SPLICEDVAE: LEARNING SPLICING RATIOS FROM SCRNA-SEQ TO ENHANCE RNA VELOCITY AND CELLULAR TRAJECTORIES

Sahil Gupta, Andrew Ni & Jiaqi Zhang
Department of Computer Science
Brown University
Providence, RI 02912, USA
{sahil_gupta,andrew_ni,jiaqi_zhang6}@brown.edu

ABSTRACT

Single-cell RNA sequencing (scRNA-seq) provides high-resolution snapshots of cellular states, yet inferring temporal dynamics such as RNA velocity and differentiation trajectories remains challenging. To address this, we present SplicedVAE, a supervised generative framework that augments the scVI variational autoencoder with a dedicated decoder for predicting per-gene splicing ratios only from raw counts. SplicedVAE jointly optimizes gene expression reconstruction and splicing-ratio prediction, enabling biologically informed regularization. Our model achieves improved splicing-ratio prediction accuracy (RMSE 0.1271, Pearson r = 0.67), enhanced latent representations, and substantially more coherent velocity fields compared to standard scVI. When reconstructed S/U counts are passed into scVelo, SplicedVAE recapitulates developmental flow patterns and yields high cosine similarity to ground-truth velocities. These results demonstrate multi-task learning can improve velocity-based trajectory reconstruction and establishes a foundation for future models capable of generating cellular trajectories.

1 INTRODUCTION

Understanding cellular dynamics from static snapshots remains a fundamental challenge in single-cell biology. Traditional approaches such as scVelo (Bergen et al., 2020) attempt to infer RNA velocity (the rate and direction of transcriptional change) by modeling the relationship between spliced and unspliced (S/U) mRNA counts (La Manno et al., 2018). However, most widely-used scRNA-seq protocols (particularly 3'-tagging methods) do not capture S/U information, severely limiting applicability. Recent computational approaches have attempted to predict velocity without S/U counts (Zeng et al., 2022; Mahajan & Maslov, 2024), but their performance remains limited by assumptions of cellular kinetics and producing discrete rather than continuous predictions.

Beyond unreliable velocity prediction, existing methods are also unable to model cellular trajectories. Supervised approaches like Velo-Predictor (Wang & Zheng, 2021) and TFvelo (Li et al., 2024) predict velocity direction but lack the generative structure to model cell state transitions. Meanwhile, generative models like diffusion-based methods (Luo et al., 2024; Ho et al., 2020) and flow-matching (Klein et al., 2025) have shown promise for capturing complex scRNA-seq distributions, but they have primarily been applied to generate static cell profiles rather than model temporal dynamics.

In response to these limitations, we propose SplicedVAE. Our approach extends the scVI variational autoencoder (Lopez et al., 2018) by adding a secondary task: predicting the ratio of spliced to total mRNA per gene. The intuition is if we train a model to both reconstruct gene expression counts (the standard scVI objective) and predict splicing ratios (our auxiliary objective), the model's internal representation should encode information about cellular dynamics that expression alone miss. This represents a key step toward unified generative frameworks capable of *generating* novel cellular trajectories, enabling in silico perturbation experiments and prospective experimental design.

2 METHODS

2.1 MODEL ARCHITECTURE

The full SplicedVAE architecture includes: (1) an scVI encoder capturing nonlinear gene–gene dependencies, (2) a negative binomial (NB) decoder reconstructing the gene-expression likelihood, and (3) a custom MLP decoder head to predict splicing ratios. The joint training objective combines the evidence lower bound (ELBO) for reconstruction with a weighted splicing prediction loss:

$$\mathcal{L} = \mathcal{L}_{\text{ELBO}} + \lambda \mathcal{L}_{\text{splice}}$$

where λ controls the contribution of the auxiliary task. Multiple loss formulations were evaluated for $\mathcal{L}_{\text{splice}}$, including MSE, weighted MSE (gene-specific weighting by total counts), binomial likelihood, and L1 loss. MSE was selected as it yielded the lowest validation RMSE and $\lambda = 1$ was chosen to balance the two objectives. This multitask formulation encourages the latent representation to encode information relevant to both gene expression and splicing dynamics.

