Single-cell multi-modal GAN reveals spatial patterns in single-cell data from triple-negative breast cancer

Highlights

- Integrating data from multiple modalities into one analysis

- Using data geometry to regularize cycle-consistent GANs

- Quantifying uncertainty through noise augmentation

Authors

Matthew Amodio, Scott E. Youlten, Aarthi Venkat, Beatriz P. San Juan, Christine L. Chaffer, Smita Krishnaswamy

Correspondence

smita.krishnaswamy@yale.edu

In brief

Combining single-cell data from different modalities is important for creating a holistic understanding of the system under study. This is demonstrated on a variety of example modalities with the scMMGAN model introduced in this paper.

CellPress

Article

Single-cell multi-modal GAN reveals spatial patterns in single-cell data from triple-negative breast cancer

Matthew Amodio,[1] Scott E. Youlten,[2,5] Aarthi Venkat,[3] Beatriz P. San Juan,[2,4,5] Christine L. Chaffer,[2,4,5] and Smita Krishnaswamy[1,6,7,*]

[1]Yale University Department of Computer Science, New Haven, CT, USA
[2]Garvan Institute of Medical Research, Darlinghurst, NSW, Australia
[3]Yale University Computational Biology and Bioinformatics, New Haven, CT, USA
[4]St Vincent's Clinical School, UNSW Medicine, UNSW Sydney, Sydney, NSW, Australia
[5]The Kinghorn Cancer Centre, Darlinghurst, NSW, Australia
[6]Yale University Department of Genetics, New Haven, CT, USA
[7]Lead contact
*Correspondence: smita.krishnaswamy@yale.edu
https://doi.org/10.1016/j.patter.2022.100577

THE BIGGER PICTURE Biological experimental data are increasingly being generated along multiple different axes, with new and more complex technologies specializing in particular measurements being developed every year. Measuring a single subject or system with multiple specialized data-collecting tools creates a natural interest in integrating the results of these individual instruments to form a single unified view. The model introduced here presents a computational technique designed for this purpose. With the single-cell multi-modal GAN (scMMGAN), there is an opportunity to measure along many different omic directions and synthesize the information from each into one larger understanding of the system under study.

[1][2][3][4][5] **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Exciting advances in technologies to measure biological systems are currently at the forefront of research. The ability to gather data along an increasing number of omic dimensions has created a need for tools to analyze all of this information together, rather than siloing each technology into separate analysis pipelines. To advance this goal, we introduce a framework called the single-cell multi-modal generative adversarial network (scMMGAN) that integrates data from multiple modalities into a unified representation in the ambient data space for downstream analysis using a combination of adversarial learning and data geometry techniques. The framework's key improvement is an additional diffusion geometry loss with a new kernel that constrains the otherwise over-parameterized GAN. We demonstrate scMMGAN's ability to produce more meaningful alignments than alternative methods on a wide variety of data modalities and that its output can be used to draw conclusions from real-world biological experimental data.

INTRODUCTION

Integrating data gathered from different sources is a critical challenge in computational genomics. Currently there are several single-cell technologies including RNA sequencing (RNA-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), Hi-C, ChIP-seq (chromatin immunoprecipitation

sequencing), and CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) as well as proteomic technologies such as cytometry by time of flight (CyTOF), imaging CyTOF, and multiplexed ion beam imaging that offer complementary cellular information.[1–7] Of these modalities, only a fraction are available as simultaneous measurements—often with quality degradation factors such as reduced gene dimensions, lower throughput,

and increased noise. The remaining measurements must be done on distinct cellular subsamples from the same population. This is the key problem that we tackle in this article: the prediction of missing or non-simultaneous modalities in order to generate a more complete set of features. Thus, given modalities such as single-cell RNA-seq (scRNA-seq), scATAC-seq, and spatial transcriptomics measured separately on different cells (from the same population), our single-cell multi-modal generative adversarial network (scMMGAN) generates a complete set of simultaneous measurements for downstream analyses.

Aligning the separately measured data computationally has many advantages over analyzing the data modalities individually. Combining data modalities allows us to leverage the advantages of each and mitigate the disadvantages. For example, combining a modality with a higher signal-to-noise ratio such as proteomic CyTOF measurements with one that has a lower signal-to-noise ratio such as scRNA-seq gives us the opportunity to resolve cell populations to a finer degree in the noisier dataset.[7] Even more compellingly, combining modalities allows us to measure variables only available in one domain combined with variables only available in another domain, thus simulating jointly measured technologies.

We base our method on the framework of cycle-consistent generative adversarial networks (CycleGANs).[8–11] In GAN-based domain adaptation frameworks, a generator network is trained to map data points of one modality into data points from another modality. During training, a discriminator is used to ensure that generated points are sampled from the high-dimensional distribution representing the second modality. In CycleGANs there are two back-to-back generators, one going from the first modality to the second modality and another going from the second modality to the first. A reconstruction error enforces that the result of two back-to-back domain adaptations results in the original distribution again, i.e., that the generators are inverses of each other over the regions of the data spaces where there are training points.

While CycleGAN frameworks can successfully generate points in each modality, the mapping they produce is not constrained enough. For instance in the original CycleGAN paper, images of zebras were mapped to images of horses, but nothing ensured that the background would be unaltered. While this may not be detrimental for natural image applications, it can be untenable for scientific applications where the scRNA-seq measurement must correspond with and corroborate the scATAC-seq measurement. Noting this key weakness, in earlier work we proposed the use of a correspondence loss and gave anecdotal examples on flow cytometry panels with overlapping measured markers.[12] However, here we both extend the application to multi-modal integration and specify a more powerful, generally applicable correspondence loss: the geometry-preserving loss. This loss enforces that the diffusion geometry, performed with a new kernel designed to pass gradients better than the Gaussian kernel, is preserved throughout the mapping. We note that this loss can be utilized even in cases where no measurements overlap.

We demonstrate the power of aligning data modalities with scMMGAN on a wide array of data types. We start by validating it on datasets where simultaneous measurements are available and use those as ground truth in evaluations. We then use scMMGAN to perform a thorough investigation into a novel triple-negative breast cancer dataset, where we have cells from the breast cancer culture HCC38 xenografted into mice and allowed to metastasize from the primary to secondary tumor locations. We show that scMMGAN can infer spatial locations of cellular structures.

RESULTS

scMMGAN model results

The scMMGAN framework is depicted in Figure 1A. Each pair of data domains or modalities has a pair of generator networks that map in either direction between them, forming a diversified multi-modal mapping. For a generator mapping from Domain i to Domain j which we denote G_{ij}, it functions as a traditional GAN guided by a discriminator in Domain j which we denote D_j. The discriminator tries to distinguish between samples from the real data for that domain x_j and samples from the generator $G_{ij}(x_i)$, while the generator tries to fool the discriminator. They alternate trying to optimize the following minimax objective:

$$\min_{G_{ij}} \max_{D_j} \mathbb{E}_{x_j \sim P(x_j)} \log(D_j(x_j)) + \mathbb{E}_{x_i \sim P_{x_i}} \log(1 - D_j(G_{ij}(x_i))).$$

(Equation 1)

In addition to this loss of the discriminator guiding the generator to transform its input modality into the output modality L_{GAN}, there are two other terms in the loss that ensure the learned mapping is informative and meaningful. These are depicted in Figure 1B. The reconstruction loss L_r is the mean-squared error (MSE) between the original data x_i and the composition of the two paired generators between the domains i and j:

$$L_r = \|x_i - G_{ji}(G_{ij}(x_i))\|^2.$$

(Equation 2)

The correspondence loss L_c imposes a constraint on a single point's representation in Domain i and Domain j, as opposed to the reconstruction loss which imposes a constraint on points within the same domain:

$$L_c = correspondence(x_i, G_{ij}(x_i)).$$

(Equation 3)

The motivation for the correspondence loss comes from the fact that previous models using cycle consistency for domain mapping with GANs included only two restrictions: (1) that the generators be able to reconstruct a point after it moves to the other domain and back; and (2) that the discriminators not be able to distinguish batches of true and mapped points. The generators can accomplish these goals in many different ways, including by learning arbitrarily complex mappings: as long as they align the two data manifolds at a distribution level. The family of paired inverse functions that can match the target distributions is large, and with existing frameworks the particular pair that results from training is determined by the vagaries of random weight initialization and mysterious biases in gradient descent.

