INTERPRETABLE SELF-SUPERVISED PROTOTYPE LEARNING FOR SINGLE-CELL TRANSCRIPTOMICS

Fatemeh S. Hashemi G., Till Richter, Alejandro Tejada, Lennard Halle
Institute of Computational Biology
Helmholtz Zentrum München, Germany

{fatemehs.hashemig, till.richter, alejandro.tejada, lennard.halle}@helmholtz-munich.de

Mohammad Lotfollahi
Wellcome Sanger Institute, UK
Cambridge Stem Cell Institute, University of Cambridge, UK

ml19@sanger.ac.uk

Fabian J. Theis
Institute of Computational Biology, Helmholtz Center, Munich, Germany
TUM, School of Computation, Information and Technology, Technical University of Munich, Germany
TUM School of Life Sciences Weihenstephan, Technical University of Munich, Germany

fabian.theis@helmholtz-munich.de

ABSTRACT

Single-cell transcriptomics is inherently noisy and sparse, posing significant challenges for uncovering underlying biological mechanisms. Addressing this issue requires effective denoising strategies to enhance the reliability of biological interpretation. Self-supervised learning has emerged as a powerful approach for learning robust representations across large single-cell datasets, improving denoising and facilitating more accurate biological insights. In this work, we present *scProto*, an interpretable self-supervised learning framework that learns prototypes, which are subsequently decoded into metacells—denoised representations that aggregate information from multiple similar cells across datasets. These metacells enhance robustness, mitigate noise, and provide a more stable and biologically meaningful representation of cell states. Beyond denoising, *scProto* is designed to preserve the structural relationships in the k-nearest neighbor (KNN) graph of the input space while simultaneously removing batch effects through self-supervised prototype learning. The loss function ensures that all cell populations, including rare ones, are well-represented through prototypes. We demonstrate that *scProto* metacells effectively capture marker genes, leading to improved cell-type distinction. Model performance is evaluated using *scGraph* metrics, which assess the preservation of cell similarity structures and geometric relationships in the embedding space, where *scProto* generally outperforms other methods. Additionally, batch effect removal and biological conservation are assessed using *scIB* metrics, indicating that *scProto* performs on par with the best-performing models while achieving better preservation of structural relationships in the embedding space.

1 INTRODUCTION

Single-cell transcriptomics enables detailed characterization of cellular heterogeneity, revealing previously unknown cell states and transitions Angerer et al. (2017). However, inherent noise and sparsity in these data complicate downstream analysis and biological interpretation. Furthermore, heterogeneous or incomplete cell-type annotations pose challenges for supervised learning approaches.

Self-supervised learning (SSL) methods leverage intrinsic data structure to extract meaningful biological features from large-scale single-cell datasets without relying on explicit labels. Recent applications of SSL in single-cell genomics have demonstrated effectiveness in transfer learning, zero-shot classification, cross-modality prediction, and data integration tasks Richter et al. (2024). A prominent class of SSL methods, contrastive learning, builds representations by aligning similar samples and separating dissimilar ones. In prototype-based contrastive learning frameworks like SwAV Caron et al. (2020), data points are compared against learnable prototypes, which serve as aggregated representations encapsulating shared features among related samples. This approach provides a data-driven alternative to traditional clustering methods for generating metacells, effectively aggregating biologically similar cells and reducing batch effects.

While traditional approaches for metacell generation effectively denoise data, they struggle to integrate information across multiple datasets due to their reliance on batch-sensitive clustering methods. To overcome this limitation, we propose a self-supervised prototype learning framework that simultaneously learns metacells and corrects batch effects. This approach enables seamless cross-dataset integration, yielding more robust and biologically meaningful metacells. Our method, as shown in Figure 1, is built on a composite loss function that integrates three key objectives:

- **Self-supervised prototype learning loss**, inspired by SwAV Caron et al. (2020), ensures that augmented versions of the same cell are assigned to the same prototype. In SwAV, both the encoder and prototype weights are trained simultaneously, allowing prototypes to aggregate information from multiple similar cells while the encoder removes batch effects. Our KNN-based augmentation strategy, inspired by graph-based metacell approaches, ensures that prototypes capture biologically meaningful information while preserving structural similarity in the learned embedding space.
- **Rare cell prototype loss**, a minmax loss which ensures that at least one prototype is assigned to every cell, even in low-density regions, enabling the model to preserve rare cell types that might otherwise be overlooked.
- **Conditional variational autoencoder (CVAE) loss**, used within the scPoli Lotfollahi et al. (2023a) architecture, to decode prototypes into metacell representations while preserving biologically relevant information for reconstruction, enhancing interpretability.