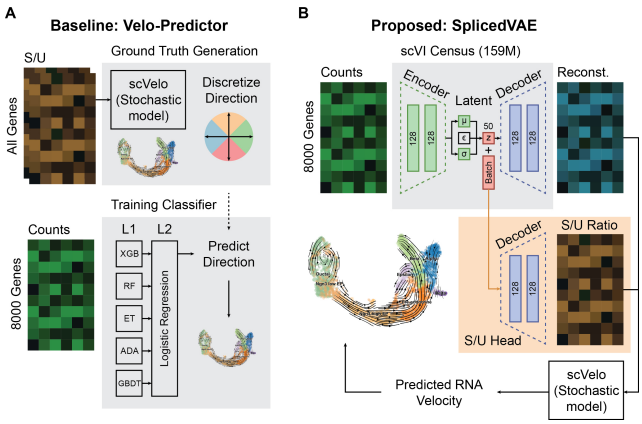

Figure 1: (A) Velo-Predictor baseline. Genes with S/U counts are preprocessed with the stochastic model from scVelo and discretized into 4 classes as ground truth labels. An ensemble of models uses raw counts followed by logistic regression to predict discretized UMAP velocity directions. (B) Architecture of our proposed SplicedVAE. Based on scVI, an S/U prediction head is added as a separate decoder for predicting splicing ratios. The stochastic scVelo model is applied to compute RNA velocity by multiplying predicted S/U ratios with the reconstructed total count.

2.2 DATASETS AND PREPROCESSING

We use scRNA-seq data from two sources. For controlled experiments and ablation studies, we use the pancreas 15-day endocrinogenesis dataset, a well-characterized developmental system with annotated cell types and known trajectories. We retain relevant covariates (sample, batch, species) to account for technical variation. For large-scale experiments, we plan on using the Arc Virtual Cell Atlas scBaseCount resource (Youngblut et al., 2025), which provides raw gene-level counts together with Velocyto-derived S/U matrices across 230M cells from 21 species.

2.3 TRAINING CONFIGURATION

To minimize training costs, we initialize the encoder and NB decoder from pretrained scVI models trained on CellXGene Census data (release 2025-11-08). For the pancreas dataset, we use the *Mus musculus* model (43.7M cells) and for Arc Atlas experiments, we use the *Homo sapiens* model (159M cells). The splicing decoder head is trained from scratch. We adopt a standardized model configuration: 150 epochs, 70:30 train/test split, latent dimensionality of 50, encoder and decoder hidden layers of size 128. Unless otherwise noted, $\mathcal{L}_{\text{splice}}$ uses MSE with unit weight ($\lambda = 1$), and the S/U head uses a 2-layer MLP architecture. Hyperparameter optimization was performed via Optuna grid search over 90 configurations; the top 5 models were retrained for 50 epochs.

