TopoGAN: unsupervised manifold alignment of single-cell data

Akash Singh[1], Marcel J.T. Reinders[1,2,3], Ahmed Mahfouz[1,2,3] and Tamim Abdelaal[1,2,4*]

[1]Delft Bioinformatics Lab, Delft University of Technology, 2628 XE Delft, The Netherlands, [2]Leiden Computational Biology Center, Leiden University Medical Center, 2333ZC Leiden, The Netherlands, [3]Department of Human Genetics, Leiden University Medical Center, 2333ZC Leiden, The Netherlands, [4]LKEB, Department of Radiology, Leiden University Medical Center, 2333ZC Leiden, The Netherlands

*To whom correspondence should be addressed.

Abstract

Motivation: Single-cell technologies allow deep characterization of different molecular aspects of cells. Integrating these modalities provides a comprehensive view of cellular identity. Current integration methods rely on overlapping features or cells to link datasets measuring different modalities, limiting their application to experiments where different molecular layers are profiled in different subsets of cells.

Results: We present TopoGAN, a method for unsupervised manifold alignment of single-cell datasets with non-overlapping cells or features. We use topological autoencoders to obtain latent representations of each modality separately. A topology-guided Generative Adversarial Network then aligns these latent representations into a common space. We show that TopoGAN outperforms state-of-the-art manifold alignment methods in complete unsupervised settings. Interestingly, the topological autoencoder for individual modalities also showed better performance in preserving the original structure of the data in the low-dimensional representations when compared to using UMAP or a variational autoencoder. Taken together, we show that the concept of topology preservation might be a powerful tool to align multiple single modality datasets, unleashing the potential of multi-omic interpretations of cells.

Availability and implementation: Implementation available on GitHub (https://github.com/AkashCiel/TopoGAN). All datasets used in this study are publicly available.

Contact: t.r.m.abdelaal@lumc.nl

1 Introduction

A growing number of single-cell technologies allow the characterization of distinct molecular features of cells, such as single-cell RNA-sequencing (scRNA-seq) or measuring chromatin accessibility at single-cell resolution (scATAC-seq). Despite advances in multimodal technologies (Zhu *et al.*, 2020), these molecular features are mostly measured from different subsets of cells. Sometimes the measured modalities share common features, for example when spatial transcriptomics and scRNA-seq are applied on the same tissue. Because the datasets are not measured from the same cells, they have to be aligned into a common space (Argelaguet *et al.*, 2021). These unimodal alignment methods are not applicable when distinct modalities are measured from distinct cells. Few methods, such as UnionCom (Cao *et al.*, 2020), MMD-MA (Singh *et al.*, 2020) and SCIM (Stark *et al.*, 2020), have attempted to address this challenging integration task by assuming a similar cellular composition between unimodal datasets collected from the same tissue. These multi-modal alignment methods, however, suffer from several limitations. First, UnionCom and MMD-MA were tested on single-cell multi-omics data, in which the multiple modalities were measured from the same cell, with perfect cell-to-cell correspondences. Although they did not exploit this correspondence in their methods, the integration performance drops significantly (as we show later) when, more realistic, datasets lacking this correspondence are used. Second, SCIM is not fully unsupervised as it requires (partial) cell type annotations for each of the different modalities in order to align the data.

Considering different single-cell modalities measured from the same biological sample, the main assumption in integration is that the different modalities lie on the same underlying manifold (Sun *et al.*, 2018). Preserving the topology of the datasets is crucial when constructing and integrating the different manifolds. Since the different modalities are measuring distinct features, it is necessary to first find a low-dimensional representation of each modality separately. Topological autoencoders have been recently introduced to project high-dimensional data into a low-dimensional latent space while preserving the data topology (Moor *et al.*, 2020). Next, these low-dimensional manifolds have to be aligned into a common space with minimal distortion to the original topology of each data modality. Generative Adversarial Networks (GANs) were successfully used in the computer vision field (Gui *et al.*, 2021). GANs were previously used to project biological datasets onto each other (Amodio and Krishnaswamy, 2018), however, based on correspondence information between the datasets, and not in a fully unsupervised setting.

We propose TopoGAN, a topology-preserving multi-modal alignment of two single-cell modalities with non-overlapping cells or features. TopoGAN first finds topology-preserving latent representations of the different modalities, which are then aligned in an unsupervised way using a topology-guided GAN. TopoGAN outperforms state-of-the-art methods and is less computationally expensive, allowing it to scale to datasets with a large number of cells. Moreover, TopoGAN is fully unsupervised with no requirement for cell type annotations.

