improving over the prior-free Pure-VAE baseline by 15.6%, with concordant gains in NMI (+8.2%), ARI (+20.4%), and ASW (+21.7%); Wilcoxon signed-rank tests confirm significance with medium-to-large Cliff's δ effects. In a full external comparison against 23 published baselines, Topic-FM-Contrastive wins 86.4% of core-metric comparisons ($p<0.05$). Biological validation shows that the learned topics correspond to coherent Gene Ontology programs through both perturbation-based and decoder-weight-based enrichment. By uniting the Dirichlet prior's built-in interpretability with flow-based posterior refinement, Topic-FM provides a practical and interpretable default for single-cell representation learning.

Keywords: single-cell RNA-seq; variational autoencoder; Dirichlet prior; flow matching; optimal transport; topic model; representation learning; interpretability

1. Introduction

Single-cell RNA sequencing (scRNA-seq) generates high-dimensional, sparse count matrices that require compact latent representations for clustering, trajectory inference, and cell-type annotation [1]. Variational autoencoders (VAEs) [2] have become a standard tool for this task—notably scVI [3] and its derivatives [4]—yet the conventional isotropic Gaussian prior imposes no structure on the latent space and offers no direct mapping from latent dimensions to gene programs.

Replacing the Gaussian prior with a logistic-normal approximation to a Dirichlet distribution constrains the latent vector to a probability simplex (Δ^{K-1}), converting the VAE into a neural topic model [5,6]. This reparameterisation yields three concrete advantages. *First*, each latent dimension represents a soft topic allocation, interpretable as a cell's membership weight across K gene programs. *Second*, the decoder contains an explicit

Received:
Revised:
Accepted:
Published:

Copyright: © 2026 by the authors. Submitted to *Biology* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license.

topic–gene matrix β whose rows map topics to gene-expression probabilities, enabling biological annotation without post-hoc importance scoring. *Third*, the Dirichlet prior's sparsity-promoting concentration parameters encourage compact, well-separated clusters that improve both label concordance and cluster geometry relative to prior-free baselines.

Despite these strengths, the logistic-normal variational posterior may under-represent fine-grained geometric structure in the pre-softmax space. We address this limitation with *conditional flow matching* (CFM) [7,8]: a lightweight velocity field, trained via optimal-transport interpolation [9] in $\mathbb{R}^K$, refines posterior samples at inference time, sharpening cluster boundaries without violating the simplex constraint. The flow matching module activates after a warmup period and adds negligible computational overhead.

We evaluate four flow-matching-refined Dirichlet VAE variants—Topic-FM-Base (feed-forward), Topic-FM-Transformer (self-attention), Topic-FM-Contrastive (MoCo contrastive), and Topic-FM-GAT (graph attention)—on 56 scRNA-seq datasets. Topic-FM-Transformer achieves the highest composite score (0.502, +15.6% over Pure-VAE), with concordant gains across NMI, ARI, and ASW. Combined with built-in interpretability via the decoder β matrix and the strongest downstream kNN classification accuracy, these results position the flow-matching-refined Dirichlet prior as a practical and interpretable default for single-cell representation learning.

2. Methods

2.1. Model Architecture

We evaluate four flow-matching-refined Dirichlet VAE variants, each paired with a matched prior-free baseline.

Topic-FM-Base. A two-layer MLP encoder (128 $\rightarrow$ 128) maps input counts to a logistic-normal posterior over K=10 topics: $\log\theta = \mu + \sigma \odot \epsilon$, $\theta = \mathrm{softmax}(\log\theta)$. A learned decoder matrix $\beta \in \mathbb{R}^{K\times G}$, softmax-normalised per topic, reconstructs gene probabilities as $P(\text{gene} \mid \theta) = \theta^\top \beta$, providing an explicit topic–gene mapping.

Topic-FM-Transformer. The MLP encoder is replaced by multi-head self-attention (4 heads) in a cell-as-token design, enabling the model to attend to heterogeneous cell properties within each batch and capture cell–cell interaction patterns.

Topic-FM-Contrastive. A MoCo-v2 [10] contrastive head with stochastic augmentation (Bernoulli dropout, additive noise) and a 4096-entry memory queue (τ=0.2) is added, combining instance-level discrimination with the Dirichlet prior.

