Enhancing Multimodal Spatial Omics Analysis with Contrastive Learning

Vikas Yadav and Swastik Singhal
Supervisor: Prof. Hamim Zafar

Abstract

Spatial omics is an emerging field that combines spatial and molecular data to analyze cell-cell communication, tissue organization, and disease markers within their native environment. This project addresses challenges in multimodal spatial omics by enhancing GEM-VAE with contrastive learning and benchmarking it against other established models. Simulated datasets and real-world data were utilized to validate the improvements, demonstrating a meaningful advance in the integration and analysis of multi-omics data.

1 Introduction

1.1 Spatial Omics: Bridging Molecular Biology and Spatial Data

Advancements in high-throughput sequencing technologies have revolutionized molecular biology, allowing for comprehensive profiling of genomes, transcriptomes, and proteomes. However, traditional omics approaches often lack spatial context, which is crucial for understanding the complex architecture and functional heterogeneity of tissues. **Spatial omics** bridges this gap by integrating molecular profiling with spatial information, enabling the study of biological systems at unprecedented resolution [10].
Unlike conventional genomics, spatial omics technologies capture the precise location of biomolecules within tissues, providing invaluable insights into:

- **Cellular Interactions and Communication**: Understanding how cells interact and communicate within their native microenvironment.
- **Tissue Architecture and Development**: Elucidating the spatial organization of cells during tissue development and differentiation.
- **Spatial Organization of Disease Markers**: Mapping the distribution of disease-associated biomarkers, aiding in diagnostics and personalized treatment planning.

Key techniques in spatial omics include:

- **Spatial Transcriptomics**: Measures RNA expression across tissue sections with spatial resolution, enabling the identification of gene expression patterns *in situ* [11].
- **Spatial Proteomics**: Maps proteins within tissues, revealing functional protein networks and post-translational modifications [6].
- **Spatial Epigenomics**: Analyzes DNA modifications, chromatin accessibility, and epigenetic markers at specific genomic loci within the spatial context of tissues [2].

These technologies have profound implications for fields such as developmental biology, neuroscience, and oncology, where spatial context is essential for understanding complex biological processes.

1.2 The Need for Multi-Omics Integration in Spatial Context

Multi-omics integration involves the comprehensive analysis of diverse datasets—such as genomics, transcriptomics, proteomics, and metabolomics—to provide a holistic view of biological systems [7]. Integrating these datasets within a spatial framework enhances our ability to link molecular variations across modalities and to understand how they collectively influence cellular functions and phenotypes.
In this research, we focus on integrating:

- **Transcriptomics**: Captures RNA-level gene activity, providing insights into gene expression dynamics and regulatory mechanisms.
- **Proteomics**: Highlights protein interactions, post-translational modifications, and spatial distributions, offering a direct link to cellular function.

By combining spatial transcriptomics and proteomics data, we aim to unravel the complex interplay between gene expression and protein function within the spatial context of tissues.

1.3 Current Computational Methods for Spatial Multi-Omics Integration

The integration of spatial multi-omics data poses significant computational challenges due to the high dimensionality, heterogeneity, and technical variability inherent in these datasets. Several computational frameworks have been developed to address these challenges:

- **SpatialGlue**: A graph-based integrative framework that leverages spatial information to align and integrate multiple omics datasets [9]. It constructs a spatial graph to model the relationships between neighboring spots or cells, enhancing integration accuracy by preserving spatial continuity.
- **STAGATE (Spatial Transcriptomics Analysis via Graph Attention Networks)**: Utilizes graph attention networks to capture spatial gene expression patterns and to impute missing data [1]. STAGATE models the spatial dependencies between cells, improving the identification of spatial domains and cell types.
- **totalVI**: A probabilistic framework based on variational autoencoders (VAEs) designed for the joint analysis of paired single-cell transcriptomics and proteomics data [5]. Although not inherently spatial, totalVI can be extended to spatial data by incorporating spatial coordinates as additional inputs.
- **GEM-VAE (Generative Embedding Model Variational Autoencoder)**: A deep learning model that captures the underlying structure of multi-omics data through latent embeddings [3]. By integrating spatial information, GEM-VAE can learn representations that reflect both molecular profiles and spatial relationships.

Despite these advancements, several limitations persist:

- **Scalability**: Many models struggle with large-scale datasets due to computational complexity.
- **Integration Accuracy**: Aligning datasets from different modalities remains challenging due to batch effects and differing feature spaces.
- **Spatial Resolution**: Capturing fine-grained spatial relationships requires sophisticated modeling of spatial dependencies.
- **Benchmarking Datasets**: The lack of standardized, ground-truth datasets hampers the objective evaluation and comparison of computational methods.

1.4 Our Contribution

To address these challenges, we propose:

- **Development of Synthetic Benchmark Datasets**: Creating realistic synthetic datasets with known ground truth to facilitate rigorous benchmarking of spatial multi-omics integration methods.
- **Enhancement of GEM-VAE with Contrastive Loss**: Incorporating contrastive loss functions into GEM-VAE to improve its ability to integrate spatial transcriptomics and proteomics data, enhancing the preservation of both molecular and spatial relationships.
- **Comprehensive Evaluation**: Benchmarking the enhanced GEM-VAE against existing methods such as SpatialGlue, STAGATE, and totalVI, using the synthetic datasets and real-world data to demonstrate its effectiveness and scalability.

Through these contributions, our research aims to advance the field of spatial multi-omics by providing robust tools and resources for data integration and analysis.