By combining prototype-based contrastive learning with a conditional variational autoencoder (CVAE), our proposed framework, *scProto*, jointly addresses noise reduction, batch correction, and interpretability, enabling robust generation of biologically meaningful metacells across multiple single-cell transcriptomic datasets.

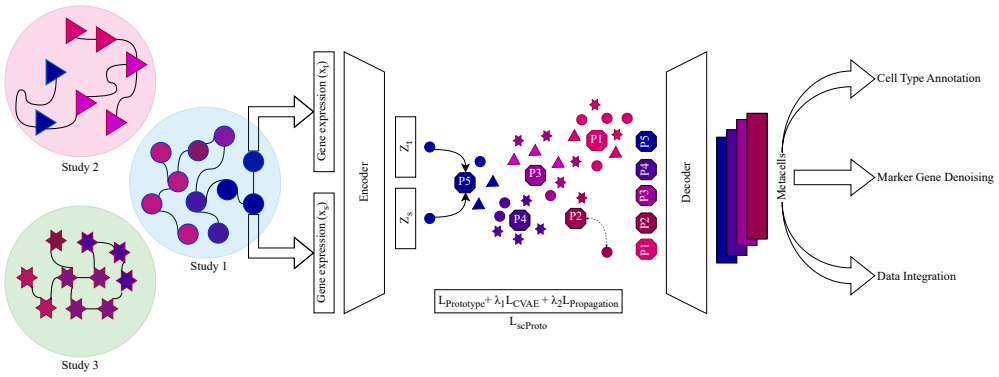

Figure 1: Overview of scProto. The pretraining objective of scProto consists of three key components: (1) Contrastive Prototype Loss, which encourages neighboring cells in the input KNN graph to map to the same prototype; (2) CVAE Loss, which enables prototype decoding into metacells while retaining important biological signals; and (3) Propagation Loss, which ensures that at least one prototype is assigned to all cell types, including rare ones. These objectives collectively train prototypes for effective single-cell aggregation while preserving essential information.

2 RELATED WORK

The increasing availability of single-cell transcriptomic data has advanced the study of cellular heterogeneity and biological processes Angerer et al. (2017). However, the high dimensionality, sparsity, and technical noise inherent in these datasets present challenges for extracting meaningful biological signals. Effective denoising and dimensionality reduction methods are essential to preserve relevant structures and uncover underlying patterns. To mitigate the impact of technical noise and sparsity, metacell-based methods focus on denoising while preserving biological variation. The MetaCell algorithm, introduced by Baran et al. (2019), constructs a KNN graph to identify small, transcriptionally coherent cell groups called metacells. These serve as denoised representations that retain biological structure without relying on explicit smoothing. An extension of this approach, SEACell, incorporates kernel archetypal analysis to better capture both discrete and continuous cell states, making it particularly effective for trajectory inference and rare cell-type identification DePasquale et al. (2023). However, both methods generate metacells using KNN graphs that retain batch effects, restricting metacell formation to within individual batches. Due to the absence of a dedicated mechanism for batch effect removal during metacell generation, additional steps for denoising and batch correction are necessary.

To overcome the high dimensionality of single-cell transcriptomic data while also addressing noise and batch effects, representation learning methods identify biologically meaningful structures by mapping data into compact latent spaces. Variational autoencoder (VAE)-based approaches, such as scVI and its semi-supervised extension scANVI, employ probabilistic modeling of gene expression to account for batch effects and facilitate dataset integration Lopez et al. (2018); Xu et al. (2021). scPoli, while also incorporating a conditional variational autoencoder (CVAE), integrates prototype learning to structure latent spaces by clustering transcriptionally similar cells around shared prototypes Lotfollahi et al. (2023a). Expanding beyond these methods, self-supervised representation learning (SSL) provides a more flexible framework for learning robust representations directly from large-scale, unlabeled data. By capturing intrinsic patterns in single-cell transcriptomics, SSL mitigates technical noise while preserving biological variation, enabling models to generalize across datasets and tasks without relying on extensive annotations. Recent studies Richter et al. (2024) have demonstrated the effectiveness of self-supervised learning in single-cell transcriptomics, enabling models to adapt to new datasets with minimal labeled data. Approaches such as contrastive learning Li et al. (2024) and masked reconstruction Richter et al. (2024) have been instrumental in applications like cross-modality alignment Tang et al. (2023), cell-type classification Zhang et al. (2022), and batch effect correction Kosuru et al. (2024). Building on these advances, foundation models trained on large-scale single-cell datasets further showcase the power of self-supervised learning (SSL) in capturing complex biological relationships. For example, scGPT employs a masked token strategy to infer gene-gene and gene-cell interactions without explicit labels, enhancing cell-type annotation, batch correction, and perturbation response prediction Cui et al. (2024). Likewise, NichFormer integrates dissociated and spatial transcriptomics data to construct spatially informed cell representations Schaar et al. (2024). These models exemplify how scaling SSL to foundation models enables the extraction of biologically meaningful patterns, reduces technical noise and batch effects, and provides a scalable, generalizable framework for single-cell analysis.