2 Methods

2.1 TopoGAN overview

TopoGAN is designed to align two datasets measuring two different single cell modalities, each measured on different non-matching cells. TopoGAN consists of two steps: 1) manifold projection, and 2) manifold alignment (Fig. 1). Assuming a lower-dimensional manifold structure for single-cell datasets (Bac and Zinovyev, 2019), TopoGAN first finds the manifold for each modality separately, with explicit preservation of the data topology. Then, the latent space representation of the two modalities are aligned in a topology-preserving manner, exploiting the assumption that the topology of the cells in the two modalities is the same. This alignment step should preserve relevant inter-modality correspondence such that similar cell types should be aligned between different modalities.

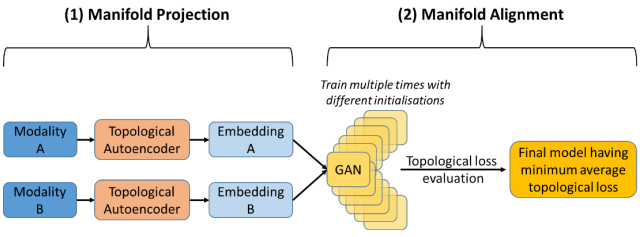

Fig. 1 TopoGAN overview. TopoGAN consists of two stages: (1) Manifold projection using Topological autoencoders to obtain a low dimensional embedding (manifold) for each modality independently. (2) Manifold alignment using a GAN. Hereto, 20 different GAN models are trained with random initializations. Then that model is selected which has the minimum average topological loss. The selected model is further trained for 1000 additional epochs to produce the final alignment.

2.1.1 Manifold projection

To project each modality to a lower-dimensional latent space, TopoGAN uses a topological autoencoder (Moor *et al.*, 2020), which chooses point-pairs that are crucial in defining the topology of the manifold instead of trying to optimize all possible point-pairs. A topological autoencoder is based on the concept of persistence homology (Edelsbrunner and Harer, 2008) which selectively considers edges connecting point-pairs below a certain distance threshold. These edges are used to construct local neighborhoods together constituting large-scale topological features. By repeating this procedure for a wide range of distance thresholds, persistent topological features are defined, where the point-pairs constituting them are known as persistence pairings. Preserving the distances between these pairings in a lower-dimensional projection of the data preserves the data topology. The loss function of the topological autoencoder is defined as:

$$L = L_r + \lambda L_t \quad (1)$$

where L_r is the reconstruction loss between the input and reconstructed output of the autoencoder across all cells, and L_t represents the topological loss, with λ is the weight of the topological loss. The topological loss is defined as:

$$\begin{aligned} L_t &= L_{XZ} + L_{ZX} \\ L_{XZ} &= \frac{1}{2}\|A^X[\pi^X] - A^Z[\pi^X]\|^2 \\ L_{ZX} &= \frac{1}{2}\|A^Z[\pi^Z] - A^X[\pi^Z]\|^2 \end{aligned} \quad (2)$$

where X is the original input data and Z is the encoded latent representation, A^X and A^Z are the distance matrices in the original and latent spaces respectively, π^X and π^Z are the persistence pairings in the original and latent spaces respectively. $A^{\cdot}[\pi^{\cdot}]$ represent subset of distances in the space $A^{\cdot}$ defined by the topologically relevant edges in that space $\pi^{\cdot}$. The term L_{XZ} ensures that persistence pairings relevant to the original manifold are equidistant in both the original and the latent spaces, while L_{ZX} ensures that persistence pairings relevant to the latent manifold are equidistant in both spaces.

To train the topological autoencoders, we used a learning rate of 0.001, batch size of 50, latent size of 8 dimensions, and an architecture of two hidden layers each of the size of 32 followed by batch normalization and ReLU activation. For the hyperparameter λ, we tested values ranging from 0.5 to 3.0 as recommended by (Moor et al., 2020).

2.1.2 Manifold alignment

We used a GAN (Goodfellow *et al.*, 2014) to align one modality (source) to the other modality (target). The generator part of the GAN aims to project the source modality onto the target modality, resulting in a combined dataset. We use a single hidden layer generator network against a double hidden-layer discriminator network. We followed previous work using GANs to stabilize the training process (Radford *et al.*, 2016) by sampling discriminator weights and biases from a normal distribution N(0,0.02), and using a Leaky ReLU as the activation function for the discriminator with an activation value of 0.2. The GAN was trained for 1000 epochs.

