

---

# MuSe-GNN: Learning Unified Gene Representation From Multimodal Biological Graph Data

---

**Tianyu Liu**  
Yale University  
tianyu.liu@yale.edu

**Yuge Wang**  
Yale University  
yuge.wang@yale.edu

**Rex Ying**  
Yale University  
rex.ying@yale.edu

**Hongyu Zhao\***  
Yale University  
hongyu.zhao@yale.edu  
\*: Corresponding author

## Abstract

Discovering genes with similar functions across diverse biomedical contexts poses a significant challenge in gene representation learning due to data heterogeneity. In this study, we resolve this problem by introducing a novel model called Multimodal Similarity Learning Graph Neural Network, which combines Multimodal Machine Learning and Deep Graph Neural Networks to learn gene representations from single-cell sequencing and spatial transcriptomic data. Leveraging 82 training datasets from 10 tissues, three sequencing techniques, and three species, we create informative graph structures for model training and gene representations generation, while incorporating regularization with weighted similarity learning and contrastive learning to learn cross-data gene-gene relationships. This novel design ensures that we can offer gene representations containing functional similarity across different contexts in a joint space. Comprehensive benchmarking analysis shows our model’s capacity to effectively capture gene function similarity across multiple modalities, outperforming state-of-the-art methods in gene representation learning by up to 97.5%. Moreover, we employ bioinformatics tools in conjunction with gene representations to uncover pathway enrichment, regulation causal networks, and functions of disease-associated or dosage-sensitive genes. Therefore, our model efficiently produces unified gene representations for the analysis of gene functions, tissue functions, diseases, and species evolution.

## 1 Introduction

Progress in biological technology has broadened the variety of biological data, facilitating the examination of intricate biological systems. A prime example of such technology is single-cell sequencing, which allows for the comprehensive characterization of genetic information within individual cells [38, 78]. This technology provides access to the full range of a cell’s transcriptomics, epigenomics and proteomics information, including gene expression (scRNA-seq) [46, 114], chromatin accessibility (scATAC-seq) [20, 17], methylation [60, 68] and anti-bodies [85]. By sequencing cells from the same tissue at various time points, we can gain insight into patterns of cellular activity over time [30]. Moreover, spatial information of cellular activity represents an equally vital, additional dimension. Such data are defined as spatial data. All of these data are known as multi-omics data, and integrating multi-omics data for combined analysis poses a significant challenge.

However, the traditional idea of multi-omics data integration known as using cells as anchors [56, 87, 39, 12, 21, 76] is only partially suitable because of the following two challenges. 1. Different omics

data pose their own challenges. For example, the unit of observation of the spatial transcriptomic data is different from other single-cell data, as a single spatial location contains mixed information (see the left part of Figure 1 a) from various cells [23] and it is not appropriate to generate spatial clusters based on gene expression similarity [16]. Current research [86] also indicates that chromosome accessibility feature is not a powerful predictor for gene expression at the cell level. 2. The vast data volume in atlas-level studies challenges high-performance computing, risking out-of-memory or time-out errors [58, 101]. With nearly 37.2 trillion cells in the human body [77], comprehensive analysis is computationally infeasible. 3. Batch effects may adversely impact analysis results [54] by introducing noise. Consequently, an efficient and powerful method focusing on multi-omics and multi-tissue data (referred to as multimodal biological data) analysis is urgently needed to address these challenges.

Acknowledging the difficulties arising from the cell-oriented viewpoint, prior work shifted focus to the gene perspective. Using gene sets as a summary of expression profiles, based on natural selection during the species evolution, may provide a more robust anchor [8]. Protein-coding genes are also thought to interact with drugs [19], which are more relevant to diseases and drug discovery. Gene2vec [24] is a method inspired by Word2vec [18], which learns gene representations by generating skip-gram pairs from the co-expression network. Recently, the Gene-based data Integration and ANalysis Technique (GIANT) [16] has been developed, based on Node2vec [34] and OhmNet [119], to learn gene representations from both single-cell and spatial datasets. However, as shown in Figure 1 (b), significant functional clustering for genes from different datasets but the same tissue was not observed based on the gene embeddings from these two models, because these methods do not infer the similarity of associated genes from different multimodal data. Additionally, they did not offer metrics to quantitatively evaluate the performance of gene embeddings compared to baseline models.