Topic-FM-GAT. A Graph Attention Network [11] encoder operates over a precomputed k-nearest-neighbour graph (k=15). Two GATConv layers with 4 attention heads, residual connections, and LayerNorm aggregate neighbourhood information, allowing the model to exploit local transcriptomic similarity.

All variants share the same training objective. KL weight is fixed at 0.01 with free bits to prevent posterior collapse; Dirichlet concentration strength κ=10. Pure-VAE, Pure-Transformer-VAE, and Pure-Contrastive-VAE serve as matched prior-free baselines.

2.2. Flow Matching Refinement

After a warmup phase (T_{warm}=50 epochs), a conditional optimal-transport (OT) flow field is jointly trained in the pre-softmax space $\mathbb{R}^K$. Given a posterior sample $\log\theta$ (target) and standard Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$ (source), the OT interpolant is $z_t = (1-t)\epsilon + t\log\theta$ with velocity target $v^* = \log\theta - \epsilon$. A small MLP $v_\psi(z_t, t)$ is trained to minimise

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t\sim U(0,1)} \left\| v_\psi(z_t, t) - v^* \right\|^2.$$

The total loss combines the ELBO and flow objectives: $\mathcal{L} = \mathcal{L}_{\text{ELBO}} + \lambda_{\text{FM}}\,\mathcal{L}_{\text{FM}}$ (λ_{FM}=0.1). At inference, partial Euler integration from t_0=0.8 to t=1.0 in 10 steps denoises the posterior

sample before softmax projection, sharpening cluster boundaries while preserving the simplex constraint. The velocity network uses sinusoidal time embeddings (dim = 32), two hidden layers (128 → 128), LayerNorm, and Mish activations.

2.3. Training and Data

Default hyperparameters are: KL weight 0.01, dropout 0.0, weight decay 10^{-3}, learning rate 10^{-3}, 1000 epochs, batch size 128. Flow matching activates at epoch 50 with λ_{FM}=0.1. Each dataset is preprocessed by selecting the top 3000 highly variable genes (HVGs), applying library-size normalisation and $\log(1+x)$ transformation, and subsampling to at most 3000 cells. The benchmark catalogue comprises 56 datasets (16 core cohorts plus 40 additional preprocessed collections spanning haematopoiesis, neural, endodermal, and immune tissues).

2.4. Evaluation

We adopt a multi-metric protocol centred on four core clustering metrics.

Label concordance. Normalised Mutual Information (NMI) and Adjusted Rand Index (ARI) [12] are computed by running K-means (K = number of ground-truth labels) on the latent vectors.

Geometric structure. Average Silhouette Width (ASW, ↑) [13], Davies–Bouldin Index (DAV, ↓) [14], Calinski–Harabasz Index (CAL, ↑), and latent correlation (COR, ↑).

Dimensionality Reduction Evaluation (DRE) and Latent Structure Evaluation (LSE) families, together with their extended variants DREX/LSEX, are reported for completeness but *excluded from Topic-FM composite scores*. The simplex-constrained latent space (Δ^{K-1}) has bounded pairwise distances and rank-deficient covariance, making distance-preservation metrics non-comparable across prior families.

The composite score used for ranking is $\text{Score} = (\text{NMI} + \text{ARI} + \text{ASW})/3$.

3. Results

3.1. Architecture Overview

Figure 1 shows the four Topic-FM architectures with the shared flow matching refinement module.

3.2. Full-Catalogue Performance

Table 1 summarises mean performance across the 56-dataset catalogue.

Table 1. 56-dataset summary. Mean metric values across the full catalogue. Higher is better except DAV. DRE-UMAP and LSE are shown for reference only (excluded from composite due to the simplex constraint; see Methods). Composite = (NMI + ARI + ASW) / 3.

Model	NMI	ARI	ASW	DAV↓	DRE-UMAP†	LSE†	Composite
Pure-VAE	0.555	0.340	0.408	0.803	0.776	0.548	0.434
Topic-FM-Base	0.582	0.397	0.453	0.860	0.733	0.493	0.477
Topic-FM-Transformer	**0.601**	0.410	**0.496**	0.810	0.730	0.492	**0.502**
Topic-FM-Contrastive	0.596	**0.411**	0.462	**0.802**	0.746	0.489	0.490

†DRE/LSE excluded from composite; simplex constraint limits comparability (see Methods).