2 Problem Statement

Spatial omics technologies have revolutionized our understanding of biological systems by enabling the high-resolution mapping of molecular profiles within their native spatial context. Techniques such as spatial transcriptomics and proteomics provide comprehensive insights into the spatial distribution of genes and proteins, respectively, across tissues and organs. Despite their transformative potential, several significant challenges impede the full exploitation of spatial omics data:

- **Limited Computational Tools for Spatial Multi-omics Integration**: Currently, there is a scarcity of computational methods capable of effectively handling and integrating multi-omics data (e.g., RNA and protein expression) within a spatial framework. The complex and high-dimensional nature of spatial multi-omics datasets demands advanced analytical tools that can capture intricate biological relationships while accounting for spatial heterogeneity.
- **Lack of Rigorous Benchmarking Datasets**: Existing methods, including prominent ones like *SpatialGlue*, have not been thoroughly evaluated due to the absence of robust and realistic benchmarking datasets. The lack of datasets with known ground truth hampers the objective assessment and comparison of different computational approaches, slowing down methodological advancements in the field.
- **Challenges in Data Quality and Complexity**: Spatial omics data often suffer from technical noise, batch effects, and missing values (e.g., dropout events in single-cell RNA sequencing). Developing methods that are robust to these issues while maintaining sensitivity to true biological signals remains a critical challenge.
- **Scalability and Computational Efficiency**: As spatial omics technologies advance, datasets are becoming increasingly large and complex. There is a pressing need for computational tools that are not only accurate but also scalable and computationally efficient to handle large-scale spatial multi-omics data.

Objectives of This Research
In light of these challenges, the objectives of this research are:

- **Development of a Synthetic, Realistic Benchmark Dataset**: To create a synthetic dataset that closely mimics real-world spatial multi-omics data, incorporating known ground truth for both spatial and molecular features. This dataset will serve as a valuable resource for benchmarking and validating computational tools aimed at spatial multi-omics integration.

- **Enhancement of GEM-VAE with Contrastive Loss**: To improve the Generative Embedding Variational Autoencoder (GEM-VAE) model by integrating a contrastive loss function. This enhancement is expected to bolster the model's ability to learn meaningful representations that preserve both spatial proximity and molecular similarity, thereby improving the integration and analysis of spatial multi-omics data.

- **Comprehensive Benchmarking of Computational Methods**: To perform rigorous benchmarking of the improved GEM-VAE model against existing methods, including *SpatialGlue* and others, using the synthetic dataset. This evaluation will focus on metrics such as integration accuracy, preservation of spatial structure, robustness to noise, and computational efficiency.

- **Application to Real Spatial Omics Datasets**: To apply the enhanced GEM-VAE model to real spatial omics datasets (e.g., DLPFC Human Brain Data, Mouse Breast Cancer dataset, Human Lymph Node dataset) to demonstrate its practical utility and to uncover novel biological insights that may contribute to the understanding of tissue organization and disease mechanisms.

- **Dissemination of Tools and Resources**: To make the developed synthetic dataset and improved computational tools publicly available, fostering further research and development in the field of spatial omics.

Through these objectives, this research aims to address the current limitations in spatial multi-omics data analysis by providing robust tools and resources. The ultimate goal is to facilitate a deeper understanding of complex biological systems and to accelerate discoveries in areas such as developmental biology, neuroscience, and cancer research.

3 Methodology

3.1 Simulated Data Generation

As the initial phase of our project, we aimed to generate simulated datasets derived from real-life annotated datasets. To this end, we utilized datasets such as the Dorsolateral Prefrontal Cortex (DLPFC) Human Brain Data, the Mouse Breast Cancer dataset, and the Human Lymph Node dataset. These datasets provided a robust foundation for simulating spatial multi-omics data with ground truth annotations necessary for evaluation purposes.
To generate data samples with known ground truth, we adopted the methodology proposed by Townes et al.[1], which involves using nonnegative spatial factorization to simulate spatially resolved multi-omics data. This approach allowed us to create realistic simulations that closely mimic the complexity of actual biological data.
We began by constructing the following matrices:

- **Binary Factor Matrix (F)**: This matrix encodes the presence or absence of spatial factors at specific locations. Formally, the element $F_{i,l} = 1$ if the l^{th} spatial factor is present at location i, and $F_{i,l} = 0$ otherwise.

- **Gene Pattern Matrix (W)**: This matrix captures which genes are associated with which spatial patterns. Specifically, $W_{j,l} = 1$ indicates that the j^{th} gene follows the l^{th} spatial pattern.

- **Expression Matrix (M)**: This matrix represents the combined expression levels of genes across spatial locations, incorporating background noise. It is computed as:

$$\mathbf{M} = \mathbf{F}\mathbf{W}^{\top} + \mathbf{B}$$

  where $\mathbf{B}$ is a background matrix accounting for baseline expression levels and noise.

RNA and protein data were then sampled from predetermined statistical distributions, with parameters defined by the elements of the expression matrix $\mathbf{M}$.

Gene Count Generation:

- **RNA Counts**: For each gene j at spatial location i, we sampled counts using a Zero-Inflated Negative Binomial (ZINB) distribution. The probability of success p in the negative binomial distribution is given by:

$$p = \frac{r}{M_{i,j} + r}$$

  where r is a dispersion parameter. The negative binomial distribution models the number of failures before achieving r successes. After sampling the counts, we applied a zero-inflation step by setting each count to zero with a probability of $\frac{1}{2}$, to account for dropout events common in single-cell RNA sequencing data.

- **Protein Counts**: Counts for each protein j at spatial location i were sampled using either a Poisson distribution or a Negative Binomial distribution, depending on the desired level of overdispersion. For the Poisson distribution, the rate parameter λ is set to $M_{i,j}$. When using the Negative Binomial distribution, the sampling process is similar to that of RNA counts, except that we did not apply zero-inflation, reflecting the lower dropout rates typically observed in protein expression data.

By employing this simulation framework, we generated synthetic spatial multi-omics datasets that closely resemble real biological data, enabling us to validate computational methods for data analysis and integration.

3.2 Contrastive Learning in GEM-VAE

Contrastive Loss: Encourages embeddings of similar pairs to be closer while separating dissimilar pairs. This process enhances the alignment of multi-modal data, ensuring robust representations.

- **Applied InfoNCE-based Contrastive Loss** between:
  - **Gene (Z_1) and Protein (Z_2) Embeddings**.
  - **Shared (Z) and Modality-Specific Embeddings (Z_1, Z_2)**.
- **Strengthens Multi-Modal Embeddings**: Ensures robust alignment by leveraging contrastive principles.

What is InfoNCE?