Contrastive learning, a fundamental technique in self-supervised learning (SSL), was originally introduced to enforce similarity between augmented views of the same sample while separating representations of different samples Chen et al. (2020). By learning discriminative features without relying on labels, contrastive learning helps structure data in a way that preserves meaningful variations. Clustering-based contrastive learning further refines this process by jointly optimizing feature representations and cluster assignments, leading to more robust and well-separated clusters. For example, SeLa Asano et al. (2019) enforces balanced clustering to ensure that learned representations capture the most informative and diverse structures in the data. SwAV Caron et al. (2020), another contrastive clustering approach, improves upon standard contrastive learning by matching different augmentations of the same sample to learnable prototypes instead of relying on direct pairwise comparisons. Unlike methods that depend on explicit positive and negative pairs, SwAV clusters samples dynamically, enforcing consistency across augmentations in an unsupervised manner, allowing for more flexible and effective representation learning.

Interpretability is essential in machine learning, particularly in biological applications where understanding model decisions can provide valuable insights into underlying biological processes. How-

ever, in SwAV, prototypes are not explicitly evaluated; instead, only the learned embeddings are assessed through downstream tasks. In contrast, the prototype-based learning method introduced in Li et al. (2018) integrates an autoencoder with a prototype layer, enabling case-based reasoning where each prototype can be decoded to reveal the factors influencing the model's decision. In single-cell analysis, expiMap builds upon a variational autoencoder (VAE) framework to structure latent spaces around known and de novo gene programs, improving data integration and biological interpretability Lotfollahi et al. (2023b). DRVI, on the other hand, enhances interpretability by employing an additive decoder to disentangle biological signals, facilitating rare cell-type identification and decomposing cellular heterogeneity into distinct latent dimensions Moinfar & Theis (2024). These methods demonstrate the importance of integrating interpretability into representation learning, particularly in domains where understanding data-driven insights is as crucial as achieving high predictive performance.

Inspired by these ideas, we present a self-supervised prototype learning approach for metacell construction, where learned prototypes can be decoded to form biologically meaningful metacells. We build upon scPoli Lotfollahi et al. (2023a) as the base architecture, extending it to incorporate self-supervised prototype learning, thereby removing reliance on labels and making the method more flexible and data-driven. Additionally, we leverage scPoli's conditional variational autoencoder (CVAE) loss to enable prototype decoding, ensuring that the learned prototypes serve as metacells that capture key biological variation. By integrating learning-based clustering strategies, our approach enhances denoising while preserving biological diversity, effectively addressing limitations in cross-dataset metacell construction.

3 Self-Supervised Prototype Learning

To learn prototypes that aggregate information from multiple datasets, we employ a KNN graph-based contrastive learning framework. In this framework, cells in close proximity within the KNN graph are encouraged to map to the same prototype, ensuring consistency and enabling prototypes to integrate information from multiple similar cells. We use cosine similarity in the embedding space to measure the similarity between cell embeddings and prototypes, ensuring that each cell is assigned to the most relevant prototype. We employed CVAE loss to decode prototypes into metacells while preserving key biological signals during encoding. Additionally, propagation loss, a min-max loss function, ensures that all cell types, including rare populations, are assigned at least one prototype. The full learning objective of scProto is illustrated in Figure 1. In the following sections, we describe our augmentation strategy, the contrastive prototype learning loss, which preserves geometric structure while mitigating batch effects, and the propagation loss designed to improve rare cell type representation. We further detail how prototypes are decoded and interpreted as metacells, ensuring biologically meaningful and denoised representations of single-cell data.