To ensure a topology-preserving alignment of the two modalities, we trained 20 different GANs and selected the GAN which best preserves the topological loss (Eq. 2) between the source data and its projection in the target data space. To do so, for each GAN, the topological loss is calculated from epoch 500 every 100th epoch until epoch 1000 (6 values) and then averaged. The topological loss was calculated for a batch size of 1000 to balance between coverage of global structure in each batch and compute memory requirements. The generator network of the selected GAN is then loaded into a new GAN model with a new discriminator network as its adversary. This final model is then trained for an additional 1000 epochs to obtain the final aligned manifolds. For evaluation purposes, we also retrained the final model 10 times with different initializations to assess the stability of the alignment performance.

2.2 Datasets

2.2.1 Peripheral blood mononuclear cells (PBMC) dataset

The PBMC dataset consists of healthy human PBMCs, simultaneously profiling gene expression (RNA) and chromatin accessibility (ATAC) from the same cells using the 10x multiome protocol. The dataset was downloaded from the 10x Genomics website (https://www.10xgenomics.com/resources/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-10-k-1-standard-2-0-0). The Full PBMC dataset contained 10,412 cells, profiling 15,494 genes and 85,468 peaks, having one-to-one correspondence between the two modalities, and including 8 major cell classes (CD4 T cells, CD8 T cells, Monocytes, NK cells, Dendritic cells, B cells, MAIT and HSPC) which are further divided into 19 cell subclasses.

To simulate a realistic data in which the cell-cell correspondences do not exist between the two modalities, we generated the Partial PBMC dataset where we randomly removed 30% of the cells from both RNA and ATAC independently stratified across the different cell classes. This results in a total of 7,329 cells for each modality, including 2,142 cells which have no corresponding cells in the other modality.

2.2.2 Bone marrow (BM) dataset

The BM dataset consists of human bone marrow cells, simultaneously profiling gene expression (RNA) and protein expression (antibody-derived tags, ADT) from the same cells using the CITE-seq protocol (Stoeckius *et al.*, 2017). The Full BM dataset contained 30,672 cells, profiling 17,009 genes and 25 ADT, with cell-cell correspondence between both modalities, and including 5 major cell classes (T cells, B cells, Mono/DC cells, NK cells and Progenitor cells), further categorized into 27 different cell subclasses.

Similar to the PBMC dataset, we generated the Partial BM dataset breaking the cell-cell correspondences across modalities. We randomly selected 9,053 cells from each modality in a stratified manner across the cell classes, in this case, such that there are no cell-cell correspondences between the two modalities left.

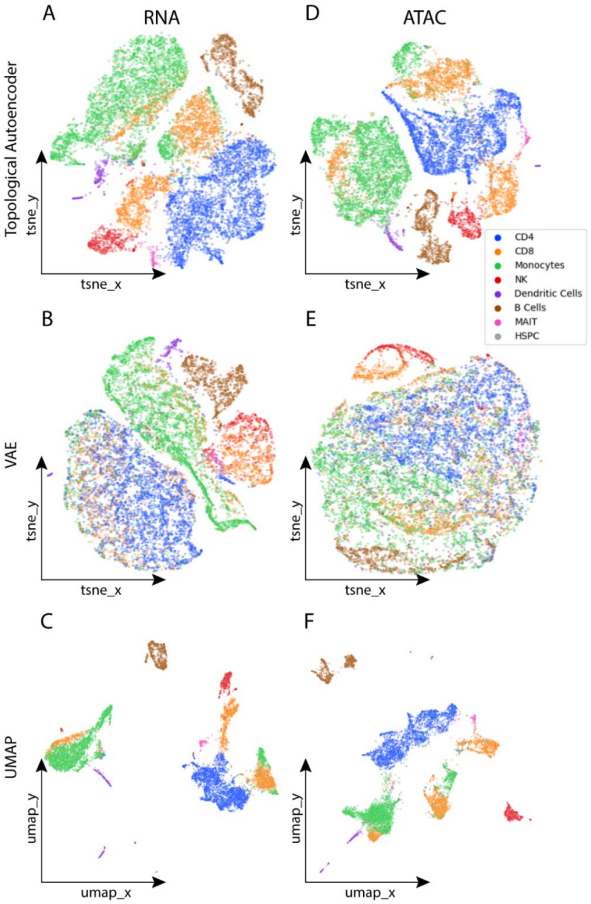

Fig. 2 Qualitative comparison of the manifold projection. Plots showing two-dimensional tSNE embeddings of the 8-dimensional manifolds obtained using **(A,D)** topological autoencoder, **(B,E)** VAE, and **(C,F)** 2-dimensional embedding obtained using UMAP, for the RNA modality and the ATAC modality of the PBMC dataset, respectively. All plots are colored according to the cell classes.