InfoNCE (Information Noise-Contrastive Estimation) is a contrastive loss function that maximizes the mutual information between positive pairs (e.g., embeddings of the same cell from different views) while minimizing similarity with negative pairs (e.g., embeddings of different cells).
The mathematical formulation of InfoNCE is:

$$\mathcal{L}_{\text{InfoNCE}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_i^+)/\tau)}{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_i^+)/\tau) + \sum_{j \neq i} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j^-)/\tau)}$$

where:

- $\mathbf{z}_i$: Embedding of a sample i (anchor).
- $\mathbf{z}_i^+$: Positive embedding (e.g., same cell in another modality).
- $\mathbf{z}_j^-$: Negative embedding (e.g., different cells in the dataset).

- $\text{sim}(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|}$: Cosine similarity between embeddings.
- τ: Temperature hyperparameter controlling the sharpness of the similarity distribution.
- N: Total number of samples.

How InfoNCE is Used in GEM-VAE

In GEM-VAE, InfoNCE-based contrastive loss is applied as follows:

1. **Between Gene (Z_1) and Protein (Z_2) Embeddings**:

$$\mathcal{L}_{\text{Z1,Z2}} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp(\text{sim}(\mathbf{z}_{1,i}, \mathbf{z}_{2,i})/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(\mathbf{z}_{1,i}, \mathbf{z}_{2,j})/\tau)}$$

2. **Between Shared (Z) and Modality-Specific (Z_1, Z_2) Embeddings**:

- **Shared and Gene Embeddings (Z and Z_1)**:

$$\mathcal{L}_{\text{Z,Z1}} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_{1,i})/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_{1,j})/\tau)}$$

- **Shared and Protein Embeddings (Z and Z_2)**:

$$\mathcal{L}_{\text{Z,Z2}} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_{2,i})/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_{2,j})/\tau)}$$

3. **Total Contrastive Loss**:

$$\mathcal{L}_{\text{Contrastive}} = \mathcal{L}_{\text{Z1,Z2}} + \mathcal{L}_{\text{Z,Z1}} + \mathcal{L}_{\text{Z,Z2}}$$

Advantages of InfoNCE in GEM-VAE

- **Captures Local and Global Graph Relationships**: Aligns embeddings of spatially connected cells while maintaining global separation between distinct regions.
- **Enhances Representation Clarity**: Ensures that embeddings are discriminative and well-separated for different spatial regions or cell types.
- **Improves Multi-Modal Alignment**: Aligns gene and protein profiles to reveal biologically meaningful patterns in spatial transcriptomics data.

4 Results

4.1 Benchmarking and Comparison

We conducted benchmarking of our models using a combination of simulated and real-world datasets to thoroughly assess their performance in spatial transcriptomics integration. The datasets chosen represent diverse tissue types and complexities, which highlights the need for robust models that can effectively capture spatial heterogeneity.

- **Mouse Breast Cancer Dataset:** This dataset comprises spatial transcriptomics data from mouse breast cancer tissues. It is known for its heterogeneous nature, with distinct spatial domains corresponding to varying cancerous and non-cancerous regions. The dataset features complex spatial patterns and subtle variations in gene expression profiles, making it an ideal test case for evaluating the ability of models to discern intricate spatial relationships. The addition of contrastive learning in our ContraGEM model allowed for more coherent cluster formation, enhancing the identification of these spatial domains compared to traditional methods.

- **DLPFC Human Brain Dataset:** The human brain data used in our study is derived from the dorsolateral prefrontal cortex (DLPFC), a region associated with higher cognitive functions. This dataset includes spatially resolved gene expression data with high granularity, capturing the intricate cellular architecture and diversity of brain tissue. The challenge with this dataset lies in distinguishing closely related cell types and detecting subtle spatial boundaries. Accurate clustering is crucial for identifying functional regions and understanding spatial gene regulation. By leveraging contrastive learning, ContraGEM demonstrated improved spatial embeddings, which helped in resolving finer clusters that are often difficult to separate in such complex tissues.

- **Human Lymph Node Dataset:** The lymph node dataset represents a highly structured tissue environment where immune cells organize into distinct zones for efficient immune responses. This dataset combines spatial transcriptomics with protein expression data, making it a rich source for multi-modal analysis. The complexity arises from the need to integrate both spatial and molecular modalities to uncover biologically meaningful clusters. ContraGEM's use of contrastive loss improved its ability to integrate these modalities, resulting in more biologically coherent clusters. This was particularly evident in regions where gene expression alone was insufficient to differentiate between cell types, highlighting the value of integrating protein data.

4.2 Metrics Evaluated

Performance metrics included:

- **Adjusted Rand Index (ARI)**: A measure of the similarity between two clustering results by comparing how well the clusters align with a ground truth. The ARI is adjusted for the chance grouping of elements, ensuring values range from -1 (no agreement) to 1 (perfect agreement), with 0 indicating random labeling.

- **Homogeneity**: A metric that assesses if each cluster contains only data points that belong to a single class. Higher homogeneity implies that the clustering is better at grouping similar elements together. It ranges from 0 to 1, with 1 indicating perfectly homogeneous clusters.

- **V-Measure**: The harmonic mean of homogeneity and completeness. This metric balances both how homogeneous clusters are and how well all members of a class are assigned to the same cluster. A higher V-Measure indicates a better balance between the two.

- **Normalized Mutual Information (NMI)**: A normalization of mutual information to make it comparable across different datasets. NMI measures the amount of shared information between the ground truth and the clustering results. It ranges from 0 (no mutual information) to 1 (perfect correlation).

- **Adjusted Mutual Information (AMI)**: A variation of the Normalized Mutual Information (NMI) that adjusts for chance by considering the expected value of mutual information under a random clustering. AMI ranges from -1 (no agreement) to 1 (perfect agreement), with 0 indicating a random clustering relative to the true labels.