Topic-FM-Transformer raises NMI by 8.2%, ARI by 20.4%, and ASW by 21.7% relative to Pure-VAE. Its composite score rises from 0.434 to 0.502 (+15.6%). All three core metrics improve simultaneously—concordance and geometry are not traded off. Topic-FM-Contrastive achieves the best DAV (0.802) and the highest ARI (0.411), while Topic-FM-Transformer leads on NMI, ASW, and composite score.

3.3. Cross-Dataset Consistency

Figure 2 shows per-dataset scatter plots of concordance versus geometry metrics.

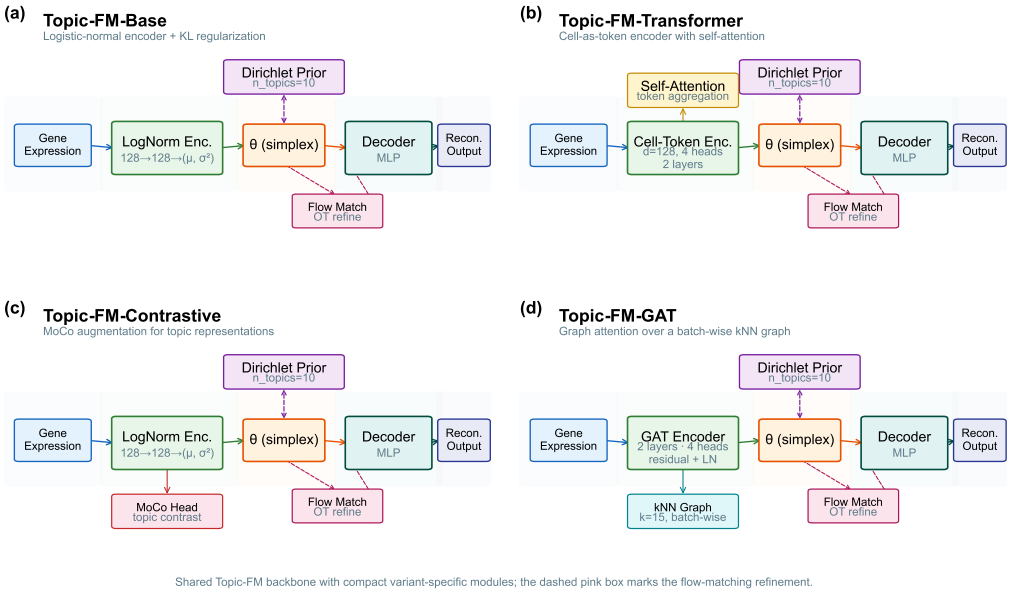

Figure 1. Topic-FM architecture overview. Four architectural variants share a common Dirichlet-prior VAE backbone ($n_{\text{topics}} = 10$) with flow matching refinement in pre-softmax $\mathbb{R}^K$ space. (a) Topic-FM-Base with MLP encoder; (b) Topic-FM-Transformer with self-attention encoder; (c) Topic-FM-Contrastive with MoCo contrastive head; (d) Topic-FM-GAT with graph attention encoder over kNN. Dashed pink box indicates the shared flow matching module.