2.3 Data preprocessing

We performed all data preprocessing using the Seurat v4.0 R package (Hao *et al.*, 2021). For the PBMC dataset, we filtered out cells with RNA count below 1,000 or above 25,000, cells with ATAC count below 5,000 or above 70,000, and cells with mitochondrial percentage above 20%. Further, the RNA modality is normalized using `SCTransform` (Hafemeister and Satija, 2019), selecting the top 3000 variable genes. The ATAC modality was normalized using the `RunTFIDF` function using a scaling factor of 10,000, followed by finding the top peaks using the `FindTopFeatures` function with min.cutoff = q0. Next, we reduced the dimensionality of the RNA and ATAC data to 50 dimensions using Principal Component Analysis (PCA) and Latent Semantic Indexing (LSI), respectively. These 50-dimensional datasets are used as input to the TopoGAN workflow.

For the BM dataset, the RNA modality was normalized using a scaling factor of 10,000 followed by log-transformation. The top 2000 variable genes were selected, next the data was scaled and centered. The ADT modality was centered log-ratio (CLR) normalized. The RNA data was reduced to 50 dimensions using PCA, while dimensionality reduction was not necessary for the ADT modality which only had 25 features.

2.4 Benchmarking methods

For the manifold projection step, we compared the performance of the topological autoencoder with a standard variational autoencoder (VAE) (Kingma and Welling, 2014), which is used for manifold projection in SCIM (Stark *et al.*, 2020). Further, we used UMAP (McInnes *et al.*, 2018) as a base-line for the manifold projection evaluation. Next, we compared the alignment performance of TopoGAN with the state-of-the-art methods UnionCom and MMD-MA. Default hyperparameter settings were used for both methods, and UnionCom was trained for 1000 epochs.

2.5 Evaluation metrics

To evaluated the manifold projection, we used the Silhouette score (Rousseeuw, 1987) which assesses the separation between the cell classes. The Silhouette score ranges from -1 to 1, where a higher value indicates better separable classes. Additionally, we calculated the Kullback-Leibler divergence KL_σ between the density estimates of the input data and the it's latent space representation (Moor *et al.*, 2020). The KL_σ value quantifies the dissimilarity between the density estimates, thus lower values (≈ 0) indicate better manifold projection performance. We chose a σ of 0.01, which represents the length scale of the Gaussian kernel.

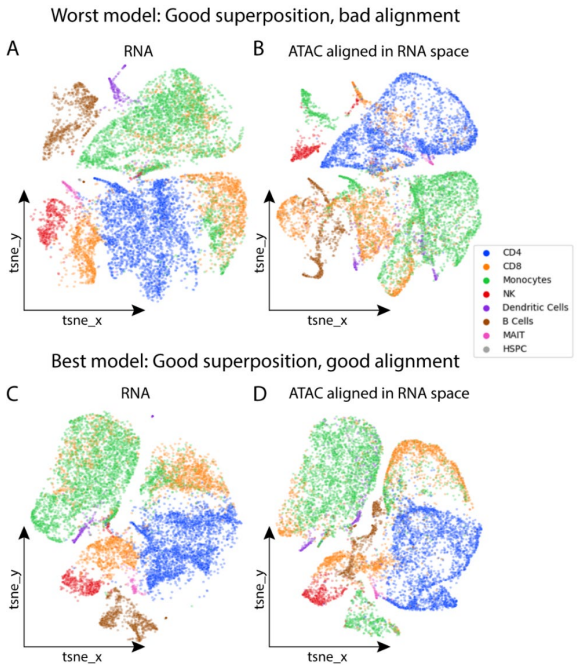

Fig. 3 GAN alignment vs superposition. Plots show tSNE embeddings of the final alignment of two GAN models, showing **(A,C)** the RNA modality (target) and **(B,D)** the ATAC modality (source) after being projected and aligned in the RNA space. **(A,B)** Worst GAN model performing good superposition of the two manifolds, but bad alignment in terms of cell classes (6.1% Celltype matching score). **(B)** Best GAN model with good alignment projecting the correct cell classes across the two modalities (75.6% Celltype matching score).