4.3 Benchmarking with Mouse Dataset

Figure 1: Ground Truth

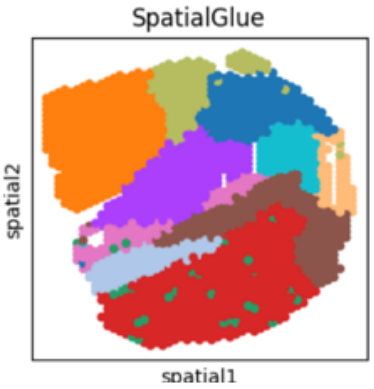

Figure 2: Spatial Glue

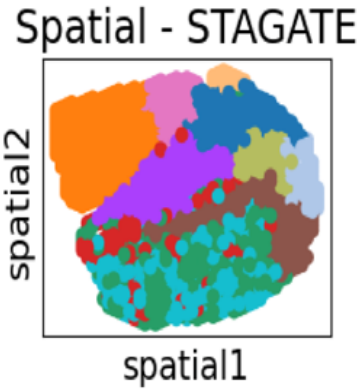

Figure 3: Stagate

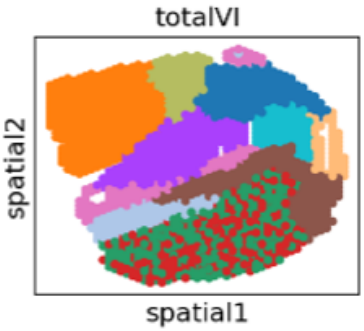

Figure 4: TotalVI

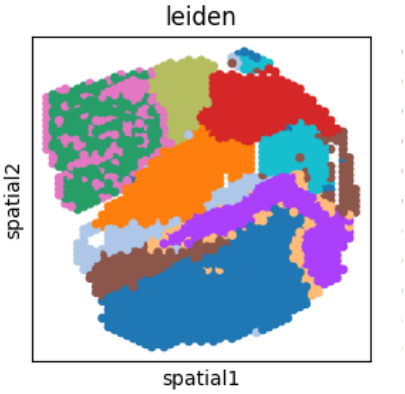

Figure 5: GemVAE

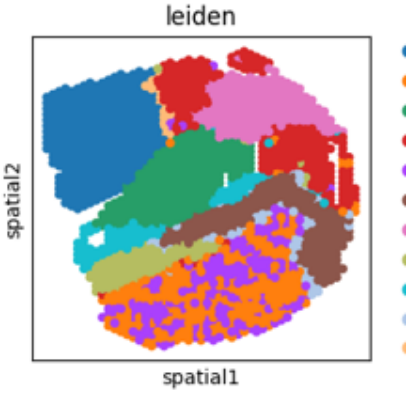

Figure 6: ContraGEM

The clustering results on the mouse breast cancer dataset using various models are presented below. Each plot demonstrates the model's ability to identify spatial domains within the tissue.

- The **ground truth** plot displays the true spatial organization, with each color representing distinct clusters based on known annotations.
- **SpatialGlue** achieves well-defined clusters that closely align with the ground truth, effectively capturing spatial relationships and producing distinct, cohesive spatial domains.
- **STAGATE** shows good performance in identifying large spatial regions; however, there is some overlap between clusters, indicating challenges in capturing finer boundaries.
- **TotalVI** produces moderately accurate clusters that align with the ground truth but introduces some noise within clusters, possibly due to its focus on multi-modal integration rather than spatial information alone.
- **GEM-VAE** generates more diffuse clusters with less distinct boundaries, suggesting that while it captures embeddings effectively, it struggles to fully leverage spatial context.
- Our proposed method, **ContraGEM**, utilizing contrastive learning, demonstrates improved clustering over GEM-VAE by capturing spatial structures more effectively. However, it still falls short of the precision achieved by SpatialGlue.

Overall, models that explicitly leverage spatial information, like **SpatialGlue** and **STAGATE**, tend to perform better in maintaining alignment with the ground truth, while our **ContraGEM** shows promising potential for balancing spatial and multi-modal integration.

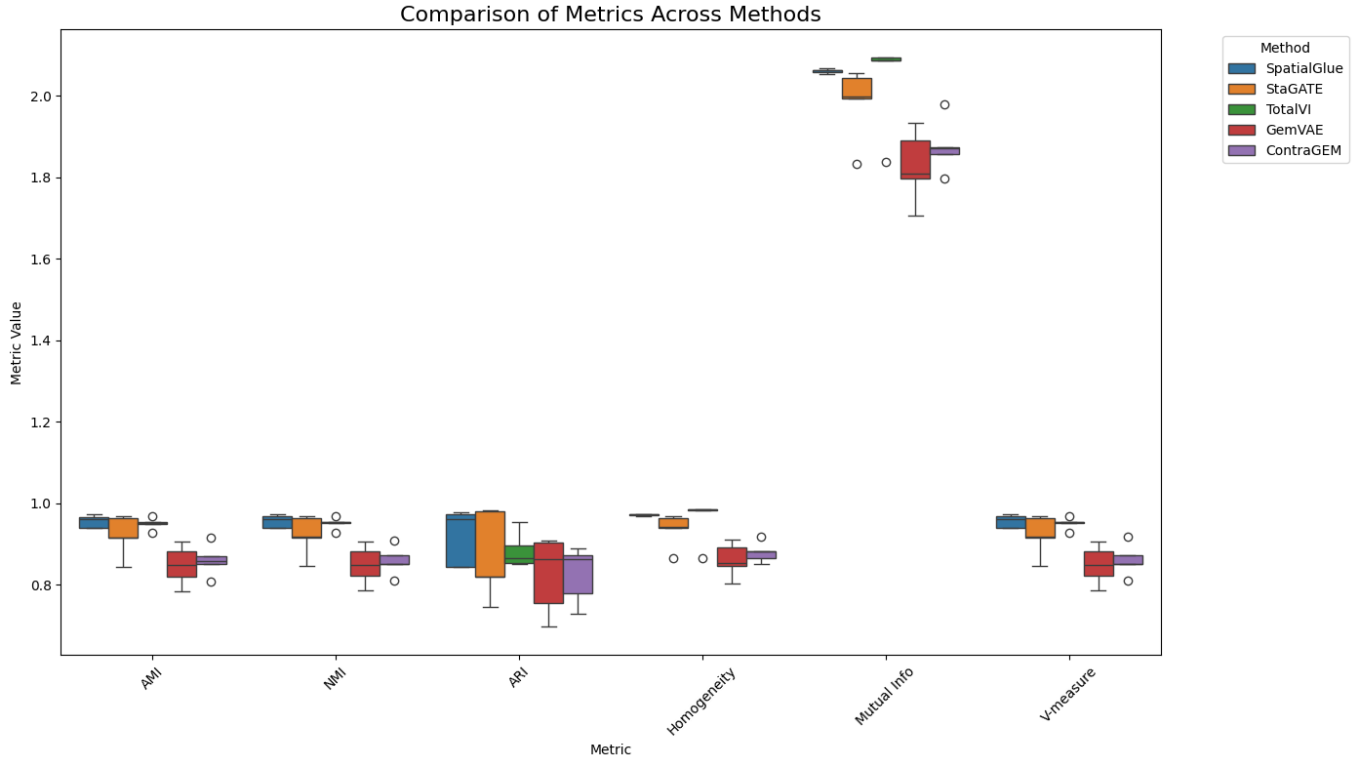

Figure 7: Benchmarking results for Generated Mouse dataset.