The GAN loss is a probability-distribution matching objective.[13] In previous work it has been proven that under certain optimality conditions a GAN discriminator provides a Jensen-Shannon divergence between the true and generated distributions.[14] A Wasserstein GAN (WGAN), on the other hand, contains
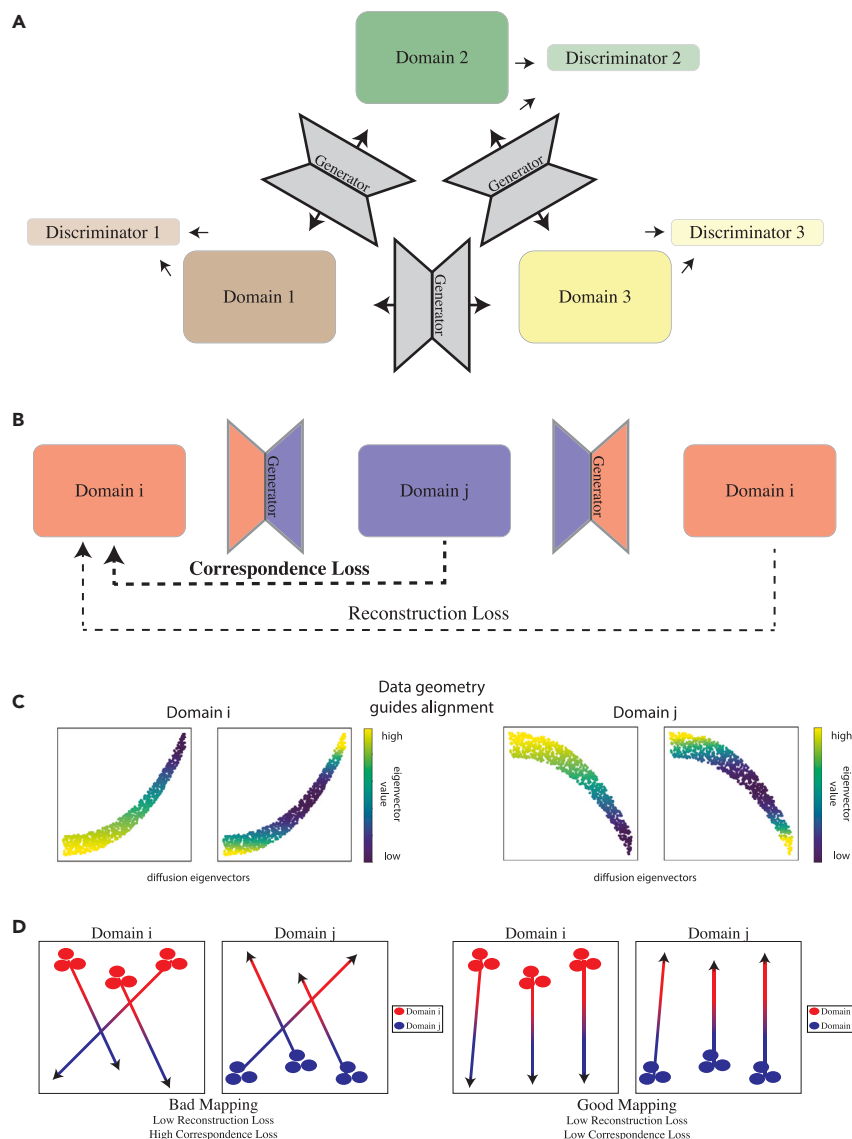
modifications that result in a Wasserstein distance being provided by the discriminator.[15]

However, simply matching probability distributions can result in incoherent cell states (Figure 1D). A key insight we bring is that distributions must only be matched within correspondence constraints. These correspondences are essentially invariances in the underlying system that are reflected in every modality. In our previous work we used customized correspondence losses for each dataset. However, here we note that when matching single-cell data we can use a nearly universal constraint—that of manifold geometry preservation.

While our model incorporates signal from a data geometry loss into a larger framework, the data geometry is too rigid to be used on its own to guide alignment. It is heavily influenced by the properties of the domain data space, and thus when the two domains are very different it does not allow for sufficient flexibility in changing the shape of the distribution. Methods that use only the data geometry struggle to align domains that are significantly different.[16] An ideal mapping would have both the flexibility of a

mapping that matches the probability distribution (as the GAN does) but preserves the data geometry as much as possible while doing so. By combining the existing GAN-based loss and a data geometry loss, the network can balance the tradeoff between these goals.

We thus introduce a correspondence loss that ensures the mappings have point-wise as well as distributional alignment by preserving the data geometry through the learned mapping. To do this, we use the diffusion map representation of the original data.

Here we give a brief overview of diffusion maps. Diffusion maps are a kernel-based method frequently used in manifold learning to produce low-dimensional embeddings that preserve intrinsic structure in the data.[17,18] The eigenvectors of the diffusion operator form an embedding where Euclidean distances correspond to diffusion distance, or the probability of getting from one point to another via random walk, on the original manifold.[19] Because these new coordinates represented by the diffusion eigenvectors abstract away much of the data-domain specific properties, they present a way of ensuring the underlying data geometry is preserved in the mapping.

By calculating the eigenvectors of the diffusion operator for the points in their original domain φ_{x_1} and the eigenvectors of the diffusion operator for the points after being mapped to the other domain $\varphi_{G(x_1)}\|_2$, we can directly compare the ith eigenvectors to enforce that intrinsic data geometry, as measured by diffusion, be preserved by the mapping. For further detail about the calculation of φ, see experimental procedures.

The correspondence geometry loss then penalizes the L2 distance between the two representations of each point:

$$L_c = \|\varphi_{x_1}^i - \varphi_{G(x_1)}^i\|_2. \qquad \text{(Equation 4)}$$
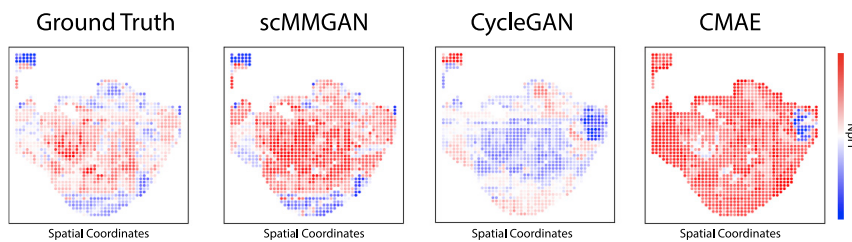
Figure 2. Results comparison from the DBIT-seq experiment
On the DBIT-seq data, shown are corresponding proteomic and transcriptomic expression for the gene shown. The x axis and y axis plotted are the measured spatial coordinates taken directly from the data. The ground-truth transcriptomic values are plotted alongside the generated proteomic values for each model, where we see scMMGAN best model the data.

By enforcing this loss in the scMMGAN framework, we ensure that the intrinsic structure of the data is preserved in the otherwise underconstrained GAN setting.

Experimental results
Mapping between spatial, scRNA-seq, and proteomic data

As an initial validation of scMMGAN, we utilize simultaneously measured multi-modal data from the newly developed deterministic barcoding in tissue sequencing (DBIT-seq) technology as ground truth. DBIT-seq uses deterministic barcoding in tissue for spatially resolved measurements of both transcriptomics and proteomics.[20] Thus, in DBIT-seq, three things are measured jointly on every cell: an scRNA-seq profile, a protein profile, and spatial coordinates. The system being studied in these data is that of mouse embryos, particularly focused on early tissue development and organogenesis.

Often, in transcriptomic/proteomic alignment problems, no "ground truth" is available because each technology measures a distribution of cells in a destructive process. As a result, models such as scMMGAN that learn to map between two distributions without needing point-wise pairings are called unsupervised alignment models. We design an experiment with these data to show how scMMGAN could have been used to obtain this information without needing them to be measured jointly. We treat the spatially located scRNA-seq data and the spatially located protein data as two separate measurements and learn to map between them. We then utilize the fact that they were measured jointly and that some of the columns in each dataset are related (corresponding genes and proteins) to evaluate the accuracy of the learned mapping. We compare against both an autoencoder-based alignment method (cross-modal autoencoder [CMAE]) and a standard CycleGAN without a correspondence loss.[9,21] For detailed descriptions of the model architectures, see experimental procedures.

Figure 2 shows example results of scMMGAN and baseline models on these data. Plotted on the given spatial coordinates, we show the ground-truth transcriptomic value along with generated proteomic values. There we see scMMGAN best models the ground-truth data. We further evaluate scMMGAN's performance on this application quantitatively. To quantify the aspect of the generated distribution matching the target distribution as a whole, we employ the metric maximum mean discrepancy (MMD), a distance defined on distributions frequently used in both deep-learning and biological contexts to distinguish between distributions.[22–24] To quantify the aspect of preserving information about the individual observation through the alignment, we use correlation between columns in the transcriptomic

space and the proteomic space known to correspond to the same gene. Since these values correspond to the same gene, we would expect there to be a correlation between a point's value before mapping and its value after mapping.