3.1 KNN-Based Data Augmentation

For data augmentation, we leverage the graph introduced by the SEACell algorithm DePasquale et al. (2023). First, we apply a dimensionality reduction algorithm to lower the data's dimensionality and construct a KNN graph that captures the local cell-cell relationships. To generate augmented samples, we randomly select a neighboring cell from the graph for each cell. In the contrastive learning framework, this augmentation encourages the model to preserve geometric structures, ensuring that cells that are close in the input space remain close in the learned embedding space

3.2 Batch-Independent Prototype Learning

For prototype learning, we employ the SwAV Caron et al. (2020) algorithm, which clusters embeddings by assigning each sample to its closest prototype. Prototypes are trained so that augmented versions of the same sample are assigned to the same prototype. A common challenge in clustering-base contrastive learning is collapse, where all embeddings converge to the same representation. SwAV mitigates this by enforcing a balanced distribution of samples across prototypes. When the number of prototypes is sufficiently large, this assumption holds and remains largely unaffected by the underlying data distribution. Even in imbalanced datasets, which are common in biological studies, the prototype allocation naturally adapts, ensuring that classes with more samples receive

a proportionally larger number of prototypes. To achieve this balance, the clustering problem is formulated as an optimal transport task, where sample-to-prototype assignments approximate a uniform distribution. The Sinkhorn-Knopp Cuturi (2013) algorithm is then used to solve this problem, iteratively adjusting the assignments to balance prototype utilization while preserving feature similarity.

In equation 1, z_t and z_s represent embedded representations of neighboring cells, obtained through a shared encoder. Instead of directly comparing features, the model assigns them to prototypes, facilitating self-supervised learning. The optimal cluster assignment, denoted as q, is computed via the Sinkhorn-Knopp algorithm. The probability p_t^k represents the softmax-based assignment of z_t to prototype k, as defined in equation 2. The objective is to align this probability distribution with q_s^k, treating it as the ground truth. The swapped loss, given in equation 2, is formulated as a cross-entropy objective, minimizing the divergence between the model's predicted and Sinkhorn-derived assignments. This promotes consistency across augmentations, ensures balanced prototype utilization, and prevents overrepresentation of certain prototypes.

$$L_{\text{SwAV}}(\mathbf{z}_t, \mathbf{z}_s) = \ell(\mathbf{z}_t, \mathbf{q}_s) + \ell(\mathbf{z}_s, \mathbf{q}_t) \tag{1}$$

$$\ell(\mathbf{z}_t, \mathbf{q}_s) = -\sum_k q_s^{(k)} \log p_t^{(k)}, \quad \text{where} \quad p_t^{(k)} = \frac{\exp\left(\frac{1}{\tau}\mathbf{z}_t^\top \mathbf{c}_k\right)}{\sum_{k'} \exp\left(\frac{1}{\tau}\mathbf{z}_t^\top \mathbf{c}_{k'}\right)}. \tag{2}$$

When running the SwAV algorithm with KNN-based augmentation, the trained embedding tends to exhibit strong batch effects. This occurs because the input KNN graph itself contains batch effects, leading the model to learn batch-specific prototypes rather than aggregating information from multiple datasets. To mitigate this, we modify the loss calculation to enforce an even distribution of samples within each batch across prototypes, preventing the formation of batch-specific clusters and reducing batch effects in the learned embeddings. As described in equation 3, for each batch of data, we separate the samples by study (batch) and compute the SwAV loss independently for each subset. We then average the SwAV loss values across all batches, ensuring that the model learns batch-independent prototypes.

$$\mathcal{L}_{\text{batchSwAV}} = \frac{1}{B}\sum_{b=1}^{B} \mathcal{L}_{\text{SwAV}}^{(b)} \tag{3}$$

3.3 Representation of Rare Cell Types Through Prototypes

The SwAV loss encourages the model to allocate more prototypes to regions with a high density of cells, which can lead to a lack of prototypes for rare cell populations. To address this, we introduce a min-max loss that minimizes the maximum distance between cells and their closest prototype. This ensures that every cell, including rare populations, has at least one prototype nearby.

$$\mathcal{L}_{\text{propagation}} = \max_i \min_j d(x_i, p_j) \tag{4}$$

By applying this loss, we force the model to distribute prototypes more evenly across the data space, preventing underrepresentation of rare cell types and improving their biological interpretability.