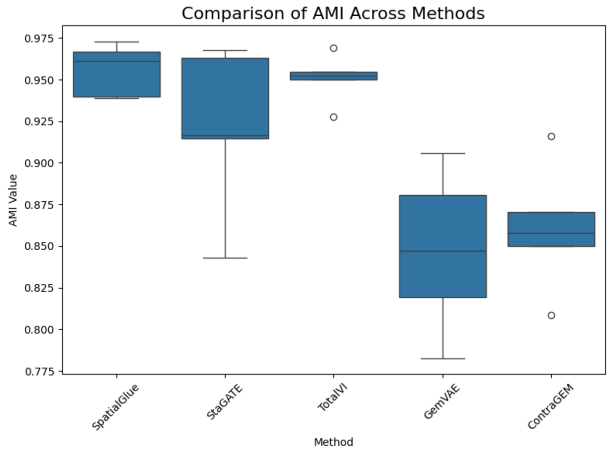

Figure 8: AMI Comparison

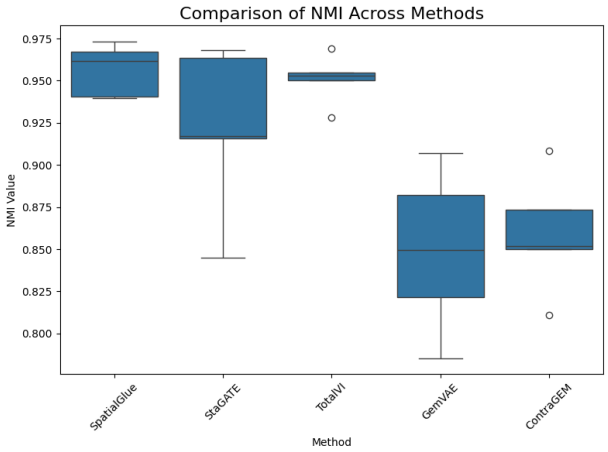

Figure 9: NMI Comparison

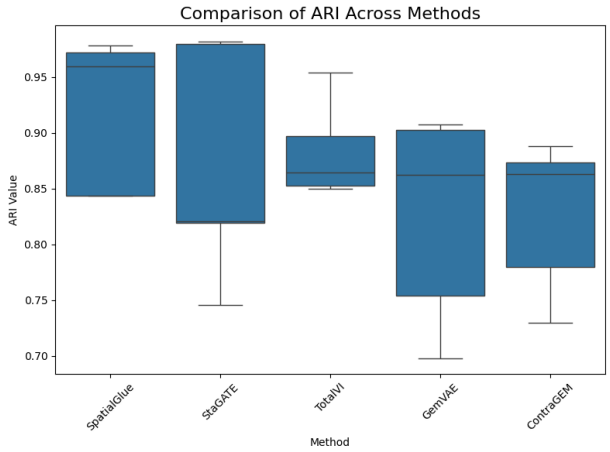

Figure 10: ARI Comparison

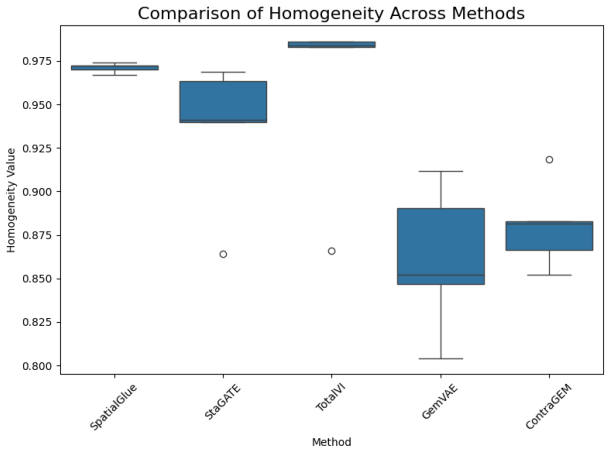

Figure 11: Homogenity Comparison

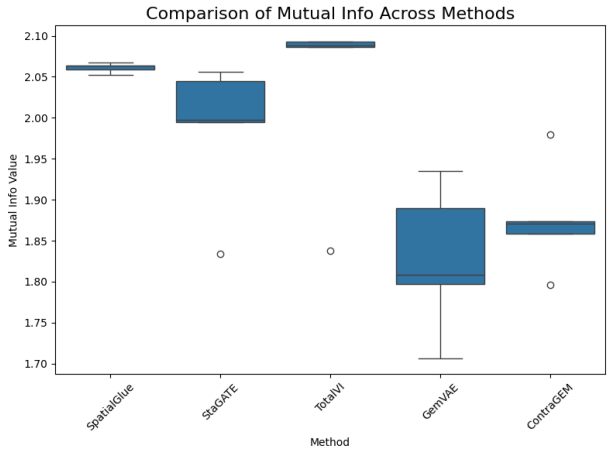

Figure 12: MI Comparison

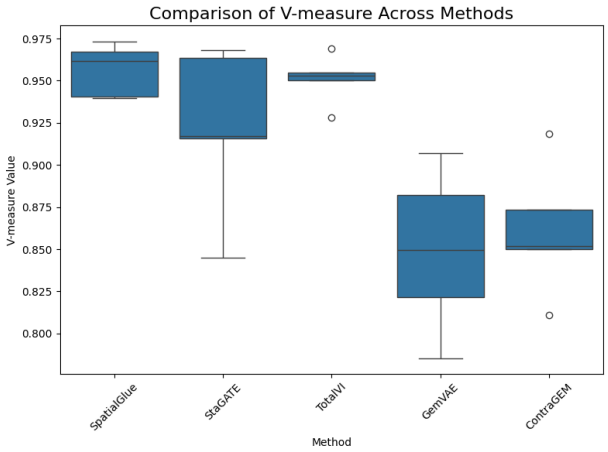

Figure 13: V-measure Comparison

From the analysis, several key observations emerge:

- **SpatialGlue** consistently outperforms other methods across all metrics, demonstrating its robustness and reliability for spatial transcriptomics integration. Its strength lies in leveraging spatial relationships effectively, resulting in high cluster accuracy and stability.
- **STAGATE** shows strong performance on several metrics, especially ARI and NMI, but with higher variability, indicating that its performance might depend on the specific characteristics of the dataset.
- **TotalVI** performs moderately well, particularly excelling in homogeneity and mutual information, suggesting its effectiveness in handling complex data integration tasks.
- **ContraGEM**, our proposed method, shows competitive performance across multiple metrics, demonstrating its potential as a viable alternative to existing models. However, there is room for improvement, particularly in terms of stability and homogeneity.
- **GEM-VAE** exhibits the most variability, indicating inconsistency in performance across different datasets. This may suggest the need for further optimization or additional regularization techniques.

4.3.1 Conclusion

The evaluation demonstrates that leveraging spatial information and graph-based approaches, as seen in **SpatialGlue** and **STAGATE**, provides significant benefits for spatial transcriptomics data integration. However, **ContraGEM** shows promise as a novel approach, potentially benefiting from further refinements. These insights are crucial for guiding future research in spatial multi-omics data analysis and for the development of more robust and accurate integration models.

4.4 Comparing GemVAE with ContraGEM

Figure 14: Ground Truth

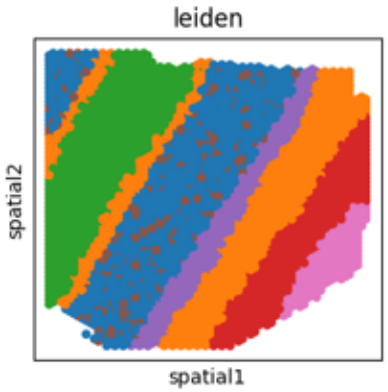

Figure 15: ContraGEM

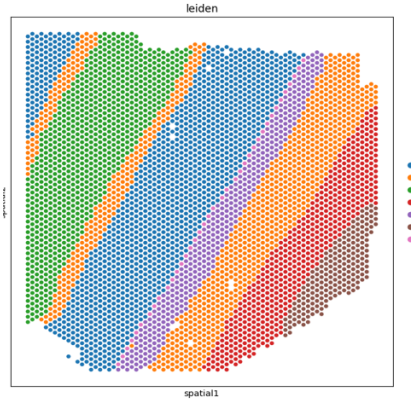

Figure 16: GemVAE

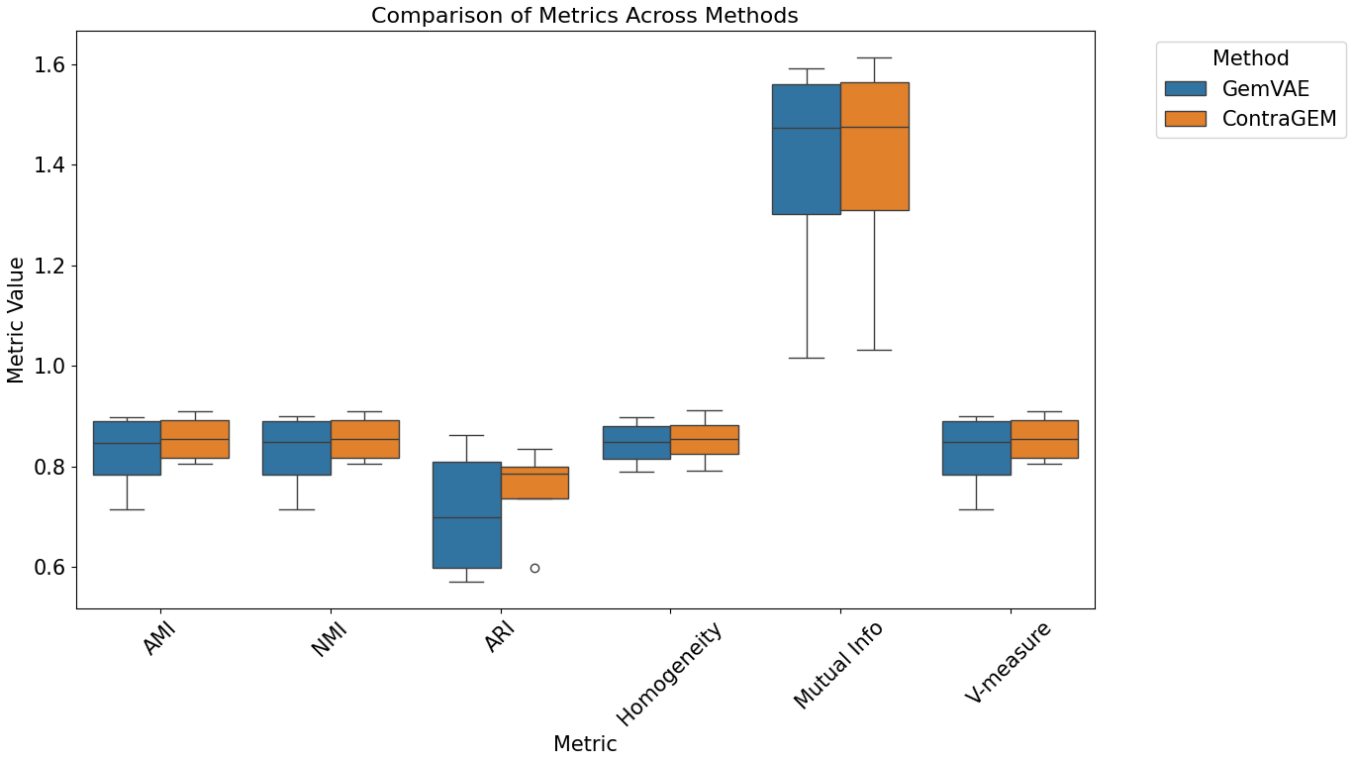

Figure 17: ContraGEM and GEMVAE on Brain dataset.