To evaluate the manifold alignment, for each cell in one modality, we determine its k-neighboring cells from the other modality in the final aligned common space (k=5, Euclidean distance). Next, we compare the class/subclass annotation of that cell with the majority vote of its neighbors and check whether it is a match or not. We report the percentage of cells with matching cell class/subclass donated as the Celltype matching and the Subcelltype matching scores, respectively.

3 Results

3.1 Topological autoencoder produced better manifold projections compared to VAE and UMAP

Before integrating different data modalities, it is crucial to acquire a proper low-dimensional embedding of each modality separately. For this manifold projection task, we used a topological autoencoder which has been shown to produce reliable topology approximations (Moor *et al.*, 2020). To the best of our knowledge, topological autoencoders have not been applied on biological datasets which, compared to classical datasets used in machine learning, contain continuous topological structures (Rizvi *et al.*, 2017). Using both RNA and ATAC modalities of the Full PBMC dataset (see Methods), we compared the manifold projection performance of the topological autoencoder with VAE and UMAP methods (Table 1). All three methods were used to reduce the 50-dimensional (PCA or LSI) data to 8 dimensions, additionally UMAP was used to produce 2-dimensional embedding for visualization purpose. Furthermore, we tested different settings for the topological loss weight λ of the topological autoencoder, and the KL weight of the VAE. Results show that the topological autoencoder is the best method in preserving the original data density estimates having overall the lowest $KL_{0.01}$ value. However, UMAP obtained the highest Silhouette score producing better separation between different cell classes.

To qualitatively compare the low-dimensional manifolds produced by each method, we generated two-dimensional t-distributed stochastic neighbor embedding (tSNE) maps (van der Maaten and Hinton, 2008) of the 8-dimensional manifolds of the topological autoencoder (Fig. 2A,D) and VAE (Fig. 2B,E), in comparison with the 2-dimensional UMAP embeddings (Fig. 2C,F). We observed that the topological autoencoder and UMAP produced good separation between different cell classes, while VAE mixes CD4 and CD8 T cells in the RNA modality, and mixes most of the classes in the ATAC modality. The topological autoencoder mimics the densities in the original data better than UMAP. Taken together, the topological autoencoder showed better performance in producing low-dimensional manifolds preserving the original structure of the data.

Table 1. Manifold projection evaluation using the Full PBMC dataset

Dataset	Method	Silhouette score	$KL_{0.01}$
RNA	Topological Autoencoder ($\lambda = 2.0$)	0.175	**0.007**
	Topological Autoencoder ($\lambda = 3.0$)	0.061	**0.007**
	VAE (KL weight = 1e-4)	-0.030	0.020
	VAE (KL weight = 1e-5)	-0.123	0.020
	VAE (KL weight = 1e-6)	-0.200	0.026
	UMAP (8 dimensions)	**0.229**	0.330
	UMAP (2 dimensions)	0.172	0.312
ATAC	Topological Autoencoder ($\lambda = 0.5$)	0.061	**0.001**
	Topological Autoencoder ($\lambda = 1.0$)	0.091	**0.001**
	VAE (KL weight = 1e-5)	-0.160	0.012
	VAE (KL weight = 5e-5)	-0.124	0.021
	VAE (KL weight = 1e-6)	-0.099	0.020
	UMAP (8 dimensions)	**0.275**	0.200
	UMAP (2 dimensions)	0.229	0.173

3.2 Minimum topological loss ensured manifold alignment instead of superposition and stabilized alignment performance using GAN

After obtaining the lower-dimensional manifold of each modality using topological autoencoders, these manifolds are integrated into one common space. We applied the TopoGAN manifold alignment on the Full PBMC dataset, aligning the ATAC modality (source) to the RNA space (target). We observed an inconsistency in the alignment performance when training multiple GANs initialized with different weights. Although their different losses were more or less equal, the Celltype matching score (see Methods) of 40 different GANs was 41.4 ± 20.6% (mean ± standard deviation). We visualized the resulting alignments for the best and the worst models (Fig. 3). Both GANs achieve a good superposition of the ATAC manifold onto the RNA manifold, aligning the ATAC data to match the shape of the RNA data. However, the worst GAN produced a poor alignment of cell classes, e.g. projecting CD4 T cells to Monocytes (Fig. 3A). Whereas, the best GAN correctly aligns most cell classes (Fig. 3B).