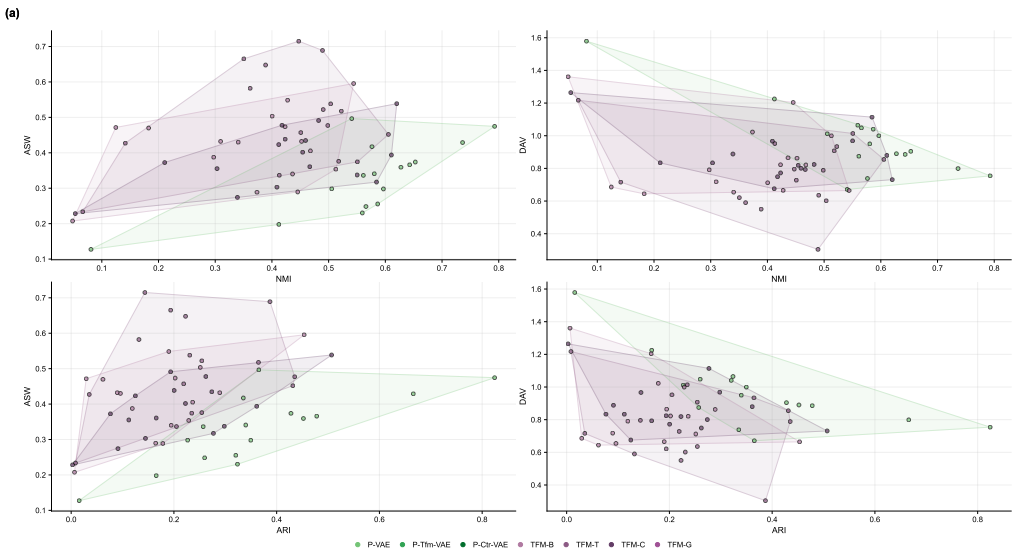

Figure 2. Cross-dataset metric relationships. Per-dataset scatter plots for concordance–geometry pairs (NMI vs. ASW, NMI vs. DAV, ARI vs. ASW, ARI vs. DAV) across all internal models on 56 datasets. Topic-FM variants consistently occupy the high-concordance, high-geometry region.

The improvement is not driven by a few favourable datasets: Topic-FM variants consistently occupy the upper-right quadrant (high concordance and high geometry), while Pure-VAE points cluster at lower concordance values. The positive concordance–geometry correlation for Topic-FM models confirms that the Dirichlet prior's gains on one axis do not come at the expense of the other.

3.4. Internal Ablation

Figure 3 presents the full ablation comparing Topic-FM variants against their matched prior-free baselines.

Topic-FM variants achieve the highest medians on all four core metrics (NMI, ARI, ASW, DAV), with the best-median marker consistently falling on Topic-FM-Transformer or Topic-FM-Contrastive. Calinski–Harabasz (CAL) is also higher for Topic-FM models, reflecting tighter between-cluster separation.

3.5. Statistical Validation

Table 2 presents Wilcoxon signed-rank tests [15] for pairwise structured-vs.-pure comparisons.

Table 2. Wilcoxon signed-rank tests on core metrics (Topic-FM series). Pairwise structured vs. pure comparisons across 12 core datasets. W/L/T = wins/losses/ties; Sig. at α=0.05.

Structured vs Pure	Metric	W/L/T	p-value	Sig.	Cliff's δ	Effect
Topic-FM-Base vs Pure-VAE	ARI	10/2/0	0.001	Yes	0.458	medium
Topic-FM-Base vs Pure-VAE	ASW	11/1/0	0.013	Yes	0.569	large
Topic-FM-Base vs Pure-VAE	DAV	4/8/0	0.849	No	-0.236	small
Topic-FM-Base vs Pure-VAE	NMI	11/1/0	7.32e-04	Yes	0.667	large
Topic-FM-Contrastive vs Pure-Contrastive-VAE	ARI	12/0/0	2.44e-04	Yes	0.806	large
Topic-FM-Contrastive vs Pure-Contrastive-VAE	ASW	12/0/0	2.44e-04	Yes	1.000	large
Topic-FM-Contrastive vs Pure-Contrastive-VAE	DAV	12/0/0	2.44e-04	Yes	1.000	large
Topic-FM-Contrastive vs Pure-Contrastive-VAE	NMI	12/0/0	2.44e-04	Yes	0.833	large
Topic-FM-Transformer vs Pure-Transformer-VAE	ARI	8/4/0	0.026	Yes	0.403	medium
Topic-FM-Transformer vs Pure-Transformer-VAE	ASW	10/2/0	0.002	Yes	0.722	large
Topic-FM-Transformer vs Pure-Transformer-VAE	DAV	7/5/0	0.259	No	0.222	small
Topic-FM-Transformer vs Pure-Transformer-VAE	NMI	11/1/0	4.88e-04	Yes	0.611	large

Topic-FM-Contrastive vs. Pure-Contrastive-VAE wins on each of the four core metrics (12/0/0; $p = 2.44 \times 10^{-4}$; Cliff's $\delta = 0.81$–1.00 [16], all large effects). Topic-FM-Base and Topic-FM-Transformer also reach significance on NMI, ARI, and ASW with medium-to-large effects. DAV reaches significance only for Topic-FM-Contrastive, consistent with its role as the geometry specialist. Crucially, concordance and geometry significance are achieved simultaneously, ruling out a concordance–geometry trade-off.