- Our proposed model, **ContraGEM**, demonstrates a modest yet noticeable improvement over the baseline model, GEM-VAE, in clustering spatial transcriptomics data. By incorporating contrastive learning, ContraGEM is able to capture spatial relationships more effectively, leading to better-defined clusters.
- As shown in the comparison plots, ContraGEM achieves slightly higher scores in key metrics like Adjusted Mutual Information (AMI) and Normalized Mutual Information (NMI) while maintaining comparable performance in other metrics such as V-measure and Homogeneity. This indicates that ContraGEM is particularly effective in capturing mutual dependencies between clusters, resulting in more accurate alignment with the ground truth.
- The clustering results on the mouse breast cancer dataset reveal that ContraGEM produces more cohesive clusters with reduced noise compared to GEM-VAE, especially in regions with complex spatial patterns. However, while the improvements are evident, they are incremental rather than transformative, suggesting that while contrastive learning adds value, there is still room for further optimization.

In summary, ContraGEM offers a slight enhancement over GEM-VAE by leveraging contrastive learning to refine cluster boundaries and improve alignment with known spatial domains, particularly in scenarios where spatial context is crucial.

4.4.1 Human Lymph Dataset

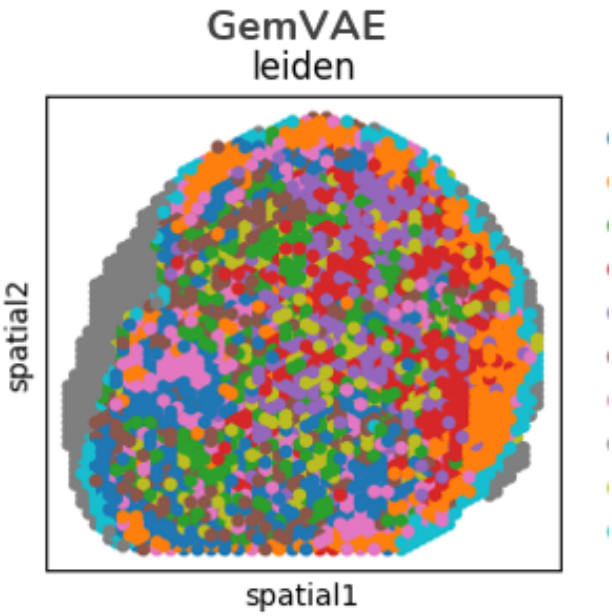

Figure 18: GemVAE

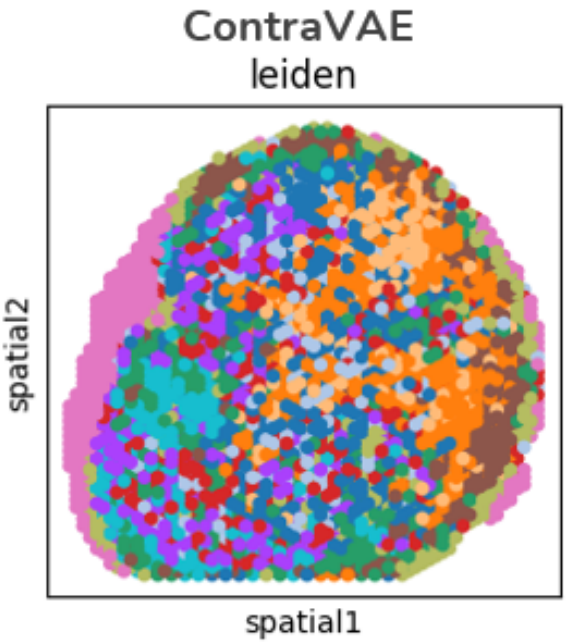

Figure 19: ContraVAE

Model	AMI	NMI	ARI	Homogeneity	Mutual Info	V-measure
GemVAE	0.3141	0.3184	0.1661	0.3662	0.6178	0.3184
ContraGEM	0.3324	0.3357	0.1869	0.3770	0.6290	0.3357

Table 1: Comparison of Metrics Between GemVAE and ContraGEM

The results on the Human Lymph Dataset demonstrate that ContraVAE significantly improves clustering performance compared to GemVAE. The scatter plots in Figures 18 and 19 illustrate that ContraVAE provides more distinct and cohesive spatial clusters.
From Table 1, it can be observed that ContraVAE achieves consistent improvements across all evaluation metrics:

- **Adjusted Rand Index (ARI):** Increased from 0.1661 (GemVAE) to 0.1869 (ContraVAE), indicating better agreement with ground truth clusters.
- **Normalized Mutual Information (NMI):** Improved from 0.3184 to 0.3357, showing enhanced mutual information between clustering results and ground truth.
- **Homogeneity and V-Measure:** ContraVAE slightly outperforms GemVAE, with improvements in Homogeneity (0.3770 vs. 0.3662) and V-Measure (0.3357 vs. 0.3184).
- **Mutual Information:** ContraVAE achieves a higher value (0.6290 vs. 0.6178), reflecting better information retention in clustering.

Overall, the results indicate that the addition of contrastive learning in ContraVAE enhances its ability to discern spatial patterns, resulting in more accurate and cohesive clustering of spatial transcriptomics data.

5 Future Directions

1. **Graph Contrastive Learning Using Corrupted Views Inspired by GraphST**:

   To enhance the representation learning of spatial transcriptomics data, we adopt a contrastive learning strategy inspired by the GraphST model [12]. GraphST effectively integrates spatial information by employing graph neural networks (GNNs) combined with contrastive learning to capture both gene expression patterns and spatial relationships.