Here we introduce a novel model called **Multimodal Similarity Learning Graph Neural Network (MuSe-GNN)**<sup>1</sup> for multimodal biological data integration from a gene-centric perspective. The overall workflow of MuSe-GNN is depicted in Figure 1 (a). Figure 1 (b) shows MuSe-GNN's superior ability to learn functional similarity among genes across datasets by suitable model structure and novel loss function, comparing to GIANT and Gene2vec. MuSe-GNN utilizes weight-sharing Graph Neural Networks (GNNs) to encode genes from different modalities into a shared space regularized by the similarity learning strategy and the contrastive learning strategy. At the single-graph level, the design of graph neural networks ensures that MuSe-GNN can learn the neighbor information in each co-expression network, thus preserving the gene function similarity. At the cross-data level, the similarity learning strategy ensures that MuSe-GNN can integrate genes with similar functions into a common group, while the contrastive learning strategy helps distinguish genes with different functions. Furthermore, MuSe-GNN utilizes more robust co-expression networks for training and applies dimensionality reduction [89] to high-dimensional data [73].

To the best of our knowledge, this is the first paper in gene representation learning that combines the Multimodal Machine Learning (MMML) concept [9, 27] with deep GNNs [49, 81] design. Both approaches are prevalent in state-of-the-art (SOTA) machine learning research, inspiring us to apply them to the joint analysis of large-scale multimodal biological datasets<sup>2</sup>. As application examples, we first used the gene embeddings generated by MuSe-GNN to investigate crucial biological processes and causal networks for gene regulation in the human body. We also applied our model to analyze COVID and cancer datasets, aiming to unveil potential disease resistance mechanisms or complications based on specific differentially co-expressed genes. Lastly, we used gene embeddings from MuSe-GNN to improve the prediction accuracy for gene functions. Here genes are co-expressed means genes are connected with the same edge.

Given the lack of explicit metrics to evaluate gene embeddings, we proposed six metrics inspired by the batch effect correction problem [54] in single-cell data analysis. We evaluated our model using real-world biological datasets from one technique, and the benchmarking results demonstrated MuSe-GNN's significant improvement from **20.1%** to **97.5%** in comprehensive assessment. To summarize the advantages of our model, MuSe-GNN addresses the outlined problem about cross-data gene similarity learning and offers four major contributions: 1. Providing an effective representation learning approach for multi-structure biological data. 2. Integrating genes from different omics and tissue data into a joint space while preserving biological information. 3. Identifying co-located

---

<sup>1</sup>Codes of MuSe-GNN: <https://github.com>HelloWorldLY/MuSe-GNN>

<sup>2</sup>Download links: Appendix M.