These scores confirm quantitatively what we saw graphically in these experiments (Table 1). All models are able to accurately match the target distribution (low MMDs), with very similar performance consisting of each model's one or two SD interval overlapping. However, when looking at the preserved correlation, we see scMMGAN achieved the best alignment with an average correlation of $r = 0.154$ between columns known to correspond. We note that the absolute value of this correlation is relatively low compared with other datasets, and this is due to limited amount of shared correlation in the underlying "ground truth" pairings of points that are jointly measured.

Unique versus common information in measurement modalities. The scMMGAN is a generative framework, but when used in nonstandard ways it can become a tool for analysis in addition to faithful generation. When measuring two aspects of a biological system with two different technologies, some of the information might be shared between the two modalities while other information might be unique to one of the modalities. For example, when mapping between a modality that measures the whole transcriptomic space such as scRNA-seq and one that measures only a subset of the proteomic space, we would expect for the genes with corresponding proteins to be more easily modeled than the genes without them.

We design our experiment as follows to test this on the transcriptomic and proteomic measurements in the DBIT-seq data (summarized in Figure 3, with further mathematical detail in experimental procedures). We train the model augmented with random noise input and then evaluate it on mapping the same points in proteomic space to the transcriptomic space, except with different random noise samples. We then calculate the variance of the different predicted values for each transcript count and for each given point in proteomic space. The mean across all points then gives us a measure of the uncertainty associated with a given transcript measurement. To factor out the influence the magnitude of counts of a given transcript would have on variance, we scale by the variance in the raw dataset for each one. We also filter out lowly expressed genes. We can then compare the stochasticity as measured in this experiment of the genes that have a corresponding proteomic measurement and those that do not.

Just as expected, we find that the average variance of genes with an analogous proteomic measurement is 0.026 while the average variance of genes without an analogous proteomic measurement is 1.419. This is a logical result, as the relationship

Table 1. Results from the DBIT-seq experiment

DBIT-seq	scMMGAN	CycleGAN	CMAE
MMD ($G(x_1),x_2$)	**0.072** ± 0.001	0.078 ± 0.006	0.082 ± 0.001
MMD ($x_1,G(x_2)$)	**0.060** ± 0.001	0.066 ± 0.002	0.079 ± 0.001
Correlation ($G(x_1),x_2$)	**0.155** ± 0.006	−0.026 ± 0.021	0.003 ± 0.082
Correlation ($x_1,G(x_2)$)	**0.152** ± 0.014	−0.088 ± 0.074	−0.012 ± 0.066

Evaluation of each model on the DBIT-seq data. While the MMDs are close for each model (meaning the predicted distribution resembles the ground-truth distribution), scMMGAN is significantly more accurate at preserving correlation between columns known to correspond. The best values are in boldface.

between transcript counts with corresponding proteomic measurements is more straightforward to model and can thus be done with more certainty. This corroborates our understanding of the process at work in this multi-modal setting.

Furthermore, this analytical process with scMMGAN allows us to inspect which genes are modeled with the most and least uncertainty and thus provide the most unique information and most common information with respect to the other modality. Unsurprisingly, the genes with the least uncertainty are related to early embryo development in mice, as that is the system being studied: for example, the three least are Gm5049, Gm37500, and Gm33051. Meanwhile, the genes with the most uncertainty are GM37686 and Rp1. It is possible that genes with higher observed uncertainty could also be of interest to the research, for example, Rp1, which is involved in the development of the retina while the study was focused on early tissue development.[25] The information that these genes had high uncertainty can help guide future experimental design decisions that would lead to the selection of proteins to measure, thus allowing for better alignment of these measurements.

In this way, scMMGAN can help provide insights into the system being studied as well as into experimental design and decisions.

Mapping between scRNA-seq and ATAC-seq data

We next perform an experiment on data consisting of paired ATAC-seq and RNA-seq measurements on the same cells. As with the previous experiment, since these two technologies both measure values related to a particular gene (chromatin availability for ATAC-seq and transcript expression for RNA-seq), we can expect there to be some correlation between the two spaces in their values for that gene, as in the prior case. The dataset we use comes from a public human blood dataset of granulocytes removed through cell sorting from peripheral blood mononuclear cells of a healthy donor.[26]

A qualitative assessment of the results via plots of the output are shown in Figure 4, with the ground-truth ATAC value plotted in the first column and the generated corresponding RNA-seq values in the subsequent columns. As before, scMMGAN's output best matches the ground truth. For the other models, while they have the appropriate amount of activation for each gene at a distribution level, they are inaccurate in terms of alignment at a point level (some populations have been inverted).

Confirming this quantitatively, as seen in Table 2, while all of the models perform adequately at matching the ground truth at a distribution level (as seen by their low MMDs), a significant difference can be seen when evaluating them at a point-wise level. scMMGAN's predictions have an average correlation with the

ground truth of $r = 0.336$ while the others are all essentially uncorrelated (0 is near the middle of each models' $1 - \sigma$ interval).

Integration of triple-negative breast cancer data

Here, we apply scMMGAN to a dataset that comprises a human xenograft model of triple-negative breast cancer (MDA-MB-231) with transcriptomic measurements jointly in both a spatial RNA-seq modality and a scRNA-seq modality lacking the spatial information. The study consists of the MDA-MB-231 cell line grown in mouse models, with the measurements taken from primary site tumors in the tissue from the mammary gland. While the experimental models are replicates, they are different organisms and thus introduce an additional source of noise for the alignment.

Each of the two measurement modalities produces transcriptomic measurements, but each also has advantages and disadvantages. The spatial RNA-seq provides the ability to analyze the physical structure of the tissue sample and localize behavior to different regions of it via (x, y) spatial coordinates. As a drawback, however, each spatial coordinate is bigger than the size of a single cell, and as a result the transcriptomics are estimates of groups of multiple cells. For example, if a cell of one type that is expressing gene A and a cell of another type that is expressing gene B are spatially adjacent, this technology would observe gene A and gene B being expressed together, even if they are never jointly expressed in a single cell.

In contrast, the scRNA-seq provides the usual single-cell granularity of measurements that would be able to distinguish between the expression of each cell. By mapping the spatial data to the scRNA-seq space, we are in essence imputing it into single-cell resolution. However, the scRNA-seq does not have spatial orientation with respect to the original tissue sample. Thus, to combine the best of each modality (spatial information at the single-cell level), we use scMMGAN to integrate them by mapping a point from the spatial RNA-seq domain to the scRNA-seq domain while considering its aligned representation of its original spatial coordinates and its generated scRNA-seq expression values.

The spatial RNA-seq dataset consists of a tissue sampled across 1,170 spatial coordinates, each coordinate with measurements on 20,092 genes. Four different scRNA-seq samples were obtained from cancerous primary site tissue (from different mice), each measured across the same 20,092 genes, and consisting of 7,606, 5,118, 8,163, and 7,591 cells, respectively.

While the scRNA-seq and spatial RNA-seq data are both transcriptomic technologies measuring gene profiles, and thus their dimensions have the same meaning, the two datasets cannot be analyzed together as is. In Figure 5A, we see that the two data distributions are completely non-overlapping prior to the use of
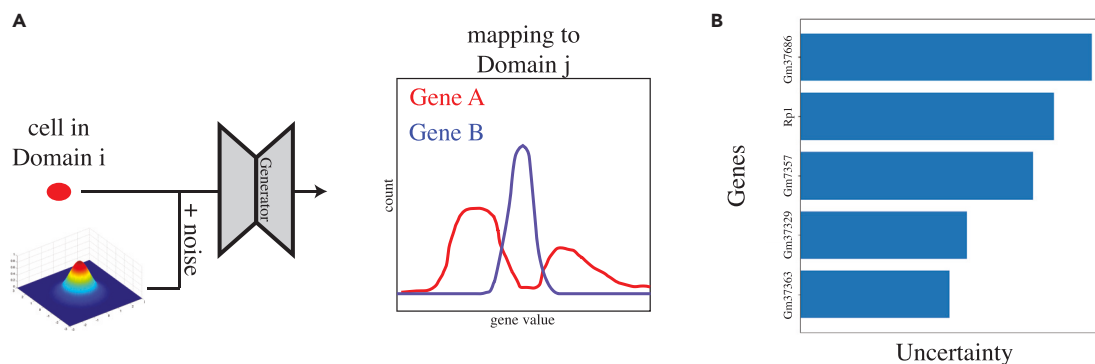
Figure 3. Design of the uncertainty quantification experiment
(A) A depiction of how scMMGAN can be used to quantify how much uncertainty is associated with the mapping to each gene. A particular cell is mapped from Domain *i* to Domain *j* along with various different noise samples. The mapped values of Gene A change significantly with the noise, while the mapped values of Gene B change little for this cell. We interpret this as a quantification of how much information there is about each gene in Domain *i*.
(B) The genes identified by scMMGAN to have the most uncertainty associated with the mapping, and thus have the least common information with the proteomic measurements in this dataset.

scMMGAN. Because the raw data are completely separable in the joint space, any downstream analysis would only be able to pick up on the difference between the two modalities and not the differences between cells within them. For an integrated analysis using information from both of them, we need the aligned output from scMMGAN (Figure 5B).