3.4 Interpretability of Learned Prototypes

To enable decoding and interpretation of learned prototypes, we incorporate a conditional variational autoencoder (CVAE) loss. This loss not only facilitates prototype decoding but also encourages the model to capture information from the most informative parts of the input, which are essential for accurate cell reconstruction. Additionally, using a CVAE framework ensures that batch-related variations are encoded in the conditional embedding, while the main embedding retains only biological information, making the learned representations more meaningful. Learned prototypes are decoded using the average of all batch embeddings, ensuring that the output is not batch-specific and instead represents a denoised version of the input that aggregates information from multiple, similar cells across different datasets.

equation 5 represents the overall loss function, which integrates the three main aspects of our model. This formulation enables prototype learning in a self-supervised manner while ensuring that prototypes capture all cell types, including rare ones. The learned prototypes are decoded as metacells and evaluated in downstream tasks

$$\mathcal{L}_{\text{scProto}} = \mathcal{L}_{\text{batchSwAV}} + \lambda_1 \mathcal{L}_{\text{propagation}} + \lambda_2 \mathcal{L}_{\text{CVAE}} \tag{5}$$

4 EXPERIMENTALS

4.1 EXPERIMENTAL SETUP

We conduct our experiments on the Immune dataset from the scIB benchmark datasetsLuecken et al. (2022), which contains 33,506 cells and 12,303 genes. The gene expression data is normalized and log1-transformed, and 4,000 highly variable genes (HVGs) are selected for analysis. We use "Freytag" and "Villani" as query studies, resulting in 29,137 cells for training the reference model and 4,369 cells as the query set. Due to time constraints, experiments on additional and larger datasets are planned for future work. For dimensionality reduction, we apply PCA with 50 components and construct a KNN graph with 50 neighbors. This graph is used for positive pair generation (KNN augmentation). The encoder-decoder model follows the scPoli Lotfollahi et al. (2023a) architecture with a latent dimension of 8, using $\lambda_1 = 0.01$ for CVAE loss scaling and $\lambda_2 = 1.0$ for propagation loss scaling. For prototype learning, we use the SwAV Caron et al. (2020) loss with the hard clustering option and $\varepsilon = 0.02$. We compare our method against scPoli, scVI Lopez et al. (2018), and PCA, evaluating performance using scIB metrics and scGraph Wang et al. (2024) metrics. To evaluate how well-trained metacells denoise marker genes, we compared the macro F1-score of a classifier using scProto metacell marker genes against one using individual cell marker genes. Additionally, we plan to compare our results with metacells identified by SEACell DePasquale et al. (2023) and MetaCell Baran et al. (2019), but due to time limitations, this analysis has not yet been conducted.

4.2 CELL TYPE DISTINCTION

Accurately distinguishing immune cell types from single-cell transcriptomic data is challenging due to the sparsity and noise of gene expression. Here, we evaluate classification performance for two biologically relevant distinctions: NK cells vs. $CD8^+$ T cells and $CD8^+$ vs. $CD4^+$ T cells.

NK cells and $CD8^+$ T cells can be distinguished by key marker genes such as *TYROBP* (for NK cells) and *CD8A* (for $CD8^+$ T cells), but classification using raw gene expression is unreliable due to data sparsity. To assess whether metacell representations enhance classification, we encoded each cell using the scProto-trained encoder, identified its closest prototype based on embedding cosine similarity, and decoded the prototype to reconstruct its metacell expression profile. We then classified cells using marker genes from the decoded prototypes and compared performance to classifiers trained on raw single-cell marker genes and scProto embeddings. As shown in Figure 2 (left panel), classifiers trained on prototype marker genes outperform those using raw single-cell marker genes, indicating that metacell representations improve classification by reducing noise while preserving biologically relevant signals. Additionally, classifiers trained directly on scProto embeddings achieve the highest F1 score, suggesting that embeddings integrate information from multiple genes beyond primary markers, further enhancing classification performance.

A similar challenge arises when distinguishing $CD8^+$ and $CD4^+$ T cells, where classification based solely on marker genes is affected by noise and sparsity. To evaluate different representations, we trained classifiers using single-cell marker genes, scProto embeddings, and scPoli embeddings. As shown in Figure 2 (right panel), classifiers trained on scProto embeddings achieve the best performance, highlighting their ability to capture subtle cellular distinctions. Interestingly, scPoli embeddings do not outperform raw marker-based classification, emphasizing the difficulty of this task and suggesting that the contrastive learning framework in scProto provides a more structured latent space for resolving fine-grained biological differences.