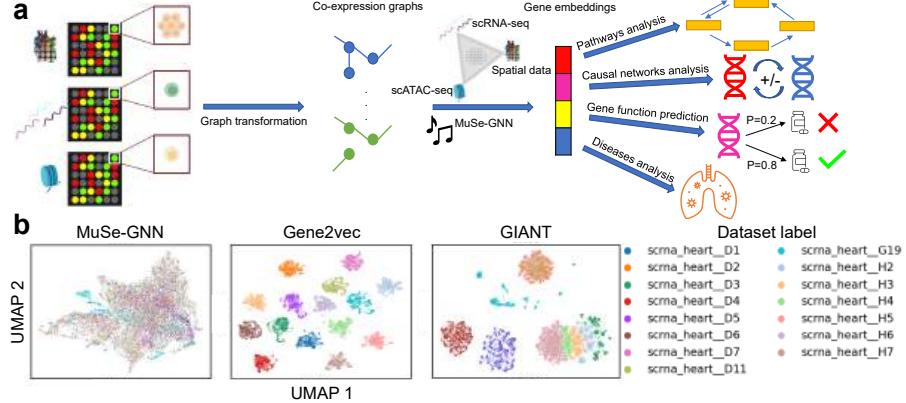


Figure 1: The workflow of MuSe-GNN, the visualization of gene embeddings, and the problems of two existing methods, GIANT and Gene2vec. **(a)** The process of learning gene embeddings by MuSe-GNN. Here we highlight the difference between single-cell data and spatial data, and the major applications of gene embeddings. Each spot from single-cell data represents one cell, while each spot from spatial data represents a mixture of cells. **(b)** UMAPs [63] for the gene embeddings of MuSe-GNN, Gene2vec and GIANT. They are colored by datasets. Based on UMAPs we can conclude that both Gene2vec and GIANT failed to learn the gene similarity based on the datasets from the same tissue, while MuSe-GNN produces meaningful embeddings agnostic to datasets.

genes with similar functions. 4. Inferring specialized causal networks of genes and the relationships between genes and biological pathways or between genes and diseases.

## 2 Related work

**Co-expression Network Analysis.** While direct attempts at joint analysis of gene functions across modalities are limited, there is relevant research in identifying correlation networks based on a single dataset. WGCNA [52] is a representative method that employs hierarchical clustering to identify gene modules with shared functions. However, as an early tool, WGCNA has restricted functionality. Its inference of co-expression networks necessitates more rigorous methods, and it cannot be directly applied to the analysis of multimodal biological data.

**Network Based Biological Representation Learning.** Apart from directly generating gene co-expression networks, learning quantitative gene representations in a lower dimensional space may better describe gene-gene relationships and facilitate downstream analyses. Gene2vec generates gene embeddings based on the co-expression network from a given database. However, it disregards the expression profile information and is based on the old Gene Expression Omnibus (GEO) up until 2017 [24]. GIANT leverages Node2vec [34] and OhmNet [119] to learn gene representations from both single-cell and spatial datasets by constructing graphs and hypergraphs for training. However, this approach still over-compresses multimodal biological datasets by removing expression profiles. Moreover, their co-expression networks created by Pearson correlation have a high false positive rate [82]. Additionally, some of the datasets used by GIANT are of low quality (see Appendix A). There are also methods sharing a similar objective of learning embeddings from other datasets. GSS [71] aims to learn a set of general representations for all genes from bulk-seq datasets [62] using Principal Component Analysis (PCA) and clustering analysis. However, it is for bulk-seq data and cannot be directly applied to single-cell datasets from different tissues. Gemini [103] focuses on integrating different protein function networks, with graph nodes representing proteins rather than genes.

**Graph Transformer.** In the deep learning domain, Transformer [95, 100] is one of the most successful models, leveraging seq2seq structure design, multi-head self-attention design, and position encoding design. Many researchers have sought to incorporate the advantages of Transformer to graph structure learning. TransformConv [81] introduces a multi-head attention mechanism to the graph version of supervised classification tasks and achieved significant improvements. MAGNA [98] considers higher-level hop relationships in graph data to enhance node classification capabilities. Graphomer [110] demonstrates the positive impact of the Transformer structure on various tasks

using data from the Open Graph Benchmark Large-Scale Challenge [44], which is further extended by GraphomerGD [111]. Recently, GPS [75] proposes a general Graph Transformer (GT) model by considering additional encoding types. Transformer architecture also contributes solutions to several biological questions [113]. scBERT [108] generates gene and cell embeddings using pre-training to improve the accuracy of cell type annotations. The substantial impact of these efforts highlights the crucial contribution of the Transformer architecture to graph data learning.

### 3 Methods

In the following sections of introducing MuSe-GNN, we will elaborate on our distinct approaches for graph construction that utilize multimodal biological data, followed by an explanation of our weight-sharing network architecture and the elements of the final loss function.

#### 3.1 Preliminaries

**GNN.** GNNs aim to learn the graph representation of nodes (features) for data with a graph structure. Modern GNNs iteratively update the representation of a node by aggregating the representations of its  $k$ -order neighbors ( $k \geq 1$ ) and combining them with the current representation. As described in [110], considering a graph  $\bar{G} = \{V, E\}$  with nodes  $V = \{v_1, v_2, \dots, v_n\}$ , the AGGREGATE-COMBINE update for node  $v_i$  is defined as:

$$a_i^{(l+1)} = \text{AGGREGATE}^{(l+1)} \left( \left\{ h_j^{(l)} : j \in \mathcal{N}(v_i) \right\} \right); h_i^{(l+1)} = \text{COMBINE}^{(l+1)} \left( h_i^{(l)}, a_i^{(l+1)} \right), \quad (1)$$

where  $\mathcal{N}(v_i)$  represents the neighbors of node  $v_i$  in the given graph, and  $h_i^{(l)}$  and  $h_i^{(l+1)}$  represent the node representation before and after updating, respectively.