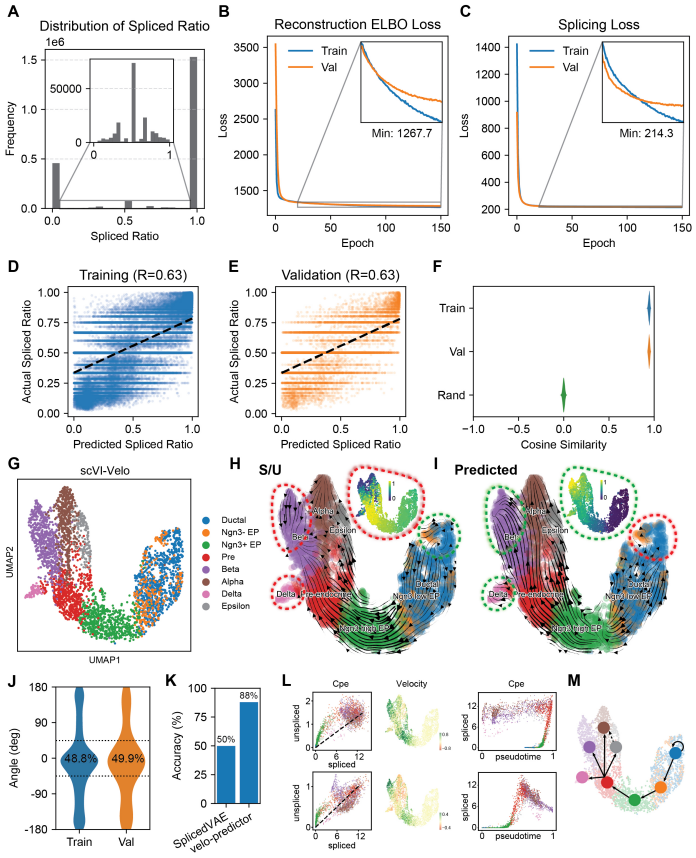

Figure 2: SplicedVAE Improves Latent Representations. (A) Splicing ratio distribution across genes and cell types. (B) Reconstruction ELBO loss and (C) splicing ratio prediction loss. Pearson correlation between predicted and ground-truth splicing ratios during (D) training and (E) validation. (F) Cosine similarity between predicted and true velocity vectors, compared to random baseline. (G) UMAP embedding of SplicedVAE latent space, colored by cell type. Velocity streamlines computed from (H) ground-truth S/U counts and (I) SplicedVAE predictions. (J) Distribution of velocity angles in UMAP space. (K) Directional classification accuracy for SplicedVAE vs. Velo-Predictor. (L) Gene-level velocity comparison (Cpe). (M) Inferred high-level differentiation trajectory map.

2.4 Evaluation Metrics

We benchmark SplicedVAE on held-out cells from the pancreas dataset, quantitatively assessing: (1) splicing ratio prediction via RMSE and Pearson correlation between predicted and ground-truth ratios, computed over all cell–gene pairs; (2) cosine similarity between velocity vectors derived from predicted versus true S/U counts; (3) directional classification accuracy for a simplified 4-class velocity task, comparing against Velo-Predictor (Wang & Zheng, 2021). Qualitative evaluation involves inspecting UMAP embeddings (computed using all cells, train and test) and RNA velocity streamlines to assess whether known developmental trajectories of pancreatic endocrine lineages are faithfully recovered.

3 Results

For splicing ratio prediction, our model achieved a test-set RMSE of 0.1271 and a positive correlation (R=0.67) between predicted and ground-truth splicing ratios on held-out test cells. This indicates the capture of meaningful signal for our *auxiliary* task, with benefits to downstream velocity estimation and trajectory inference.

3.1 Improved Latent Representations via Multitask Learning

To isolate the contribution of the splicing prediction task, we trained a standard scVI model without the auxiliary head on the same pancreas dataset. PCA of all genes produce reasonable cell type clusters in UMAP space, yet trajectory reconstruction already fails to capture fully coherent developmental flows (Fig. 3B, D; Appendix A.1). Restricting to the 5706/8000 gene subset overlapping with scVI's input features degrades trajectory quality further, with Alpha, Beta, and Delta clusters exhibiting disorganized flow patterns (red circles; Fig. 3C, E). Training a standard scVI model on this subset compounds these issues with reconstruction ELBO loss plateauing at higher values (Fig. 3F) and a latent space that yields poor clustering and inaccurate trajectory inference (Fig. 3G).

In contrast, Figure 2 reveals SplicedVAE achieves lower reconstruction ELBO loss (Fig. 2B), demonstrating that the splicing prediction head acts as an effective regularizer. The learned latent space exhibits clearer cluster boundaries and better cell type separation compared to standard scVI (Fig. 2G vs. Fig. 3G). High Pearson correlation (Fig. 2D-E) and cosine similarity (Fig. 2F) between predicted and ground-truth velocity vectors indicate that SplicedVAE captures meaningful directional information exceeding random baselines.