We analyze the scMMGAN alignment by taking the spatial RNA-seq, mapping it to the scRNA-seq space, and then clustering the generated scRNA-seq data (Figure 5B). We use spectral k-means clustering with a selected parameter of $k = 5$, and we then plot the clusters according to their original spatial coordinates. As we see, these scRNA-seq clusters preserve spatial patterns seen in the coordinates, demonstrating our ability to make new spatially informed conclusions by analyzing the generated scRNA-seq data in conjunction with the original spatial coordinates.

In Figure 5C, we look at the opposite mapping direction of taking the scRNA-seq data and generating spatial RNA-seq with it. By mapping scRNA-seq points to these generated coordinates, we can see spatial organization of particular cell types. For example, in this figure we plot the generated spatial coordinates for cells high in SLC2A1 and CDK11A and see spatial differentiation between these two types of cells. We show all of these clusters plotted in Figure 5D.

Now that we see scMMGAN has learned to map data between the two modalities in a way that preserves gene signals, we next compare this with the alternative alignment models. In Figure 6, we show the results of the learned mapping from spatial RNA-seq to scRNA-seq for each model. We first note that each model was able to generate a distribution that accurately matched the target distribution, an observation we will demonstrate quantitatively later. However, the alternative approaches to scMMGAN achieved this result by aligning a given spatial RNA-seq gene profile to an scRNA-seq observation that is very different.

In the first column of Figure 6, we plot the value of five genes across the original spatial coordinates in the spatial RNA-seq data. We then plot the generated scRNA-seq value of that gene for each spatial coordinate for each model in the subse-

quent columns, starting with scMMGAN. With FAM87b in the first row, we see that scMMGAN's generated values largely match the original spatial pattern, with some minimal changes that were necessary to match the target distribution as well. The CycleGAN matches much of the bottom half of the spatial coordinates, but the top half maps some coordinates that were low in the gene to scRNA-seq profiles that are high in the gene, and vice versa. The CMAE has even less correspondence between the original spatial RNA-seq value of the gene and the generated scRNA-seq value.

The preservation of signals by scMMGAN and not the other methods has important consequences for downstream analysis. In the first row of Figure 6, we show the values of SLC2A1. This gene encodes the glucose transporter type 1 (GLUT1) protein that is commonly upregulated in triple-negative breast cancers and is associated with high-grade tumors, having been previously shown to be a potential driver of metastasis in a broad array of breast and other cancers.[27–31] Notably, in the spatial data, SLC2A1 activity has a strong spatial pattern in which areas in the tissue express it highly. With scMMGAN mapping, the spatial observations with high SLC2A1 also have high expression in the generated scRNA-seq data. With the other models, however, the SLC2A1-high spatial observations are mapped to SLC2A1-low scRNA-seq cells. This important signal has been lost, and the downstream analysis that seeks to understand the differential spatial distribution and function will have lost this key gene signal. The scMMGAN mapping produces aligned data that best preserves the original signal.

The bottom row showing RER1 demonstrates another canonical situation motivating scMMGAN's correspondence loss. This gene is roughly bimodally distributed with equal numbers of observations high and low within it. Because flipping two populations is often as easy as introducing a single negative sign into a single weight in a neural network layer, CMAE maps all spatial coordinates high in the gene to scRNA-seq profiles low in the gene and vice versa. Only with scMMGAN's correspondence loss is one of these equally-easy-to-learn mappings specified as preferable, with the training objective significantly lower for
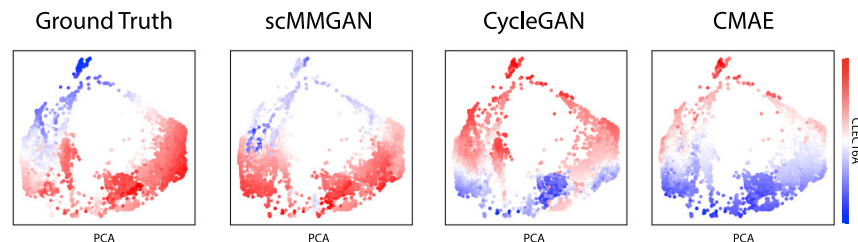
Figure 4. Results comparison from the ATAC-seq/RNA-seq experiment
Ground-truth values for held-out cells and the predictions for each model on the experiment mapping between ATAC and RNA sequencing. scMMGAN's output matches the ground truth most accurately compared with the other models, which inverted populations through the mapping. Coordinates shown are from the first two principal component analysis (PCA) dimensions.

the one that does not flip the populations as opposed to it being equal. This is further corroborated by the results of the quantitative experiments shown in Table 3.

Gene correlations. scMMGAN highlights the differences between the measurements of the two modalities by investigating the genes most highly correlated with a particular gene of interest in this system in both the original data and the generated data. Specifically, as the spatial data is an aggregate measurement of multiple cells in the same proximate area in the tissue (not a single cell), we can highlight some possibly spurious correlations by mapping them to the scRNA-seq space and recalculating the correlations.

Consider, for example, the glucose-transporter gene SLC2A1 that we have studied previously. If we look at the n most correlated genes in the original spatial data, any of them that have low correlation in the generated scRNA-seq data are candidates for spurious data artifacts. Similarly, any of the n most correlated genes in the generated scRNA-seq data that have low correlation in the original scRNA-seq are candidates for novel associations found by the scMMGAN.

Choosing $n = 700$ and defining low correlation in the other space as being less than 0.1, we obtain a list of six genes that have spurious correlations: FRAT2, CAMK2A, CANX, LRRC66, ZMIZ1-AS1, and MTRNR2L8. We then have the following five genes that have been discovered by scMMGAN: AC092115.3, P2RX7, CCDC93, UTP25, and BBS10.

Among these genes whose correlation to the glucose transmitter SLC2A1 is discovered by scMMGAN, we see P2RX7, which has been identified in the literature as a precursor to glucose transporters.[32] This provides corroborating support in favor of the scMMGAN-discovered gene correlations.

DISCUSSION

In this work we demonstrated that scMMGAN can align data from related experiments but different modalities in a way that best preserves the properties of the original cells through learned mapping. The addition of the correspondence loss in scMMGAN's architecture resolves the ambiguity created by only stating a distribution-level loss in learning a mapping. This holds across a wide array of data types and modalities, distribution shapes, and other settings that arise in practical biological experiments.

We have shown how scMMGAN can be used to measure uncertainty in the mapping and use injected stochasticity to gauge which information is unique to one of the modalities and which information is common between them. This can be used to not only answer questions about the cellular samples in an experi-

ment but also to answer questions about the technologies and modalities themselves in terms of their strengths and weaknesses.

Furthermore, we have shown how scMMGAN can be used to identify spurious correlations found in one modality as artifactual results, as opposed to real findings. Similarly, we demonstrated scMMGAN's ability to identify novel correlations that are not visible in an individual modality but become apparent when the data are mapped to another modality. In these ways, scMMGAN can be added to traditional analysis pipelines to uncover further insights from complicated, multi-modal experimental data.