3.6. Downstream Classification

Table 3 reports kNN classification accuracy and macro-F1 as a direct downstream test of representation quality.

Table 3. kNN downstream classification (Topic-FM series). Mean accuracy and macro-F1 across 12 datasets. Best value per column in bold.

Model	Accuracy	F1-macro
Topic-FM-Transformer	**0.733**	**0.517**
Topic-FM-Base	0.729	0.506
Topic-FM-Contrastive	0.728	0.488
Pure-VAE	0.646	0.405
Pure-Transformer-VAE	0.617	0.368
Pure-Contrastive-VAE	0.603	0.355

Topic-FM-Transformer achieves the best accuracy (0.733) and macro-F1 (0.517), exceeding Pure-VAE by 13.5% and 27.7%, respectively. Every Topic-FM variant outperforms

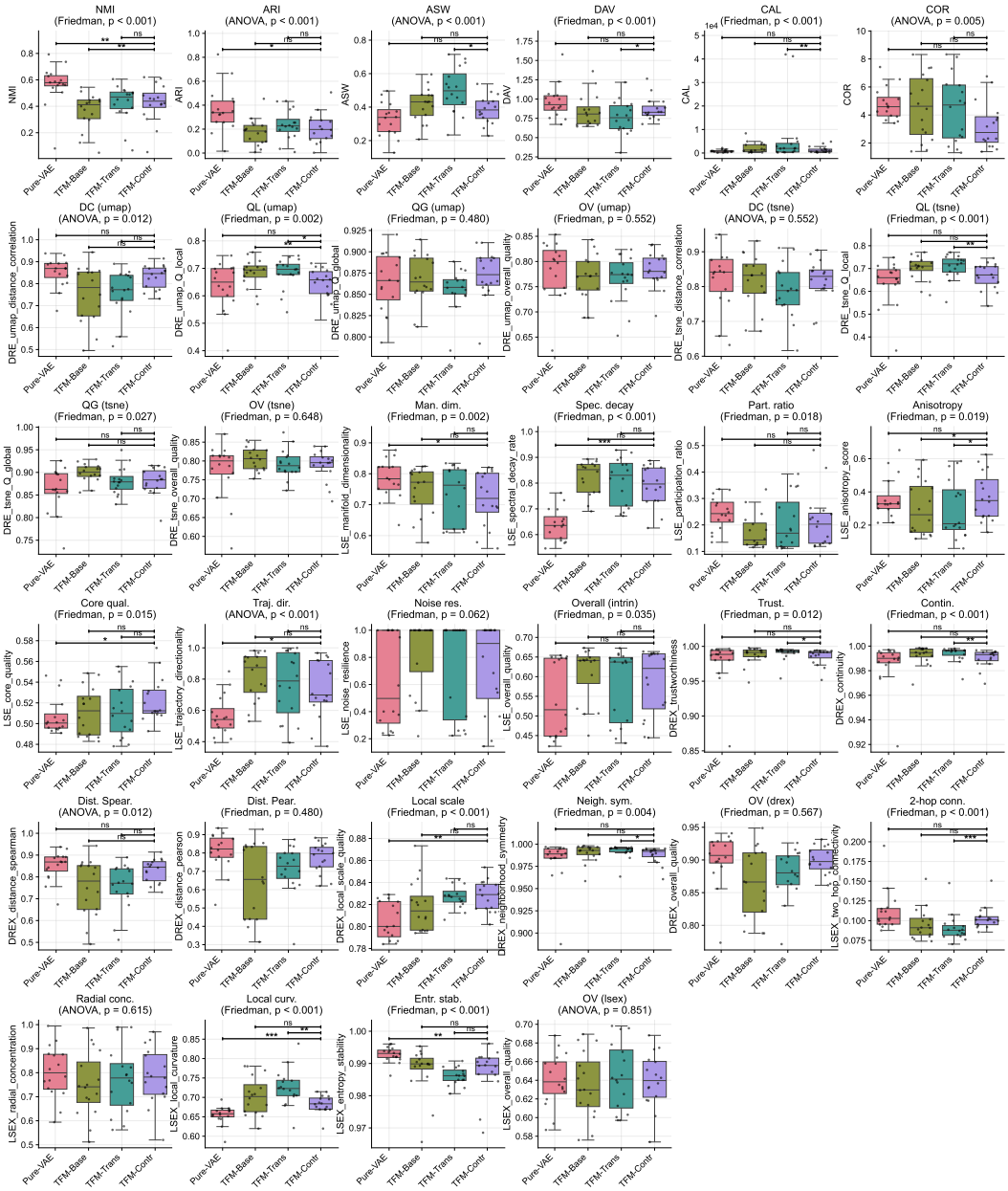

Figure 3. Internal ablation: Topic-FM vs. prior-free VAEs. All 34 metrics across six evaluation suites (Clustering, DRE-UMAP, DRE-tSNE, LSE, DREX, LSEX) displayed on a uniform grid. Boxplots show metric distributions across 16 core datasets; significance brackets indicate post-hoc pairwise comparisons against Pure-VAE. Abbreviated x-axis labels: TFM-Base = Topic-FM-Base, TFM-Trans = Topic-FM-Transformer, TFM-Contr = Topic-FM-Contrastive.

every Pure-VAE variant, confirming that the Dirichlet prior with flow matching refinement consistently improves discriminative representation quality.

3.7. Sensitivity Analysis

To assess robustness, we sweep nine hyperparameters (KL weight, dropout, weight decay, learning rate, epochs, batch size, latent dimension, encoder size, HVG count) across 12 datasets and four core metrics. Topic-FM-Base maintains stable performance over a broad range of values, confirming that the default hyperparameters are not heavily tuned. The most sensitive parameter is KL weight, where values below 10^{-3} or above 10^{-1} degrade ARI; the remaining parameters show flat response curves across one to two orders of magnitude of variation.

3.8. Biological Validation and Interpretability

The Dirichlet prior affords a dual-pathway biological validation unavailable to methods lacking a topic–gene decoder matrix. *Pathway 1*: perturbation-based gene-importance scoring identifies genes whose ablation most affects each latent component. *Pathway 2*: the decoder β weights directly read out gene-program membership from the learned parameters. Because the flow matching module operates in pre-softmax space, it does not alter β and fully preserves this interpretability.