3.2 Velocity Field Reconstruction

Velocity fields derived from SplicedVAE's predicted splicing ratios (Fig. 2I) largely recapitulate the directional patterns observed in ground-truth velocity fields computed from true S/U counts (Fig. 2H). The model preserves key trajectory structures across cell type clusters in UMAP space, suggesting successful capture of underlying developmental dynamics. However, localized discrepancies remain: several regions show disorganized flow patterns (red circles, Fig. 2H), and inferred pseudotime occasionally contradicts known biological progression, indicating room for improvement.

Gene-level velocity comparisons (Fig. 2L for gene Cpe) and high-level trajectory maps (Fig. 2M) further illustrate the model's ability to reconstruct biologically plausible differentiation flows. Directional classification accuracy reaches 50% (Fig. 2K), which falls short of the 88% baseline performance from Velo-Predictor (Wang & Zheng, 2021). While this suggests that continuous splicing ratio prediction does not yet outperform discrete classifiers, continuous outputs enable downstream velocity estimation via scVelo, providing a more flexible representation than categorical outputs.

4 Discussion and Future Work

We introduce SplicedVAE, a novel supervised generative framework that learns splicing dynamics directly from gene expression counts through multitask variational autoencoders. To our knowledge, this is the first approach to successfully predict continuous splicing ratios from expression-only data while simultaneously improving latent representations for trajectory inference. The key innovation is leveraging splicing prediction as an auxiliary task that enhances the latent space to encode dynamics-relevant variation that expression-only models cannot capture.

Our results demonstrate three core contributions: (1) moderate but significant correlation (R=0.67) between predicted and ground-truth splicing ratios without requiring S/U sequencing, (2) improved reconstruction ELBO and clearer cell type clustering through multitask learning, and (3) velocity fields that recapitulate known developmental trajectories in pancreatic endocrinogenesis. These findings suggest that splicing information—though noisy—provides a complementary signal that enhances both generative modeling and dynamical inference in single-cell data.

Several limitations should be noted: (1) the 50% directional classification accuracy compared to Velo-Predictor's 88% indicates the model does not yet reliably capture directional signal. This may stem from error accumulation when converting predicted ratios using scVelo, (2) evaluation is currently limited to a single dataset (pancreas endocrinogenesis), and (3) metrics are computed over all cell–gene pairs; performance likely varies across genes requiring more detailed analysis.

Future work will address these limitations and extend experiments to the full Arc Virtual Cell Atlas, evaluating whether patterns learned in pancreas generalize to diverse developmental systems. We also aim to implement diffusion-based and flow-matching architectures to move beyond point predictions of splicing ratios to generate continuous cellular trajectories.

ACKNOWLEDGMENTS

This research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. We also thank the reviewers for their time and comments.

REFERENCES

Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12): 1408–1414, December 2020. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-020-0591-3. URL https://www.nature.com/articles/s41587-020-0591-3.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, December 2020. URL http://arxiv.org/abs/2006.11239. arXiv:2006.11239 [cs].

Dominik Klein, Jonas Simon Fleck, Daniil Bobrovskiy, Lea Zimmermann, Sören Becker, Alessandro Palma, Leander Dony, Alejandro Tejada-Lapuerta, Guillaume Huguet, Hsiu-Chuan Lin, Nadezhda Azbukina, Fátima Sanchís-Calleja, Theo Uscidda, Artur Szalata, Manuel Gander, Aviv Regev, Barbara Treutlein, J. Gray Camp, and Fabian J. Theis. CellFlow enables generative single-cell phenotype modeling with flow matching, April 2025. URL http://biorxiv.org/lookup/doi/10.1101/2025.04.11.648220.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David Van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, August 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0414-6. URL https://www.nature.com/articles/s41586-018-0414-6.