To quantify how distorted the source manifold is after projection to the target space, we inspected the topological loss between the source data and the projected source. Using the Full PBMC dataset, we performed 40 experiments using identical GAN architectures but different random initializations, trained for 1000 epochs. GAN training showed that the generator and discriminator losses stabilized around 400 epochs. Therefore, we calculated the topological loss from epoch 500 to 1000 every 100 epochs,

between the input ATAC (source) manifold and the output ATAC manifold projected onto the RNA space. To interpret the topological losses, we correlated them with the Celltype and the Subcelltype matching scores at the same epochs, resulting in a negative Pearson correlation of -0.73 and -0.67, respectively. This negative correlation indicates that GANs with low topological loss (i.e. preserving the topology of the source data after alignment) tend to produce better manifold alignment. This observation promoted us to train 20 different GANs and select the model with the minimum average topological loss (see Methods) as the final TopoGAN model. We chose to train 20 base GANs as that showed to cover a wide range of alignment scores.

Using the final TopoGAN model and across 10 different experiments with the Full PBMC dataset, we obtained an average Celltype matching score of 67.5 ± 2.0%, and an average Subcelltype matching of 54.5 ± 2.3%. Hence, incorporating the topological loss greatly helps to stabilize the performance of the GAN model, and ensures manifold alignment instead of just superposition of datasets.

3.3 TopoGAN outperforms state-of-the-art methods

We benchmarked TopoGAN against UnionCom and MMD-MA. First, we applied the three methods to the Full PBMC dataset, and evaluated the results using the Celltype and Subcelltype matching scores (Table 2). UnionCom outperformed both TopoGAN and MMD-MA. However, the Full PBMC dataset includes perfect cell-cell correspondences between the two modalities, since they are measured from the same cell. Although none of the models are explicitly using this information, it might implicitly be informative for some of the models. To address this, we generated a Partial PBMC dataset in which the modalities are partially taken from non-overlapping cells (see Methods). We observed a large drop in the performance of UnionCom when tested on the Partial PBMC dataset (Table 2). Although the performance of TopoGAN on the Partial PBMC dataset also decreased compared to the performance on the Full PBMC, TopoGAN outperformed other methods on the Partial PBMC data.

These observations are replicated on the BM dataset which measures RNA and protein levels using the CITE-seq protocol. For the Full BM dataset, TopoGAN obtained an average Celltype matching score of 77.8 ± 2.0 %, and an average Subcelltype matching score of 43.2 ± 3.4 % over 10 different experiments. Running UnionCom and MMD-MA failed for the Full BM dataset, due to the large memory requirements of both methods. The memory requirements of TopoGAN are almost constant with the number of cells, due to the fixed batch size. For the Partial BM dataset with no cell-cell correspondences, TopoGAN similarly outperforms UnionCom and MMD-MA (Table 2).

Table 2. Benchmarking TopoGAN against UnionCom and MMD-MA*

Dataset	Method	Celltype matching (mean ± std) %	Subcelltype matching (mean ± std) %
Full PBMC	TopoGAN	67.5 ± 2.0	54.5 ± 2.3
	UnionCom	**95.7 ± 0.5**	**90.4 ± 0.9**
	MMD-MA	23.1 ± 9.5	12.4 ± 5.7
Partial PBMC	TopoGAN	**59.3 ± 3.0**	**47.4 ± 1.4**
	UnionCom	23.1 ± 6.6	13.4 ± 5.0
	MMD-MA	31.7 ± 10.9	14.5 ± 8.6
Partial BM	TopoGAN	**56.6 ± 0.4**	**32.3 ± 7.3**
	UnionCom	29.1 ± 11.8	9.9 ± 5.2
	MMD-MA	45.9 ± 10.1	15.4 ± 6.7

**Reported results are computed over 10 different runs.*

Further, we qualitatively compared the performance of TopoGAN and UnionCom in order to interpret the drop in the performance. We visualized the final alignment results for the Full PBMC, Partial PBMC and Partial BM datasets (Fig. 4). For the Full PBMC dataset, both TopoGAN and UnionCom showed good mixing of the RNA and ATAC modalities, while keeping different cell classes separable (Fig. 4A). For the Partial PBMC dataset, TopoGAN showed a similar behavior, while UnionCom is not able to properly align or mix the two modalities anymore (Fig. 4B). Similar observation can be obtained using the Partial BM dataset (Fig. 4C).

4 Discussion

We present TopoGAN, a method to integrate multi-modal single-cell data with non-overlapping cells or features. TopoGAN is fully unsupervised and relies on the assumption that different single-cell modalities measured from the same tissue have the same underlying manifold, hence the topological structure of these modalities should be similar. To perform manifold alignment, TopoGAN uses a GAN in combination with a topological loss guiding the selection of the best performing GAN. We would like to stress that although the topological loss idea was inspired based on the evaluation using the cell type annotations, these annotations are not at all used by TopoGAN (the topological loss is fully unsupervised).