Limitations

There are limitations to the proposed approach that bear mentioning. Although GANs have been useful in mapping distributions, they suffer from key drawbacks. First, they are difficult to train because of the adversarial losses, which can lead to instability.[33] This instability means that the model can deteriorate from effective to ineffective quickly across training iterations. Second, they often suffer from mode collapse because they are not penalized by distribution-level losses to match the entire distribution.[34] The additional correspondence loss does not worsen these issues. We informally observe that early stopping, as a regularization, as well as our geometric loss helps mitigate these effects, but these effects may still be present in some contexts. Additionally, with our framework based on pairwise generators in each mapping direction, the number of generators necessary grows quadratically. This means that for a large number of input modalities to align, the networks would have to be made small or would have to be trained separately. Finally, our geometry-based loss is not completely "plug-and-play" in the sense that we still require a choice of distance between data points. In cellular data, we used Euclidean distance to compute the manifold. However, in other contexts, such as two image types, more complicated measures such as the structural similarity index may be used.[35,36]

For this reason, we encourage continued evaluation of aligned results through external verification measures. For example, in this work we verified that known signals across genes and across cells are still preserved in the aligned data. Moreover, we point out that the novel gene correlations found by scMMGAN are potential discoveries that should be further investigated with experiments specifically designed for this aim.

EXPERIMENTAL PROCEDURES

In this section we further expand on the model, experimental regimes, and implementations used in this work.

Table 2. Results from the ATAC-seq/RNA-seq experiment

ATAC-seq	scMMGAN	CycleGAN	CMAE
MMD $(x_1, G(x_2))$	**0.033** ± 0.001	**0.033** ± 0.000	0.038 ± 0.000
MMD $(G(x_1), x_2)$	**0.031** ± 0.000	0.032 ± 0.000	0.051 ± 0.003
Correlation $(x_1, G(x_2))$	**0.313** ± 0.025	0.024 ± 0.140	−0.014 ± 0.108
Correlation $(G(x_1), x_2)$	**0.358** ± 0.016	0.034 ± 0.225	0.020 ± 0.111

Evaluation of each model on the ATAC-seq/RNA-seq data. The MMDs for each model are close, as each models the ground truth at a whole-distribution level. scMMGAN is the only model whose predictions preserve the known correlation, however, because its alignment is also accurate point-wise. The best or tied-for-best values are in boldface.

Resource availability

Lead contact

The lead contact is Smita Krishnaswamy (smita.krishnaswamy@yale.edu).

Materials availability

There are no newly generated materials.

Data and code availability

An implementation of the scMMGAN model written in Python and Tensorflow, which can be run on any user-loaded datasets, is available at https://github.com/KrishnaswamyLab/scMMGAN. Direct further data availability inquiries to the lead contact.

Biological methods

This section describes the methods used to acquire the dataset of triple-negative breast cancer investigated in this paper.

Animal studies

All experiments were approved by and conducted in accordance with the National Health and Medical Research Council Statement on Animal Experimentation, the requirements of New South Wales State Government legislation, and the rules for animal experimentation of the Biological Testing Facility of the Garvan Institute and the Victor Chang Cardiac Research Institute (protocol #18/12).

Ten NOD/SCID mice at 6–8 weeks of age were purchased from Australian Bioresources (ABR). Mice were 8–9 weeks of age at time of injections. MDA-MB-231-GFP cells (1×10^6 cells) were prepared in 25 μL 20% Matrigel (BD Matrigel Matrix Growth Factor Reduced)/serum-free medium and injected orthotopically into the inguinal mammary fat pads. After 9.5 weeks, animals underwent survival surgery to remove the primary tumors (700–1,000 mm^3), which were subsequently split into three parts: one chunk was used for scRNA-seq, one chunk was formalin fixed and embedded in paraffin for future analysis, and one chunk was frozen in optimal cutting temperature compound for spatial transcriptomics analysis. Animals were housed for 2 more weeks, after which the liver, lymph nodes, and lungs were harvested for analysis of metastatic cells in those tissues.

scRNA-seq preparation and analysis

Primary tumors, lungs, livers, and lymph nodes were chopped into small pieces, then incubated at 37°C for 40 min on a rotary shaker in DMEM/F12 containing collagenase A (300 U) and hyaluronidase (100 U). Following digestion, cell suspensions were pelleted, the DMEM removed, washed with PBS 2% (v/v) fetal bovine serum (FBS), and resuspended in 0.15% + 10% DNaseI trypsin for 1 min. Trypsin was quenched with 2% FBS/DMEM. Cells were resuspended in FACS buffer (2% [v/v] FBS in PBS). GFP$^+$ alive tumor cells were sorted and collected, and scRNA-seq was performed using Chromium 10X technology.

Spatial transcriptomics

Spatial transcriptomics was performed according to the published protocol.[37]

Training objectives

Here we elaborate on the training objectives used in the scMMGAN framework learning. We define the formulation considering a pair of domains, with the definitions extending to multiple domains accordingly. It is composed of distinct GAN networks, each with a generator network G with input X and output X'. We call each generator a mapping from the input domain to the output, or target, domain. Each generator attempts to make its output $G(X)$ indistinguishable by D from X'. Denote the two datasets X_1 and X_2. Let the generator mapping from X_1 to X_2 be G_{12} and the generator mapping from X_2 to X_1 be G_{21}. The discriminator that tries to separate true samples from X_1 from the generated output of $G_{21}(X_2)$ is D_1, and the discriminator that tries to separate true samples from X_2 from generated samples from $G_{12}(X_1)$ is D_2.

The loss for G_1 on minibatches x_1 and x_2 is

$$x_{12} = \mathbf{G_{12}}(x_1)$$
$$x_{121} = \mathbf{G_{21}}(x_{12})$$
$$L_r = L_{reconstruction} = \text{MSE}(x_1, x_{121})$$
$$L_d = L_{discriminator} = - \mathbb{E}_{x_1 \sim P_{X_1}}[\log \mathbf{D_2}(x_{12})],$$
$$L_c = L_{correspondence} = L(x_1, x_{12})$$
$$L_{\mathbf{G_1}} = \lambda_r L_r + \lambda_d L_d + \lambda_c L_c$$

where MSE is the mean-squared error and L is the correspondence loss discussed previously. The hyperparameters λ_r, λ_d, and λ_c are chosen to balance the reconstruction, discriminator, and correspondence losses. These can be chosen by default to be $\lambda_r = \lambda_d = \lambda_c = 1$, but λ_c increased if the observed correspondences are low and λ_r increased if the observed reconstructions are not accurate.

Similarly, the loss for G_2 is

$$x_{21} = \mathbf{G_{21}}(x_2)$$
$$x_{212} = \mathbf{G_{12}}(x_{21})$$
$$L_r = \text{MSE}(x_2, x_{212})$$
$$L_d = - \mathbb{E}_{x_2 \sim P_{X_2}}[\log \mathbf{D_1}(x_{21})].$$
$$L_c = L(x_2, x_{21})$$
$$L_{\mathbf{G_2}} = \lambda_r L_r + \lambda_d L_d + \lambda_c L_c$$

The losses for D_1 and D_2 are

$$L_{\mathbf{D_1}} = - \mathbb{E}_{x_1 \sim P_{X_1}}[\log \mathbf{D_1}(x_1) + \log \mathbf{D_1}(x_{121})]$$
$$- \mathbb{E}_{x_2 \sim P_{X_2}}[\log(1 - \mathbf{D_1}(x_{21}))]$$
$$L_{\mathbf{D_2}} = - \mathbb{E}_{x_2 \sim P_{X_2}}[\log \mathbf{D_2}(x_2) + \log \mathbf{D_2}(x_{212})].$$
$$- \mathbb{E}_{x_1 \sim P_{X_1}}[\log(1 - \mathbf{D_2}(x_{12}))]$$

Calculation of correspondence loss

For the following notation, consider one of the datasets x_1 and its representation after being mapped to the other domain, $G(x_1)$. First, matrices of pairwise distances D_{x_1} and $D_{G(x_1)}$ are constructed.

These are then transformed into matrices of pairwise affinities with an inverse-distance kernel $k(x_i, x_j) = \max(0, 1 - ||x_i - x_j|| 1 / \sigma_i)$, where σ_i is a k-nearest neighbor adaptive bandwidth. This kernel was necessary, as the standard Gaussian kernel suffered from gradients that saturated and suppressed learning. With this kernel, we are able to effectively perform diffusion geometry learning through gradient descent.