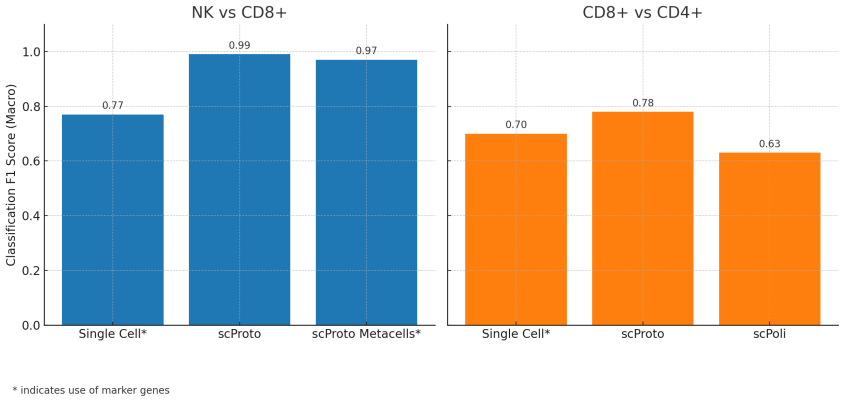

Figure 2: F1 score comparison for distinguishing different immune cell types. Left panel: F1 score for distinguishing NK cells from CD8$^+$ T cells across three different settings: scProto embeddings, single-cell marker genes, and the closest prototype marker genes. Right panel: F1 score for classifying CD8$^+$ vs. CD4$^+$ T cells across three different settings: scPoli embeddings, scProto embeddings, and single-cell marker genes.

4.3 Evaluation of Learnt Embedding Quality

To evaluate the quality of the learned embeddings, we first visualize them using UMAP (Figure 3). The visualization demonstrates that scProto effectively captures cellular heterogeneity, with prototypes representing all annotated cell types, including rare populations such as plasma cells and CD10$^+$ B cells, ensuring that even low-density cell types are assigned at least one prototype. Additionally, closely related cell types, such as CD8$^+$ and CD4$^+$ T cells, which are often difficult to distinguish in single-cell RNA sequencing data, are partially separated in the latent space and assigned distinct prototypes. This suggests that scProto enhances the resolution of subtle cellular differences while maintaining biologically meaningful clustering. To quantitatively assess the preservation of

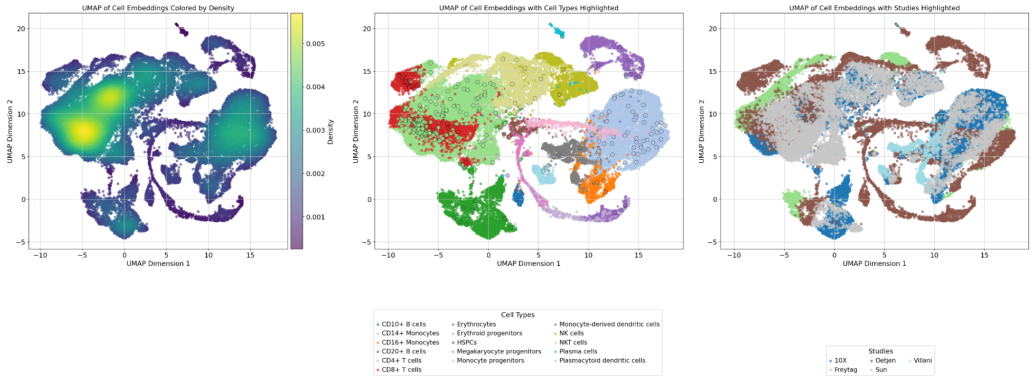

Figure 3: The left panel shows the density of cells across different regions. The middle panel presents a UMAP of cells colored by cell type, along with prototypes. Each prototype is colored by the label color of the majority label among its 10 nearest neighbors. We observe that cells of the same type cluster together, and the prototypes successfully capture all cell types, including rare ones. The right panel displays cells colored by batch, revealing that, in most areas, batches are well integrated. However, some batch effects remain in certain regions, which might reflect underlying biological differences between batches.

biological structure, we employ scGraph metrics from Metric Mirage Wang et al. (2024), which evaluate how well the learned embedding retains cell-cell relationships relative to a reference graph. The reference graph used for scGraph computation is derived from PCA of the original data, mak-

ing PCA a natural baseline for structural preservation. However, while PCA provides an unbiased representation of cell-cell relationships, it lacks batch correction capability, limiting its suitability for integrated single-cell analysis. The results, as shown in Table 1, indicate that scProto not only learns biologically meaningful prototypes that can be decoded as metacells but also better preserves cell-cell similarities within and between clusters, leading to a more faithful representation of cellular relationships. Among methods that do not suffer from strong batch effects, scProto achieves better structural preservation, demonstrating its ability to capture biologically meaningful organization while also addressing batch effects—a limitation of PCA.