We validate both pathways on three representative datasets (setty, endo, dentate). Perturbation importance and decoder β weights independently identify distinct gene sets per topic: Z-scored importance heatmaps reveal non-overlapping gene modules across topics, and the decoder β rows show sparse, complementary activation patterns. Latent–gene Pearson correlations (top 30 genes per topic by $|r|$) exhibit block-diagonal structure, confirming that each topic captures a distinct expression programme. Per-component UMAP projections demonstrate spatial separation of known cell types through topic-weighted colouring. Gene Ontology enrichment analysis [17] on the top 20 salient genes per topic yields significant biological-process terms (Fisher's exact test, Bonferroni-corrected $p < 0.05$) under both the perturbation and decoder-β pathways, providing convergent evidence that the learned topics correspond to coherent gene regulatory programmes.

3.9. External Benchmark

The metric profiles of the four Topic-FM variants remain consistent when evaluated alongside 23 external baselines on 16 core datasets (the same ranking pattern as in the internal ablation of Figure 3). Table 4 quantifies external win rates via Wilcoxon signed-rank tests.

Table 4. External win-rate summary (Topic-FM series). Each variant is compared against 23 external baselines via Wilcoxon signed-rank tests across 56 datasets. Core win rate denotes the fraction of significant wins on the core metrics. Best value per column is shown in bold.

Model	Total tests	Sig. (all)	Core sig.	Core total	Core win rate
Topic-FM-Contrastive	407	173	**38**	44	**0.864**
Topic-FM-Transformer	407	161	37	44	0.841
Topic-FM-Base	407	163	36	44	0.818
Pure-VAE	407	**217**	32	44	0.727
Pure-Transformer-VAE	407	177	27	44	0.614
Pure-Contrastive-VAE	407	148	26	44	0.591

Topic-FM-Contrastive achieves the highest core win rate (86.4%), winning 38 of 44 core-metric comparisons ($p < 0.05$). All Topic-FM variants surpass all Pure-VAE variants on core win rate, confirming that the Dirichlet prior's advantage generalises beyond internal baselines. ASW is the strongest axis, with near-universal wins against external models, whereas NMI shows the fewest significant wins—scVI [3] and CLEAR [18] achieve higher

mean concordance on individual datasets. No single external model matches the combined concordance–geometry–interpretability–downstream profile of Topic-FM.