We showed that TopoGAN outperforms current state-of-the-art methods, UnionCom and MMD-MA, when tested on datasets with partial or no cell-cell correspondence. UnionCom was trained and optimized using single-cell multi-omics datasets, and showed good performance on datasets in which the different modalities are measured from the same cell. Although the cell-cell correspondence is not used as input, UnionCom is apparently able to implicitly learn from this information as its performance largely dropped for more realistic scenarios where modalities are measured from different cells. MMD-MA relies heavily on hyperparameters tuning which needs to be done using cell annotation information to obtain good estimates. Hence, MMD-MA is not suitable for a fully unsupervised integration setting.

For manifold projection, we used topological autoencoders and showed their ability to preserve the structure of the data in the low-dimensional embedding. Topological autoencoders showed better results compared to VAE and UMAP, both quantitatively and qualitatively. Additionally, it was previously shown that topological autoencoders have superior performance to PCA and regular autoencoder (Moor *et al.*, 2020). Therefore, it might be interesting to explore the applicability of topological autoencoders in other single cell analysis tasks. One example is trajectory inference studying the differentiation trajectory of cells using scRNA-seq datasets (Saelens *et al.*, 2019). Most trajectory inference methods rely on a lower-dimensional representation of the data, where topological autoencoders can be applied to produce low-dimensional space preserving the topology of the inter-cellular relationships in the data.

The main assumption of TopoGAN, different modalities measured from the same tissue have the same underlying manifold structure, is not completely true. Although this assumption is based on the fact that the cell pool where different modalities sample from is the same, hence similar cellular structure, different modalities are measuring different molecular features capturing different views of this cellular structure. As a result, the underlying manifolds of each modality are not identical, however, we argue that there is enough similarity between these manifold that can be used to perform the data integration.