To further evaluate embedding quality, we assess scIB metrics, which jointly measure biological conservation and batch correction. As shown in Table 1, scProto is among the best-performing methods in biological conservation, ensuring that cells annotated as the same cell type in the dataset remain close to each other in the embedding space. Additionally, scProto achieves batch correction performance comparable to other state-of-the-art methods, demonstrating its ability to integrate data while preserving cell-type structure. Although PCA achieves the highest scGraph scores due to its direct alignment with the input graph, it completely lacks batch correction, making it unsuitable for integrated single-cell analysis. In contrast, scProto maintains strong scGraph performance while also achieving competitive batch correction, effectively balancing structural preservation and integration quality.

While scIB metrics assess how well annotated cells from the same cell type cluster together while also evaluating batch correction performance, scGraph metrics capture the global structure of the embedding, including relationships between different cell types. Because scProto performs well in both metrics, it successfully balances batch correction and biological structure preservation, making it a robust approach for single-cell transcriptomic analysis.

Model	**scGraph Metrics**			**scIB Metrics**		
	Rank-PCA	Corr-PCA	Corr-Weighted	scIB Total	Batch Correction	Bio Conservation
scProto	<u>0.600</u>	0.748	<u>0.593</u>	0.608	0.559	<u>0.640</u>
scPoli Fully Supervised	0.502	0.572	0.469	**0.668**	**0.647**	**0.683**
scPoli	0.468	0.664	0.443	<u>0.623</u>	<u>0.641</u>	0.610
scVI	0.537	<u>0.766</u>	0.576	0.622	0.596	<u>0.640</u>
PCA	**0.863**	**0.910**	**0.850**	0.517	0.339	0.636

Table 1: Comparison of scIB and scGraph metrics across different models

5 Conclusion

In this work, we introduced scProto, an interpretable self-supervised model that learns metacells as denoised representations by aggregating information from multiple datasets while preserving the geometric structure of single-cell data and removing batch effects. We demonstrated that metacells effectively reduce data sparsity and noise, improving classification performance based on metacell-derived marker genes compared to raw single-cell data. We also evaluated scProto's embedding quality and structural preservation using scGraph and scIB metrics, comparing it with existing integration methods. Training scProto on large-scale cell atlases could establish it as a foundation model, where prototypes aggregate information from vast datasets to reveal meaningful biological structures and provide deeper insights into cellular heterogeneity. Such a model could serve as a building block for future applications, including perturbation prediction for drug discovery, cellular trajectory inference, and disease modeling. Future research can enhance the destination distribution of cell propagation through prototypes and integrate multimodal data to improve both biological fidelity and computational efficiency, ultimately advancing single-cell analysis and uncovering complex biological mechanisms.

6 Meaningfulness Statement

Single-cell transcriptomic data is noisy, high-dimensional, and lacks robust labels, making meaningful representation challenging. We address this by (1) denoising data through self-supervised

prototype learning, aggregating information from similar cells, and decoding them into biologically meaningful metacells, and (2) reducing dimensionality while preserving structure using a latent space that maintains the KNN graph, ensuring similar cells remain close. This self-supervised approach captures intrinsic biological organization without label bias. Our results show that metacells enhance marker gene identification and cell type classification, while the latent space preserves meaningful relationships, validated by scGraph metrics, enabling structured and interpretable single-cell representations.

7 ACKNOWLEDGMENTS

Sincere gratitude is extended to Dr. Arman Sepehr for invaluable assistance in identifying issues in the initial results and optimizing hyperparameters, and to Dr. Mahmoud Ghandi for insightful discussions that resolved critical challenges, particularly regarding prototype propagation and batch-effect correction in graph-based approaches. Appreciation is also given to Artur Sztata for his guidance in identifying appropriate metrics for embedding evaluation, as well as to colleagues from TheisLab for their helpful feedback and constructive suggestions throughout this project. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) - Project number 513025799 (grant number TH 900/19-1). It was also funded by the Hamburger Wissenschaftspreis awarded to F.J.T. and the German Federal Ministry of Education and Research (BMBF) under grant no. 01IS18053A. Additionally, support was provided by the Chan Zuckerberg Initiative Foundation (CZIF; grant CZIF2022-007488, Human Cell Atlas Data Ecosystem) and the European Union (ERC, DeepCell - 101054957).