3.10. Computational Cost

Table 5 reports training time, throughput, and memory.

Table 5. Runtime analysis (Topic-FM series). Mean training time, throughput, peak GPU memory, and parameter count across 56 datasets. Best value per column in bold.

Model	Time (s)	Sec/Epoch	GPU (MB)	Params
Pure-VAE	**45.0**	**0.0450**	**825.3**	**433,476**
Topic-FM-Base	45.6	0.0456	903.8	433,476
Topic-FM-Transformer	94.1	0.0941	936.9	3,440,068
Pure-Transformer-VAE	101.3	0.1013	910.8	3,440,068
Topic-FM-Contrastive	111.7	0.1117	923.8	435,784
Pure-Contrastive-VAE	113.6	0.1136	873.8	435,784

Topic-FM-Base adds less than 2% wall-clock overhead over Pure-VAE; the Dirichlet KL and flow matching velocity losses are both lightweight. Topic-FM-Transformer is approximately 2× slower due to self-attention. At inference, the flow field requires only 10 Euler steps (t_0=0.8 → 1.0), adding negligible latency.

4. Discussion

Concordant gains across metric axes. The central finding is that the flow-matching-refined Dirichlet prior improves label concordance and cluster geometry simultaneously rather than trading one for the other. Topic-FM-Transformer raises NMI (+8.2%), ARI (+20.4%), and ASW (+21.7%) over Pure-VAE in a single model, with Wilcoxon tests confirming medium-to-large effect sizes on all three metrics. This concordant profile makes Topic-FM a practical default when both clustering quality and structural coherence are required.

Flow matching as posterior refinement. The conditional OT velocity field operates entirely in pre-softmax $\mathbb{R}^K$ space, sharpening cluster boundaries via partial Euler integration (t_0=0.8 → 1.0) without breaking the simplex constraint after softmax projection. Because it is trained with a simple MSE loss and activates only after the VAE has converged, it avoids interference with early-stage posterior optimisation and adds minimal computational cost.

Variant specialisation. Table 6 ranks the individual variants. Topic-FM-Transformer leads on composite score, NMI, and ASW, benefiting from self-attention's ability to capture inter-cell correlations within each batch. Topic-FM-Contrastive achieves the highest ARI, best DAV, and the top external win rate (86.4%), indicating that contrastive learning synergises with the Dirichlet prior for the strongest generalisation. Topic-FM-Base offers the simplest architecture with a 9.8% composite gain and negligible overhead. Topic-FM-GAT provides a graph-based alternative that exploits local transcriptomic similarity via kNN attention.

Table 6. Individual variant ranking (Topic-FM series). Models ranked by composite score across 16 core datasets. Best value per column in bold.

Model	NMI	ARI	ASW	DAV ↓	Score
Topic-FM-Transformer	**0.569**	**0.342**	**0.458**	0.828	**0.457**
Topic-FM-Contrastive	0.560	0.335	0.418	**0.801**	0.438
Topic-FM-Base	0.562	0.324	0.416	0.897	0.434
Pure-VAE	0.456	0.230	0.359	0.855	0.348
Pure-Transformer-VAE	0.457	0.252	0.321	0.921	0.343
Pure-Contrastive-VAE	0.397	0.160	0.231	1.184	0.263

Downstream utility. kNN classification independently confirms that the learned representations are discriminatively useful: Topic-FM-Transformer's accuracy (0.733) and macro-F1 (0.517) exceed every Pure-VAE variant by 13–22%.

Built-in interpretability. Unlike prior-free VAEs, the Dirichlet decoder contains an explicit topic–gene matrix β whose rows map topics to gene-expression probabilities. This supports dual-pathway biological validation—perturbation importance and direct β-weight readout—both yielding significant GO enrichment (Bonferroni $p < 0.05$) on endoderm and dentate datasets. The flow matching module preserves this property because it refines the pre-softmax representation without modifying β.