Jiachen Li, Xiaoyong Pan, Ye Yuan, and Hong-Bin Shen. TFvelo: gene regulation inspired RNA velocity estimation. *Nature Communications*, 15(1):1387, February 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-45661-w. URL https://www.nature.com/articles/s41467-024-45661-w.

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, December 2018. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-018-0229-2. URL https://www.nature.com/articles/s41592-018-0229-2.

Erpai Luo, Minsheng Hao, Lei Wei, and Xuegong Zhang. scDiffusion: conditional generation of high-quality single-cell data using diffusion model. *Bioinformatics*, 40 (9):btae518, September 2024. ISSN 1367-4811. doi: 10.1093/bioinformatics/btae518. URL https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btae518/7738782.

Tarun Mahajan and Sergei Maslov. noSpliceVelo infers gene expression dynamics without separating unspliced and spliced transcripts, August 2024. URL http://biorxiv.org/lookup/doi/10.1101/2024.08.08.607261.

Xin Wang and Jie Zheng. Velo-Predictor: an ensemble learning pipeline for RNA velocity prediction. *BMC bioinformatics*, 22(Suppl 10):419, September 2021. ISSN 1471-2105. doi: 10.1186/s12859-021-04330-1.

Nicholas D. Youngblut, Christopher Carpenter, Arshia Nayebnazar, Abhinav Adduri, Rohan Shah, Chiara Ricci-Tam, Jaanak Prashar, Rajesh Ilango, Noam Teyssier, Silvana Konermann, Patrick D. Hsu, Alexander Dobin, Dave P. Burke, Hani Goodarzi, and Yusuf H. Roohani. scBaseCount: an AI agent-curated, uniformly processed, and autonomously updated single cell data repository, March 2025. URL http://biorxiv.org/lookup/doi/10.1101/2025.02.27.640494.

Zhiliang Zeng, Shouwei Zhao, Yu Peng, Xiang Hu, and Zhixiang Yin. Cascade Forest-Based Model for Prediction of RNA Velocity. *Molecules*, 27(22):7873, November 2022. ISSN 1420-3049. doi: 10.3390/molecules27227873. URL https://www.mdpi.com/1420-3049/27/22/7873.