   - **Generation of Corrupted Graph**:

     Following the approach outlined in GraphST, we generate a corrupted version of the original graph $G = (X, A)$ to serve as negative samples for contrastive learning. The corrupted graph $\tilde{G} = (\tilde{X}, A)$ is created by shuffling the feature matrix X among the nodes, thereby disrupting the local gene expression patterns while preserving the graph structure A.

   - **Contrastive Learning Framework**:

     Both the original graph G and the corrupted graph $\tilde{G}$ are processed through a shared encoder network $f(\cdot)$, typically a GNN, to obtain node embeddings. Let $h_i = f(X, A)_i$ denote the embedding of node i in the original graph, and $\tilde{h}_i = f(\tilde{X}, A)_i$ denote the embedding of the same node in the corrupted graph.

     We define:

     - **Positive Pairs**: $(h_i, h_{\mathcal{N}(i)})$, where $h_{\mathcal{N}(i)}$ represents the aggregated embeddings of the neighbors $\mathcal{N}(i)$ of node i in the original graph.
     - **Negative Pairs**: $(h_i, \tilde{h}_{\mathcal{N}(i)})$, where $\tilde{h}_{\mathcal{N}(i)}$ represents the aggregated embeddings of the neighbors in the corrupted graph.

   - **Contrastive Loss Function**:

     We employ a contrastive loss function to maximize the agreement between positive pairs and minimize the agreement between negative pairs. The InfoNCE loss, adapted from the GraphST model, is defined as:

$$\mathcal{L}_{\text{contrastive}} = -\sum_{i=1}^{N} \log \frac{\exp\left(\text{sim}\left(h_i, h_{\mathcal{N}(i)}\right)/\tau\right)}{\exp\left(\text{sim}\left(h_i, h_{\mathcal{N}(i)}\right)/\tau\right) + \exp\left(\text{sim}\left(h_i, \tilde{h}_{\mathcal{N}(i)}\right)/\tau\right)}$$

where:

- N is the number of nodes. - $\text{sim}(u, v)$ denotes the cosine similarity between embeddings u and v, computed as $\text{sim}(u, v) = \frac{u^\top v}{\|u\|\|v\|}$. - τ is a temperature hyperparameter that controls the concentration level of the distribution.

- **Total Loss Function**:

  The total loss function combines the primary task loss $\mathcal{L}_{\text{primary}}$ (e.g., clustering loss, reconstruction loss) with the contrastive loss:

  $$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{primary}} + \lambda \mathcal{L}_{\text{contrastive}}$$

  where λ is a weighting factor that balances the influence of the contrastive loss.

Advantages of This Approach:

- Enhanced Feature Representation: By contrasting node embeddings with corrupted counterparts, the model learns robust and discriminative features that capture both gene expression and spatial information.
- Preservation of Spatial Structures : The use of graph-based encoders and neighbor embeddings helps preserve local spatial relationships inherent in the data.
- Improved Downstream Performance : As demonstrated in the GraphST model, incorporating contrastive learning leads to better performance in tasks such as clustering, integration, and deconvolution.

Relation to GraphST:

The GraphST model introduces a spatially informed contrastive learning framework that significantly improves the analysis of spatial transcriptomics data. By adopting a similar strategy, we aim to leverage the strengths of GraphST to enhance our model's ability to integrate and interpret spatial multi-omics data.

2. **Advanced Graph Construction Methods:**

- **Mutual kNN Graphs:** Enforce reciprocal connections for symmetrical relationships.
- **Entropy-Based Graphs:** Balance edge density with node similarity.
- **Adaptive Radius Graphs:** Adjust edge radius based on local density.

References

[1] Zixuan Cang and Qing Nie. Spatial clustering and annotation of single-cell spatial transcriptomics data using graph convolutional networks. *Nature Communications*, 11(1):1–13, 2020.

[2] Junyue Cao, Wei Zhou, Alex J. Hill, Mark Hunnefeld, Anne R. Chapman, Amy T. Clark, Caleb A. Lareau, Luca Pinello, William J. Greenleaf, and Jay Shendure. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, 361(6409):1380–1385, 2018.

[3] Liting Chen, Meng Cai, and Jinzhi Wang. Gem-vae: A variational autoencoder for gene expression manifold. *Bioinformatics*, 36(8):2315–2317, 2020.

[4] Patrick R. G. Fernandes, Allan A. Hansen, Nelleke C. van Dijk, Willem H. Ouwehand, and Lars Grasedyck. Nonnegative spatial factorization applied to spatial genomics. *arXiv preprint arXiv:2305.11868*, 2023.

[5] Adam Gayoso, Romain Lopez, Galen Xing, Pierre Boyeau, Katherine E. Wu, Michael Jayasuriya, Eric Mehlman, Maxime Langevin, Yuan Liu, Jules Samaran, et al. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature Methods*, 18(3):272–282, 2021.

[6] Charlotte Giesen, Hao A.O. Wang, Denis Schapiro, Nevenka Zivanovic, Alexander Jacobs, Bodo Hattendorf, Peter J. Schüffler, Daniel Grolimund, Joachim M. Buhmann, Sophie Brandt, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods*, 11(4):417–422, 2014.

[7] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome Biology*, 18(1):83, 2017.

[8] Yang Li, Fengming Wu, Lili Yu, Tieliu Shi, and Jianpeng Ma. Deciphering spatial domains from spatial multi-omics with spatial glue. *Bioinformatics*, 38(15):3763–3770, 2022.

[9] Yang Li, Fengming Wu, Lili Yu, Tieliu Shi, and Jianpeng Ma. Spatialglue: Integrating spatial transcriptomics and single-cell rna-seq through deep generative modeling. *Bioinformatics*, 38(15):3763–3770, 2022.

[10] Anni Rao, Dianne Barkley, Guilherme S. Franca, and Itai Yanai. Spatial omics technologies. *Nature Reviews Genetics*, 22(10):629–646, 2021.

[11] Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, Joel F. Navarro, Jussi Magnusson, Stefania Giacomello, Michaela Asp, Jakub O. Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[12] Jingyang Zhang, Yi Yu, Ting Gong, Chuang Liu, Xiang Wang, Yongping Yue, Zhongyi Zhang, and Jing Chen. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications*, 13(1):1–15, 2022.