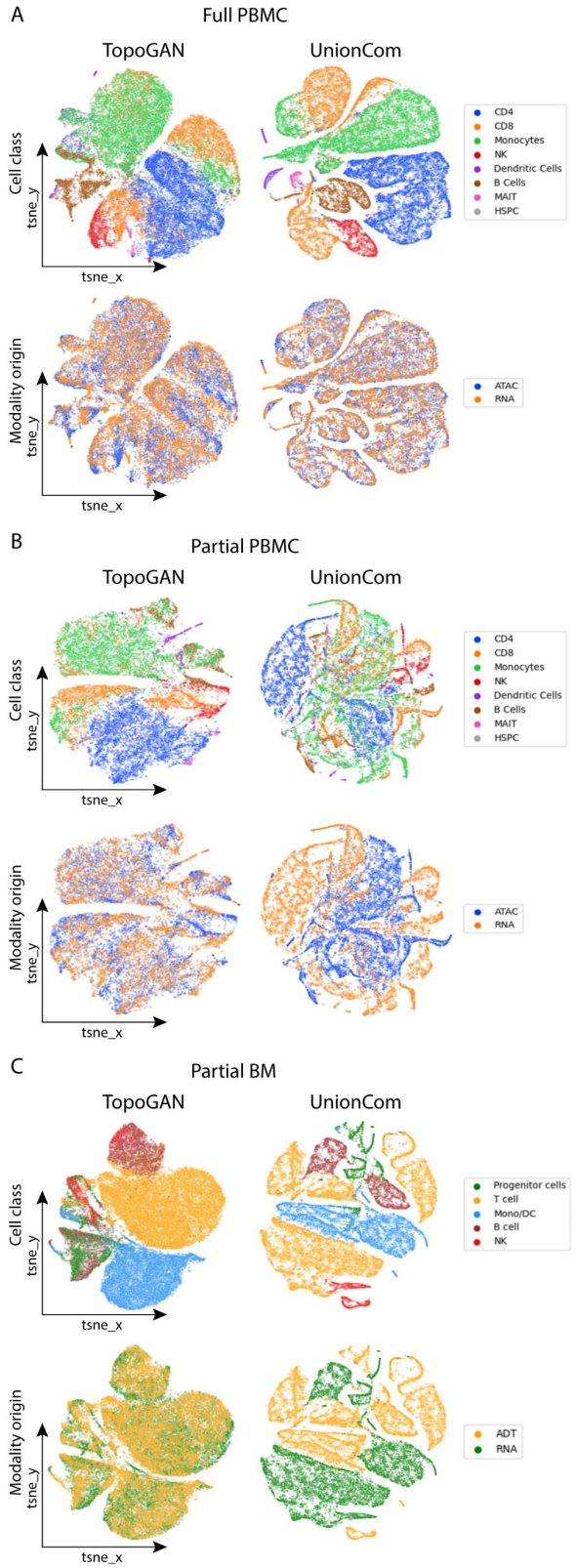

Fig. 4 Qualitative comparison of TopoGAN and UnionCom. Plots show tSNE embeddings of the final alignment produced by TopoGAN (left column) and UnionCom (right column) when applied on **(A)** Full PBMC dataset, **(B)** Partial PBMC dataset, and **(C)** Partial BM dataset. Each tSNE map is plotted twice, once colored with the cell classes showing how well different cell classes are separated, and once colored with the modality of origin showing how well different modalities are mixed.

A major limitation in the current TopoGAN workflow is the requirement of training multiple GAN networks in order to choose the best model based on the topological loss. It is evident that the quality of the alignment achieved is limited by the best alignment obtained in this set of GAN models. Here, we trained 20 different GAN models which is computationally expensive and there is no guarantee that the select GAN model is the best possible solution for the tested dataset. Future improvement is this direction can incorporate the topological loss as a regularization term in the overall loss function of the GAN. This will guide the GAN to minimize the topological loss during training, thus eliminating the need to train multiple GAN models.

In all our experiments, we used the RNA modality as the target modality to which other source modalities (ATAC or ADT) were aligned. The choice of the target modality has an impact on the final alignment performance. Furthermore, we did not fine-tune the hyperparameters used for each dataset. Optimizing these hyperparameters specifically for each dataset may improve the overall results.

In conclusion, TopoGAN opens new opportunities in studying complex tissues as it enables the integration of multiple molecular views without the restriction that these are measured from the same cell.

Funding

This project was supported by the NWO Gravitation project: BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology (NWO: 024.004.012) and the NWO TTW project 3DOMICS (NWO: 17126).

Conflict of Interest: none declared.

References

Amodio,M. and Krishnaswamy,S. (2018) MAGAN: Aligning biological manifolds. In, *35th International Conference on Machine Learning, ICML 2018*.

Argelaguet,R. *et al.* (2021) Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.*, **39**.

Bac,J. and Zinovyev,A. (2019) Lizard brain: Tackling locally low-dimensional yet globally complex organization of multi-dimensional datasets. *Front. Neurorobot.*, **13**.

Cao,K. *et al.* (2020) Unsupervised topological alignment for single-cell multi-omics integration. *Bioinformatics*, **36**, i48–i56.

Edelsbrunner,H. and Harer,J. (2008) Persistent homology—a survey.

Goodfellow,I.J. *et al.* (2014) Generative adversarial nets. In, *Advances in Neural Information Processing Systems*.

Gui,J. *et al.* (2021) A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Trans. Knowl. Data Eng.*

Hafemeister,C. and Satija,R. (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.*, **20**.

Hao,Y. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**.

Kingma,D.P. and Welling,M. (2014) Auto-encoding variational bayes. In, *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*.

van der Maaten,L. and Hinton,G. (2008) Visualizing Data using t-SNE. *J. Mach. Learn.*, **9**, 2579–2605.

McInnes,L. *et al.* (2018) UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.*, **3**, 861.

Moor,M. *et al.* (2020) Topological autoencoders. In, *37th International Conference on Machine Learning, ICML 2020*.

Radford,A. *et al.* (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. In, *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*.

Rizvi,A.H. *et al.* (2017) Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.*, **35**.

Rousseeuw,P.J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**.

Saelens,W. *et al.* (2019) A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.*, **37**, 547–554.

Singh,R. *et al.* (2020) Unsupervised manifold alignment for single-cell multi-omics data. In, *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, BCB 2020*.

Stark,S.G. *et al.* (2020) SCIM: Universal single-cell matching with unpaired feature sets. *Bioinformatics*, **36**.

Stoeckius,M. *et al.* (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods*, **14**, 865–868.

Sun,X.J. *et al.* (2018) An integrated analysis of genome-wide DNA methylation and gene expression data in hepatocellular carcinoma. *FEBS Open Bio*, **8**.

Zhu,C. *et al.* (2020) Single-cell multimodal omics: the power of many. *Nat. Methods*, **17**, 11–14.