These affinity matrices are transformed into transition probability matrices P_{x_1} and $P_{G(x_1)}$ through row normalization. Powering these matrices t times then represents taking t steps forward in the Markov chain. Let the eigenvectors of these two matrices be φ_{x_1} and $\varphi_{G(x_1)}$, respectively. The ith row of φ_{x_1} and $\varphi_{G(x_1)}$ represents the diffusion coordinates of the point x_{1i} in the original space and $G(x_{1i})$ in the generated space, respectively. Because we are only seeking to match low-frequency structure of the data, we use only the first n_e
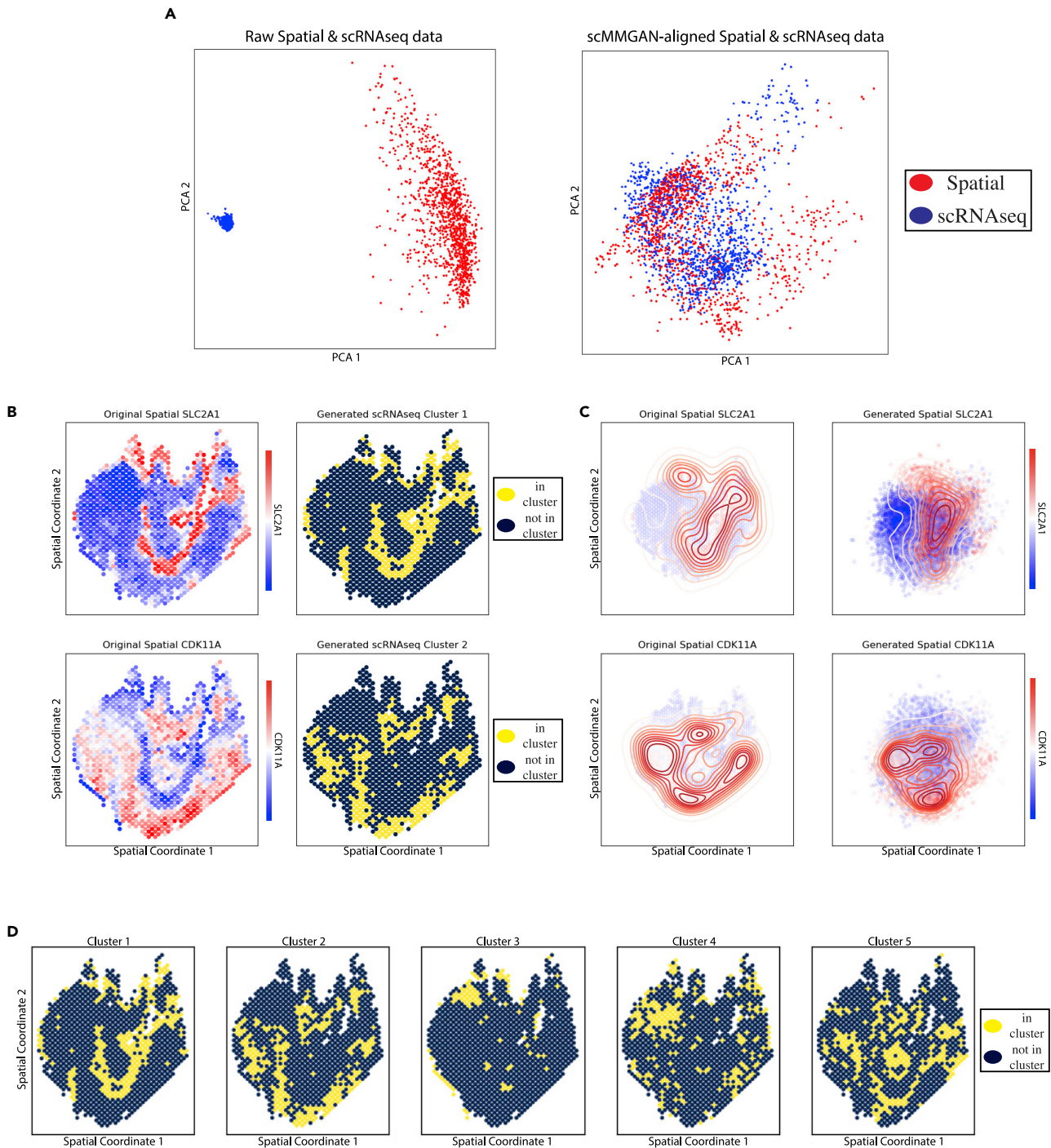
Figure 5. Analysis of scMMGAN alignment and clusters on the triple-negative breast cancer dataset

(A) Plotted are the PCA coordinates of the gene expression values from the two distributions. In the raw data, the spatial RNA-seq and scRNA-seq are not directly comparable, as they are entirely separable. After mapping with scMMGAN, they are aligned and comparable with downstream analysis.

(B) Mapping spatial RNA-seq to scRNA-seq, clustering the generated scRNA-seq values, and then plotting the cluster by the measured spatial coordinate on the x axis an y axis.

(C) Generated spatial RNA-seq data from scRNA-seq, including generated spatial coordinates. Same coordinates as previous plot.

(D) All generated clusters mapped to the spatial RNA-seq space. Same coordinates as previous plots.
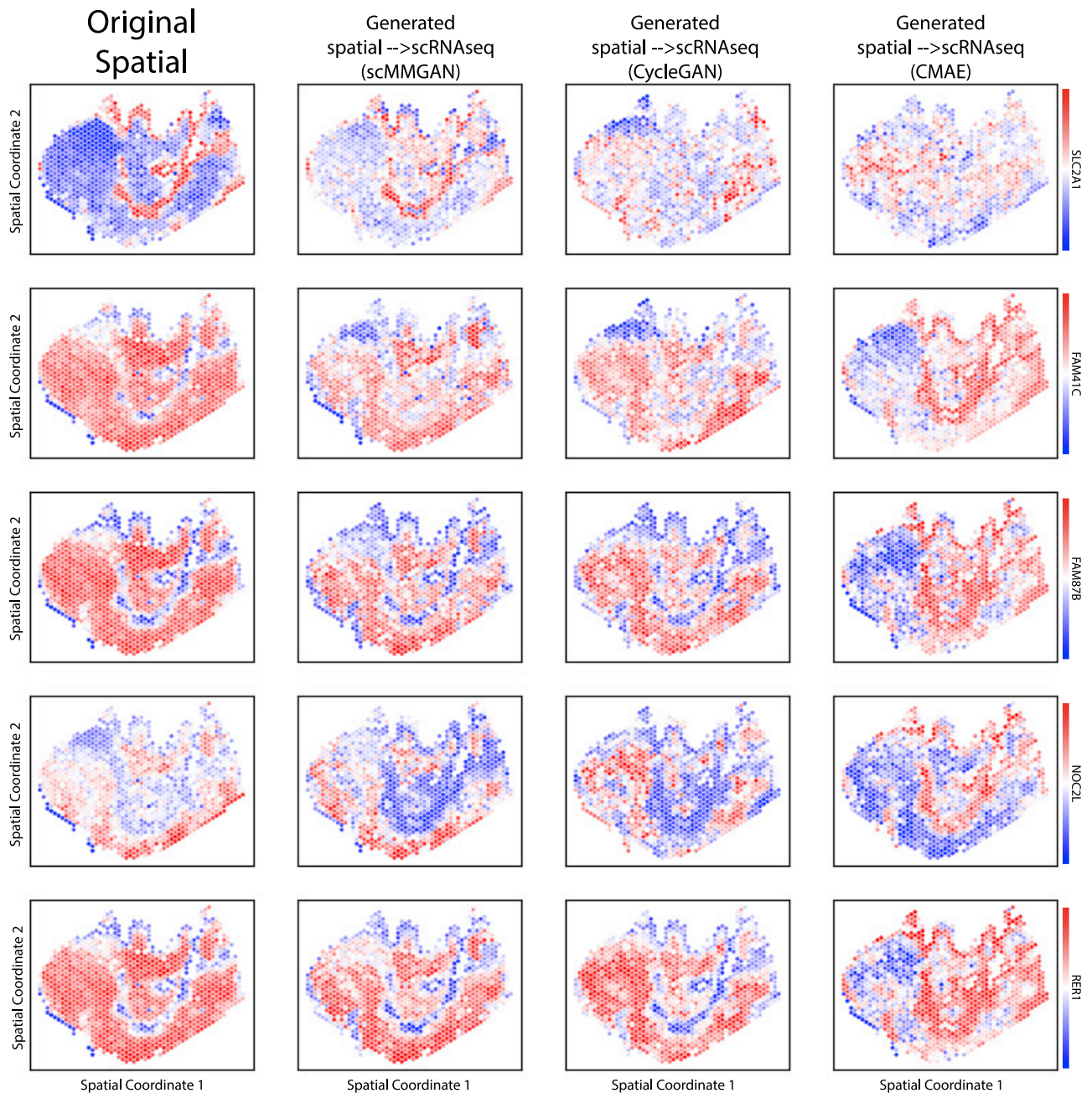
Figure 6. Generated scMMGAN expression value results plotted on the spatial coordinates
The x axis and y axis plotted are the raw measured spatial coordinates from the spatial RNA-seq. The color is expression value, where we compare the original spatial RNA-seq of a gene with each generated scRNA-seq value of that gene for each method, showing scMMGAN best aligns the original and generated values.

eigenvectors of the data as an approximation. The eigenvectors are then re-scaled to be between −1 and 1.