**Problem Definition.** We address the gene embeddings generation task by handling multimodal biological datasets, denoted as  $\mathcal{D} = (\{V_i, E_i\})_{i=1}^T$ . Our goal is to construct a model  $\mathcal{M}(\cdot, \theta)$ , designed to yield gene embeddings set  $\mathcal{E} = \{e_1, \dots, e_T\} = \mathcal{M}(\mathcal{D}, \theta)$ . In this context,  $\mathcal{D}$  represents the input,  $\theta$  represents the parameters, and  $\mathcal{E}$  represents the output. In other words, we aim to harmonize gene information from diverse modalities within a unified projection space, thereby generating consolidated gene representations.

#### 3.2 Graph construction

Before constructing gene graphs, our first contribution involves the selection of highly variable genes (HVGs) for each dataset. These HVGs constitute a group of genes with high variance that can represent the biological functions of given expression profiles. Moreover, considering co-expression networks is important for gene representation learning because it allows us to characterize gene-gene relationships. As sequencing depth, or the total counts of each cell, often serves as a confounding factor in the co-expression networks inference [11], we employ two unique methodologies, scTransform [35] and CS-CORE [88], to process scRNA-seq and scATAC-seq data, thus creating gene expression profiles and co-expression networks unaffected by sequencing depth. For spatial transcriptomic data, our focus is on genes displaying spatial expression patterns. We use SPARK-X [116] to identify such genes and then apply scTransform and CS-CORE. For detailed algorithmic information regarding these methods, please see Appendix B. Additionally, we demonstrate the immunity of CS-CORE to batch effects when estimating gene-gene correlations in Appendix C. In all our generated graphs (equivalent to co-expression datasets), nodes represent **genes** and edges represent **co-expression** relation of genes.

#### 3.3 Cross-Graph Transformer

To capitalize on the strengths of the Transformer model during our training process, we integrate a graph neural network featuring a multi-head self-attention design [81], called TransformerConv, to incorporate co-expression information and generate gene embeddings. Details of TransformerConv can be found in Appendix B.4. Incorporating multimodal information can estimate more accurate gene embeddings, supported by Appendix D. The cross-graph transformer can efficiently learn gene embeddings containing gene functions across different graphs, advocated by the comparison of different network structure choices in Appendix E.1.

**Weight sharing.** Given the variability among multimodal biological datasets, we employ a weight-sharing mechanism to ensure that our model learns shared information across different graphs, representing a novel approach for learning cross-graph relationships. We also highlight the importance of weight-sharing design in Appendix E.1.

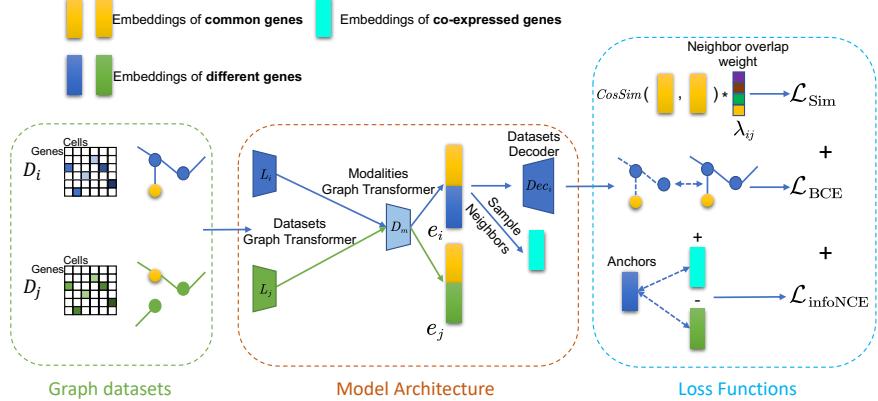


Figure 2: The overall model architecture and the design of loss functions for MuSe-GNN. The color of nodes in the green block represents common/different genes across two datasets. The brown block represents the network architecture of MuSe-GNN, and the blue block represents different loss function components of MuSe-GNN. The color gradients of the left two matrices represent different gene expression levels.