CONFLICT OF INTEREST

F.J.T. consults for Immunai Inc., CytoReason Ltd, Cellarity, BioTuring Inc., and Genbio.AI Inc., and has an ownership interest in Dermagnostix GmbH and Cellarity.

REFERENCES

Philipp Angerer, Lukas Simon, Sophie Tritschler, F Alexander Wolf, David Fischer, and Fabian J Theis. Single cells make big data: new challenges and opportunities in transcriptomics. *Current opinion in systems biology*, 4:85–91, 2017.

Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*, 2019.

Yair Baran, Aviezer Alpert, Dror Karnieli, et al. Metacell: analysis of single-cell rna-seq data using k-nn graph partitioning. *Bioinformatics*, 35(10):1766–1773, 2019. doi: 10.1093/bioinformatics/bty909.

Mathilde Caron, Ishan Misra, et al. Unsupervised learning of visual features by contrasting cluster assignments. *NeurIPS*, 2020.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods*, 21(8):1470–1480, August 2024. ISSN 1548-7105. doi: 10.1038/s41592-024-02201-0.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Emily DePasquale, Diana Schnell, Saptarshi Ghosh, et al. Seacells: Inference of transcriptional states and trajectories in single-cell rna-seq data. *Nature Methods*, 20:381–390, 2023. doi: 10.1038/s41592-023-01775-5.

Vishnu Vardhan Kosuru, Sree Virajitha Ramaraju, Sri Harshitha Anatatmula, and T Anjali. Contrastive representation learning for multimodal single-cell rna-seq data integration. In *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 627–633, 2024. doi: 10.1109/ICECA63461.2024.10800778.

Oscar Li, Hao Liu, Chaofan Chen, and Cynthia Rudin. Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Shumin Li, Jiajun Ma, Tianyi Zhao, Yuran Jia, Bo Liu, Ruibang Luo, and Yuanhua Huang. Cellcontrast: Reconstructing spatial relationships in single-cell rna sequencing data via deep contrastive learning. *Patterns*, 5(8), 2024.

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, 2018. doi: 10.1038/s41592-018-0229-2.

Mohammad Lotfollahi, Mitra Naghipourfar, Fabian J. Theis, and Guy Wolf. scpoli: A deep generative model for interpretable single-cell omics data integration. *Nature Communications*, 14:1234, 2023a. doi: 10.1038/s41467-023-45678-9.

Mohammad Lotfollahi, Sergei Rybakov, Karin Hrovatin, Soroor Hediyeh-Zadeh, Carlos Talavera-López, Alexander V Misharin, and Fabian J Theis. Biologically informed deep learning to query gene programs in single-cell atlases. *Nature Cell Biology*, 25(2):337–350, 2023b.

Malte D Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F Müller, Daniel C Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods*, 19(1):41–50, 2022.

Amir Ali Moinfar and Fabian J. Theis. Unsupervised Deep Disentangled Representation of Single-Cell Omics, November 2024.

Till Richter, Mojtaba Bahrami, Yufan Xia, David S Fischer, and Fabian J Theis. Delineating the effective use of self-supervised learning in single-cell genomics. *Nature Machine Intelligence*, pp. 1–11, 2024.

Anna C Schaar, Alejandro Tejada-Lapuerta, Giovanni Palla, Robert Gutgesell, Lennard Halle, Mariia Minaeva, Larsen Vornholz, Leander Dony, Francesca Drummer, Mojtaba Bahrami, et al. Nicheformer: a foundation model for single-cell and spatial omics. *bioRxiv*, pp. 2024–04, 2024.

Wenzhuo Tang, Hongzhi Wen, Renming Liu, Jiayuan Ding, Wei Jin, Yuying Xie, Hui Liu, and Jiliang Tang. Single-cell multimodal prediction via transformers. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 2422–2431, 2023.

Hanchen Wang, Jure Leskovec, and Aviv Regev. Metric mirages in cell embeddings. *bioRxiv*, pp. 2024–04, 2024.

Chenling Xu, Romain Lopez, Edouard Mehlman, Jeffrey Regier, Michael I Jordan, and Nir Yosef. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular systems biology*, 17(1):e9620, 2021.

Ruiyi Zhang, Yunan Luo, Jianzhu Ma, Ming Zhang, and Sheng Wang. scpretrain: multi-task self-supervised learning for cell-type classification. *Bioinformatics*, 38(6):1607–1614, 2022.