We also perform a check before comparing the eigenvectors of the original data φ_x and those of the generated data $\varphi_{G(x)}$. Because the direction of the ei-genvectors can be switched, two datasets with equivalent intrinsic geometry could have eigenvectors that are either highly correlated or highly anticorre-lated. To combat this, we calculate the correlation of each pair of eigenvectors before computing the loss and whether the correlation is below a threshold $\varepsilon_r = -0.5$, then we multiply the values of $\varphi_{G(x)}$ by −1 before computing the loss. We then also compare the eigenvectors at different scales by sum-

ming adjacent vectors and comparing the new combined representations that have half the number of vectors.

Noise-augmented model
Here we detail the noise-augmented model used in the section about distin-guishing unique and common information. The core idea is that by providing additional noise as input, the model will be able to use the stochasticity when necessary or ignore it if not. In other words, some generated values will have more certainty behind them in the model and others greater uncer-tainty. We experiment with this notion by introducing a slight modification of

Table 3. Results from the triple-negative breast cancer experiment

scRNA-seq –> Spatial	scMMGAN	CycleGAN	CMAE
	MMD $(G(x_1), x_2)$		
Sample 1	**0.072** ± 0.003	**0.072 ± 0.001**	0.076 ± 0.002
Sample 2	**0.071** ± 0.001	0.072 ± 0.002	0.075 ± 0.002
Sample 3	**0.072** ± 0.002	**0.072** ± 0.001	0.075 ± 0.002
Sample 4	**0.071** ± 0.002	0.072 ± 0.004	0.076 ± 0.002
	MMD $(x_1, G(x_2))$		
Sample 1	0.076 ± 0.002	0.080 ± 0.003	**0.075 ± 0.001**
Sample 2	0.085 ± 0.002	**0.074** ± 0.002	0.078 ± 0.003
Sample 3	**0.081** ± 0.002	0.082 ± 0.001	0.087 ± 0.004
Sample 4	0.079 ± 0.001	**0.076** ± 0.001	0.086 ± 0.006
	MSE $(x_1, G(x_2))$		
Sample 1	**0.987** ± 0.128	2.001 ± 0.525	2.021 ± 0.323
Sample 2	**0.995** ± 0.126	1.934 ± 0.235	1.771 ± 0.521
Sample 3	**0.887** ± 0.026	2.097 ± 0.548	1.591 ± 0.728
Sample 4	**1.029** ± 0.017	1.932 ± 0.353	1.823 ± 0.288
	MSE $(G(x_1), x_2)$		
Sample 1	**0.931** ± 0.101	1.972 ± 0.515	2.031 ± 0.275
Sample 2	**0.970** ± 0.108	1.931 ± 0.195	1.843 ± 0.536
Sample 3	**0.878** ± 0.024	2.068 ± 0.556	1.609 ± 0.756
Sample 4	**0.985** ± 0.010	1.905 ± 0.366	1.872 ± 0.287

Quantitative measurement of how well the generated distributions match the target distribution (MMD) and how well they preserve correspondence with the original input distribution (MSE). While all methods match the target distribution reasonably (top two sections), only scMMGAN minimally alters the points in the alignment (bottom two sections). Statistics reported on both mapping directions and across five independent trials. The best or tied-for-best values are in boldface.

the scMMGAN framework: additional random input. We calculate the generator's output as $G(x_i; z)$, where $z \sim N(0, 1) \in IR^D$, concatenating a draw from an isotropic normal distribution with the original input. The reconstruction and correspondence losses are then calculated as usual with just x_i. This allows the model to create more stochasticity around regions of the space where there is not enough information to pin down precisely the correct alignment, while it can ignore the noise and create a deterministic mapping in regions of the space where there is enough information.

Invariance and risk

In this section we connect our model in the domain alignment setting to existing literature on invariance and risk minimization.[38] Consider the domain alignment task as drawing a dataset $d_i^e \sim D^e$ from a distribution $P(D^e)$ where the environment e determines how observations manifest d_i in D_e. The domains X_1 and X_2 in our setting are drawn from $P(D^e)$, then the datasets $x_1^i, i = 1 \ldots n_1$ and $x_2^i, i = 1 \ldots n_2$ are drawn from the distributions X_1 and X_2. Thus, we have two sources of randomness from sampling to consider in learning the desired mappings G, one from the points sampled from the distribution and the other from the distribution sampled from the distribution over settings. We want to minimize the risk of alignment:

$$\mathbb{E}_P \mathbb{E}_{X_1, X_2} = L\left(G_{12}(x_1^i), x_2^i\right) + L\left(x_2^i, G_{21}(x_1^i)\right).$$

The inadequacy of solely a cycle-consistency loss should be obvious from this formation. While minimizing the cycle-consistency loss can optimize performance with respect to the expectation over X_1 and X_2, the formulation is equivalent for all datasets drawn from P. Thus, it is incapable of resolving correlations that are structurally related to the datasets versus those arising spuriously from sampling. These ideas are key to the extension beyond just using cycle consistency in the construction of domain mapping networks.

Diffusion maps

Diffusion maps define a process of Markovian diffusion over a dataset, whereby a set of local affinities capture the intrinsic data geometry as quantified by diffusion distances.[18] They operate on a matrix of pairwise distances that is transformed into a matrix of pairwise affinities, here via the commonly used Gaussian kernel $k(x_i, x_j) = \exp\{-\|x_i - x_j\|^2 / \varepsilon\}$. A Markov chain transition matrix over the dataset P is constructed from the pairwise affinity matrix A

via $P = \left(\sum_j A_{ij}\right)^{-1} A$. Powering the matrix P^t represents taking t forward steps

in the Markov chain. Diffusion maps are then defined as $\Psi_t(x) = (\lambda_1^t \psi_1, \ldots, \lambda_l^t \psi_l)$ where λ_i and ψ_i are the ith eigenvalue and eigenvector of P, and l is a hyperparameter of the number of top eigenvectors to use. The diffusion map coordinates form a space where the Euclidean distance between points approximates the diffusion distance between those points.

In previous work, diffusion operators have been used in the context of multimodal data integration.[39] This has been done for the related tasks of visualizing and denoising, rather than mapping between, the datasets. The approach there differs from ours in that it relies on combining diffusion operators from different modalities through algebraic operations as opposed to our method, which integrates them into a broader deep-learning framework.

The diffusion maps are a key foundational notion used in the construction of the data geometry loss.

Geometry-preserving correspondence loss

We now further elaborate on a few points about the geometry-preserving correspondence loss introduced in this paper. We only enforce the correspondence loss on the first eigenvectors because this ensures that basic low-frequency signals are largely aligned while still allowing the flexibility of changing high-frequency signals that are more likely to be idiosyncratic to each domain. In practice we find using 10–20 eigenvectors works best.

We note that any changes to the data geometry would cause a mismatch here, and thus the ideal alignment would not drive this term in the objective all the way to zero unless the two datasets being aligned have identical geometry. Despite the goal not being zero correspondence loss, there are many different mappings that achieve comparably low GAN losses, and among them the ones with lower correspondence losses are preferable. This is why using it as a regularization to lightly guide the transformation in addition to the GAN loss can achieve the best performance overall.

Architecture and baselines

We compare scMMGAN with alternative baseline deep-learning models used for alignment of this type: a CycleGAN, to motivate the need for the correspondence loss by showing the improper alignments obtained without it;[9] and CMAE, an autoencoder-based model that uses separate encoders/decoders that learn to map into a shared space and then generates by crossing the encoder of one domain with the decoder of another.[21] These alternative methods use distribution-level losses to ensure the generated distribution matches the target distribution, but do not impose any loss on the representation of a point and its representation in the aligned domain. As a result, they can produce alignments that unnecessarily invert signals and change values of individual points.

With scMMGAN, we use a generator consisting of three internal layers of 128, 256, and 512 neurons with batch norm and leaky rectified linear unit activations after each layer, and a discriminator consisting of three internal layers with 1,024, 512, and 256 neurons with the same batch norm and activations except with minibatching after the first layer.[33,40] We use a correspondence loss coefficient of 10, cycle-loss coefficient of 1, learning rate of 0.0001, and batch size of 256. As preprocessing steps prior to running each model on this dataset, we correct for dropout with the manifold smoothing method MAGIC,[41] zero-center and unit scale each dimension, and reduce to 50 principal components. We use these architectures and hyperparameters in all subsequent experiments except where otherwise stated.