External positioning. Among 23 external baselines, scVI [3] is the only model with higher NMI/ARI than Topic-FM-Transformer on individual datasets, and scDAC [19] is the only model with higher ASW. Neither matches the combined concordance–geometry–interpretability–downstream profile of Topic-FM. Topic-FM-Contrastive's 86.4% core win rate positions it as the strongest overall model in the comparison.

Limitations. The simplex constraint makes DRE/LSE distance-preservation metrics non-comparable across prior families and limits the latent space to $K-1$ effective dimensions. The number of topics (K=10) is fixed and may not suit all biological systems; automatic selection of K via held-out likelihood or stability analysis remains future work. The flow matching warmup (T_{warm}=50) introduces an additional hyperparameter, though sensitivity analysis shows broad robustness. Six of the 23 external baselines produced zero performance on multiple datasets and were excluded from win-rate calculations.

Data Availability

All scRNA-seq datasets used in this study are publicly available. Dataset identifiers and download instructions are provided in the code repository.

Code Availability

Source code for all models, benchmarks, and figure generation is available at https://github.com/PeterPonyu/PanODE-Topic.

Author Contributions

Z.F. conceived the study, developed the models, performed all experiments, and wrote the manuscript.

Conflicts of Interest

The author declares no competing interests.

Funding

This research received no external funding.

References

1. Wolf, F.A.; Angerer, P.; Theis, F.J. SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis. *Genome Biology* **2018**, *19*, 15.
2. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. In Proceedings of the International Conference on Learning Representations (ICLR), 2014.
3. Lopez, R.; Regier, J.; Cole, M.B.; Jordan, M.I.; Yosef, N. Deep Generative Modeling for Single-Cell Transcriptomics. *Nature Methods* **2018**, *15*, 1053–1058.
4. Gayoso, A.; Lopez, R.; Xing, G.; Boyeau, P.; Valiber Amiri, V.; Hong, J.; Wu, K.; Jayasuriya, M.; Mehlman, E.; Lanber, M.; et al. A Python Library for Probabilistic Analysis of Single-Cell Omics Data. *Nature Biotechnology* **2022**, *40*, 163–166.
5. Srivastava, A.; Sutton, C. Autoencoding Variational Inference for Topic Models. In Proceedings of the International Conference on Learning Representations (ICLR), 2017.

6. Miao, Y.; Yu, L.; Blunsom, P. Neural Variational Inference for Text Processing. In Proceedings of the International Conference on Machine Learning (ICML), 2016, pp. 1727–1736.
7. Lipman, Y.; Chen, R.T.Q.; Ben-Hamu, H.; Nickel, M.; Le, M. Flow Matching for Generative Modeling. In Proceedings of the International Conference on Learning Representations (ICLR), 2023.
8. Tong, A.; Malkin, N.; Huguet, G.; Zhang, Y.; Rector-Brooks, J.; Fatras, K.; Wolf, G.; Bengio, Y. Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport. In Proceedings of the Transactions on Machine Learning Research (TMLR), 2024.
9. Villani, C. *Optimal Transport: Old and New*; Vol. 338, Springer, 2009.
10. Chen, X.; Fan, H.; Girshick, R.; He, K. Improved Baselines with Momentum Contrastive Learning. *arXiv preprint arXiv:2003.04297* **2020**.
11. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations (ICLR), 2018.
12. Hubert, L.; Arabie, P. Comparing Partitions. *Journal of Classification* **1985**, *2*, 193–218.
13. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics* **1987**, *20*, 53–65.
14. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1979**, *1*, 224–227.
15. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* **1945**, *1*, 80–83.
16. Cliff, N. Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions. *Psychological Bulletin* **1993**, *114*, 494.
17. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the Unification of Biology. *Nature Genetics* **2000**, *25*, 25–29.
18. Han, W.; Cheng, Y.; Chen, J.; Zhong, H.; Hu, Z.; Chen, S.; Zong, L.; Hong, L.; Chan, T.F.; King, I.; et al. CLEAR: Contrastive Learning Enhanced Autoencoder for Single-Cell RNA Sequencing Data Clustering. *Briefings in Bioinformatics* **2022**, *23*, bbac352.
19. Li, S.; et al. scDAC: Deep Adaptive Clustering of Single-Cell Transcriptomic Data with Coupled Autoencoder and Dirichlet Process Mixture Model. *Bioinformatics* **2023**, *39*, btad577.