A FIGURES & VISUALIZATION

A.1 BASELINE SCVI MODEL LIMITATIONS

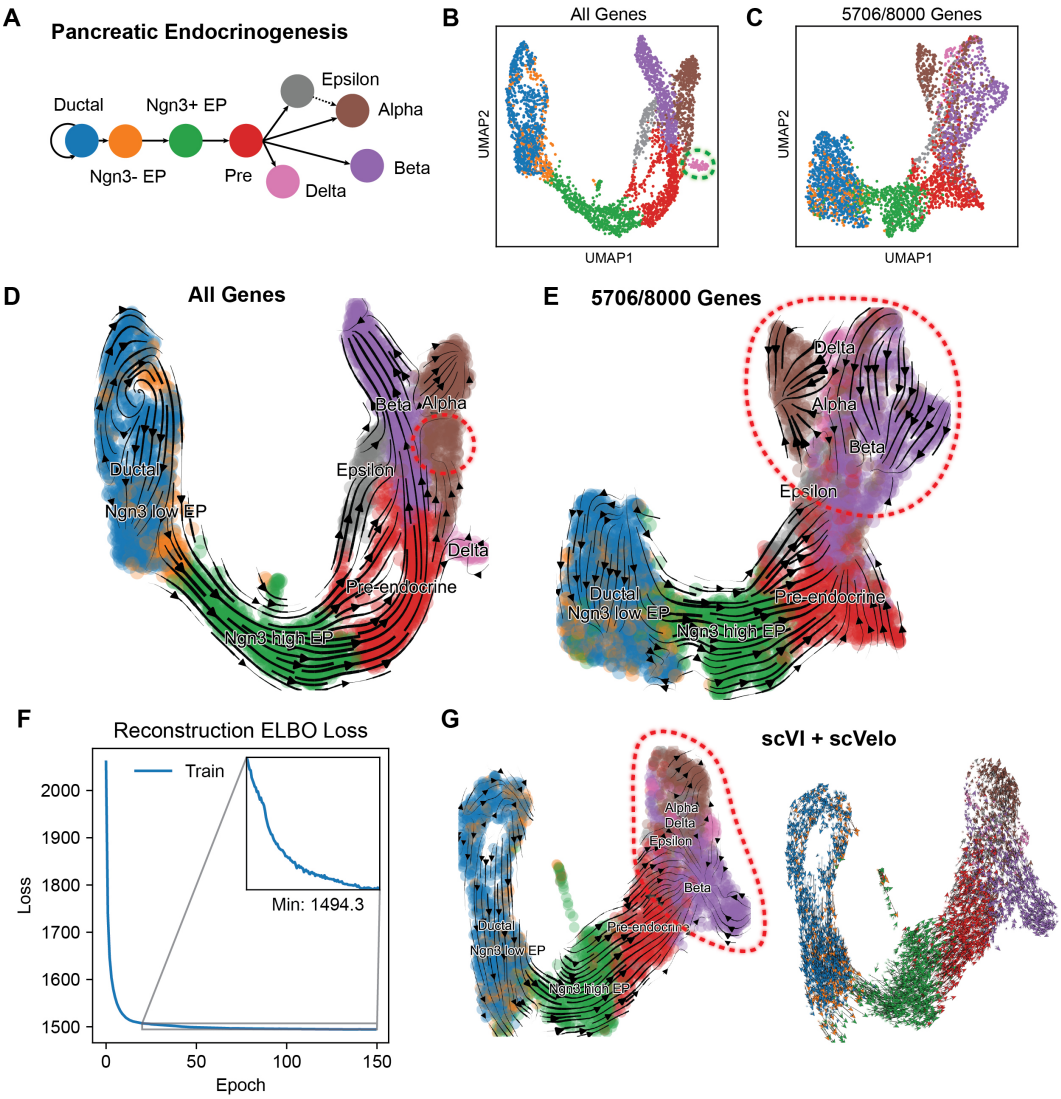

Figure 3: (A) Known developmental hierarchy of pancreatic endocrine cells. (B-C) UMAP embedding of PCA of all genes and subset of scVI genes. (D-E) Trajectory reconstructions for all genes and gene subset, showing disorganized flow patterns (red circles) in Alpha, Beta, and Delta clusters when subset. (F) Reconstruction ELBO loss during scVI training, failing to minimize sufficiently. (G) UMAP embedding from standard scVI, demonstrating poor clustering and inaccurate trajectory inference.

A.2 Related Work

scVelo (Bergen et al., 2020) generalizes RNA velocity estimation by fitting per-gene dynamical models of transcriptional kinetics to the observed spliced and unspliced count distributions. It provides three modes: deterministic, stochastic, and dynamical, with the latter recovering gene-specific rate parameters and latent time. scVelo requires experimentally measured S/U counts as input.

Velo-Predictor (Wang & Zheng, 2021) is a supervised framework that predicts discretized RNA velocity directions from total expression counts. It first computes ground-truth velocities via scVelo, discretizes them into four directional classes in UMAP space, and trains an ensemble of classifiers (random forest, gradient boosting, and neural network) followed by logistic regression to predict these classes. Velo-Predictor achieves high directional accuracy but produces categorical rather than continuous velocity outputs, limiting downstream applications such as trajectory reconstruction.

scVI (Lopez et al., 2018) is a deep generative model for scRNA-seq that uses a variational autoencoder with a negative binomial observation model. It learns a low-dimensional latent representation while accounting for library size and batch effects, and serves as the backbone encoder and expression decoder in SplicedVAE.