AUTHOR CONTRIBUTIONS

M.A. and S.K. designed the model and computational experiments, and wrote the text of the manuscript. M.A. wrote the code and performed the computational experiments. S.E.Y., A.V., and C.L.C. helped with the biological analysis of the breast cancer dataset. S.Y., B.P.S.J., and C.L.C. performed experiments to generate the breast cancer dataset.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Ozsolak, F., and Milos, P.M. (2011). Rna sequencing: advances, challenges and opportunities. Nat. Rev. Genet. *12*, 87–98.

2. Baek, S., and Lee, I. (2020). Single-cell atac sequencing analysis: from data preprocessing to hypothesis generation. Comput. Struct. Biotechnol. J. *18*, 1429–1439.

3. Forcato, M., Nicoletti, C., Pal, K., Livi, C.M., Ferrari, F., and Bicciato, S. (2017). Comparison of computational methods for hi-c data analysis. Nat. Methods *14*, 679–685.

4. Park, P.J. (2009). Chip–seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. *10*, 669–680.

5. Stoeckius, M., and Peter, S. (2017). Cite-seq.

6. Moretti, J.L., Hauet, N., Caglar, M., Rebillard, O., and Burak, Z. (2005). To use mibi or not to use mibi? that is the question when assessing tumour cells. Eur. J. Nucl. Med. Mol. Imag. *32*, 836–842.

7. Cheung, R.K., and Utz, P.J. (2011). Cytof—the next generation of cell detection. Nat. Rev. Rheumatol. *7*, 502–503.

8. Kim, T., Cha, M., Kim, H., Lee, Jung K., and Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. In International Conference on Machine Learning (PMLR), pp. 1857–1865.

9. Zhu, J.-Y., Park, T., Isola, P., and Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.

10. Almahairi, A., Rajeshwar, S., Sordoni, A., Bachman, P., and Courville, A. (2018). Augmented cyclegan: learning many-to-many mappings from unpaired data. In International Conference on Machine Learning (PMLR), pp. 195–204.

11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. Commun. ACM *63*, 139–144.

12. Amodio, M., and Krishnaswamy, S. (2018). Magan: aligning biological manifolds. In International Conference on Machine Learning (PMLR), pp. 215–223.

13. Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In International conference on machine learning (PMLR), pp. 214–223.

14. Gulrajani, I., Ahmed, F., Martin, A., Vincent, D., and C Courville, A. (2017). Improved training of wasserstein gans. Adv. Neural Inf. Process. Syst. *30*.

15. Wu, J., Huang, Z., Thoma, J., Acharya, D., and Van Gool, L. (2018). Wasserstein divergence for gans. In Proceedings of the European Conference on Computer Vision (ECCV)), pp. 653–668.

16. Stanley, J.S., III, Scott, G., Wolf, G., and Krishnaswamy, S. (2020). Harmonic alignment. In Proceedings of the 2020 SIAM International Conference on Data Mining (SIAM), pp. 316–324.

17. Coifman, R.R., and Lafon, S. (2006). Diffusion maps. Appl. Comput. Harmon. Anal. *21*, 5–30.

18. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., and Zucker, S.W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc. Natl. Acad. Sci. USA *102*, 7426–7431.

19. De la Porte, J., Herbst, B.M., Hereman, W., and Van Der Walt, S.J. (2008). An introduction to diffusion maps. In Proceedings of the 19th symposium of the pattern recognition association of South Africa (PRASA 2008), Cape Town, South Africa, pp. 15–25.

20. Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell *183*, 1665–1681.e18.

21. Yang, K.D., Belyaeva, A., Venkatachalapathy, S., Damodaran, K., Katcoff, A., Radhakrishnan, A., Shivashankar, G.V., and Uhler, C. (2021). Multi-domain translation between single-cell imaging and sequencing data using autoencoders. Nat. Commun. *12*, 31.

22. Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H.-P., Schölkopf, B., and Smola, A.J. (2006). Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics *22*, e49–e57.

23. Gao, R., Liu, F., Zhang, J., Han, B., Liu, T., Niu, G., and Sugiyama, M. (2020). Maximum mean discrepancy is aware of adversarial attacks. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.11415.

24. Amodio, M., van Dijk, D., Srinivasan, K., Chen, W.S., Mohsen, H., Moon, K.R., Campbell, A., Zhao, Y., Wang, X., Venkataswamy, M., et al. (2019). Exploring single-cell data with deep multitasking neural networks. Nat. Methods *16*, 1139–1145.

25. Bowne, S.J., Daiger, S.P., Hims, M.M., Sohocki, M.M., Malone, K.A., McKie, A.B., Heckenlively, J.R., Birch, D.G., Inglehearn, C.F., Bhattacharya, S.S., et al. (1999). Mutations in the rp1 gene causing autosomal dominant retinitis pigmentosa. Hum. Mol. Genet. *8*, 2121–2128.

26. Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral t cell exhaustion. Nat. Biotechnol. *37*, 925–936.

27. Hussein, Y.R., Bandyopadhyay, S., Semaan, A., Ahmed, Q., Albashiti, B., Jazaerly, T., Nahleh, Z., and Ali-Fehmi, R. (2011). Glut-1 expression correlates with basal-like breast cancer. Transl. Oncol. *4*, 321–327.

28. Min, K.-W., Kim, D.-H., Son, B.K., Moon, K.M., Kim, S.M., Intazur Rahaman, M., Kim, S.W., Kim, E.K., Kwon, M.J., Koh, Y.W., and Oh, I.H. (2021). High slc2a1 expression associated with suppressing cd8 t cells and b cells promoted cancer survival in gastric cancer. PLoS One *16*, e0245075.

29. Noguchi, Y., Saito, A., Miyagi, Y., Yamanaka, S., Marat, D., Doi, C., Yoshikawa, T., Tsuburaya, A., Ito, T., and Satoh, S. (2000). Suppression of facilitative glucose transporter 1 mrna can suppress tumor growth. Cancer Lett. *154*, 175–182.

30. Krzeslak, A., Wojcik-Krowiranda, K., Forma, E., Jozwiak, P., Romanowicz, H., Bienkiewicz, A., and Brys, M. (2012). Expression of glut1 and glut3 glucose transporters in endometrial and breast cancers. Pathol. Oncol. Res. *18*, 721–728.

31. Wu, Q., Ba-Alawi, W., Deblois, G., Cruickshank, J., Duan, S., Lima-Fernandes, E., Haight, J., Tonekaboni, S.A.M., Fortier, A.M., Kuasne, H., et al. (2020). Glut1 inhibition blocks growth of rb1-positive triple negative breast cancer. Nat. Commun. *11*, 4205–4212.

32. Arguin, G., Bourzac, J.F., Placet, M., Molle, C.M., Paquette, M., Beaudoin, J.F., Rousseau, J.A., Lecomte, R., Plourde, M., and Gendron, F.-P. (2017). The loss of p2x7 receptor expression leads to increase intestinal glucose transit and hepatic steatosis. Sci. Rep. *7*, 12917–13016.

33. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. Adv. Neural Inf. Process. Syst. *29*, 2234–2242.

34. Liu, K., Tang, W., Zhou, F., and Qiu, G. (2019). Spectral regularization for combating mode collapse in gans. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6382–6390.

35. Brunet, D., Vrscay, E.R., and Wang, Z. (2012). On the mathematical properties of the structural similarity index. IEEE Trans. Image Process. *21*, 1488–1499.

36. Hore, A., and Ziou, D. (2010). Image Quality Metrics: Psnr vs. Ssim. In In 2010 20th International Conference on Pattern Recognition (IEEE), pp. 2366–2369.

37. Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., et al. (2019 Oct). High-definition spatial transcriptomics for in situ tissue profiling. Nature methods *16*, 987–990.

38. Arjovsky, M., Bottou, L., Gulrajani, I., and Lopez-Paz, D. (2019). Invariant Risk Minimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1907.02893.

39. Kuchroo, M., Godavarthi, A., Tong, A., Wolf, G., and Krishnaswamy, S. (2021). Multimodal data visualization and denoising with integrated diffusion. In 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP) (IEEE), pp. 1–6.

40. Santurkar, S., Tsipras, D., Ilyas, A., and Mądry, A. (2018). How does batch normalization help optimization? In Proceedings of the 32nd international conference on neural information processing systems, pp. 2488–2498.

41. Van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell *174*, 716–729.e27.