**Datasets & Modalities Graph Transformer.** Drawing inspiration from the hard weight-sharing procedure [15, 115], we not only employ dataset-specific Graph Transformer (GT) layers  $L_1, L_2, \dots, L_n$  for each graph ( $G_1, G_2, \dots, G_n$ ) from the same modality  $m$ , but also connect all these dataset-specific layers to a set of shared GT layers, denoted as  $D_m$ . This design showcases our novel approach to incorporating weight-sharing into the GT framework. The forward process of MuSe-GNN, given dataset  $i$  with network parameter  $\theta_*$ , is defined as follows:

$$X'_i = D_m(L_i(G_i; \theta_{L_i}); \theta_{D_m}). \quad (2)$$

**Datasets Decoder.** Here we propose a dataset-specific decoder structure based on Multi-layer Perceptrons (MLP). This decoder model is crucial in reconstructing the co-expression relationships among different genes, showcasing our inventive use of MLP for this purpose. Given a graph  $G_i$  and its corresponding gene embedding  $e_i$ , the decoding process of MuSe-GNN, with network parameter  $\theta_{dec,i}$ , is defined as follows:

$$E_{rec} = \text{DECODERMLP}(e_i e_i^T; \theta_{dec,i}), \quad (3)$$

where  $E_{rec}$  represents the reconstructed co-expression network.

### 3.4 Graph Reconstruction Loss ( $\mathcal{L}_{\text{BCE}}$ )

Within a single graph, we implement a loss function inspired by the Graph Auto-encoder (GAE) [48]. This function is designed to preserve two key aspects: 1. the similarity among genes sharing common functions, and 2. the distinctions among genes with differing functions. This innovative use of a GAE-inspired loss function constitutes a significant contribution to the methodological design. For a graph  $G_i = \{V_i, E_i\}$ , the loss function for edge reconstruction is defined as:

$$\begin{aligned} e_i &= \text{ENCODERGNN}(\{V_i, E_i\}; \theta_{enc}) = D_m(L_i(G_i; \theta_{L_i}); \theta_{D_m}), \\ E_{rec} &= \text{DECODERMLP}(e_i e_i^T; \theta_{dec,i}); E'_{rec} = \text{FLATTEN}(E_{rec}); E'_i = \text{FLATTEN}(E_i), \\ \mathcal{L}_{\text{BCE}} &= -\frac{1}{|E'_{rec}|} \sum_{t=1}^{|E'_{rec}|} [E'_i[t] \cdot \log E'_{rec}[t] + (1 - E'_i[t]) \cdot \log (1 - E'_{rec}[t])], \end{aligned} \quad (4)$$

where ENCODERGNN and DECODERMLP represent the encoder and decoder parts of our model, respectively.  $\mathcal{L}_{\text{BCE}}$  denotes the computation of binary cross entropy loss (BCELoss) for the given input data.  $E_{rec}$  is the reconstructed adjacency matrix and  $|E'_{rec}|$  is the length of its flatten version. Further justification of the model design can be found in Appendix F.

### 3.5 Weighted-similarity Learning ( $\mathcal{L}_{\text{Sim}}$ )

To integrate shared biological information across graphs from disparate datasets, we fuse the reconstruction loss of the input graph structure with a cosine similarity learning loss. In this process, we treat common HVGs between each pair of datasets as anchors. Our aim is to maximize the cosine similarity ( $\text{COSSIM}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|_2 \cdot \|\mathbf{b}\|_2}$ ) for each common HVG across two arbitrary datasets (represented by the yellow blocks in Figure 2), utilizing their gene embeddings. However, in practice, different common HVGs may have different levels of functional similarity in two datasets, which is hard to be directly quantified. Consequently, we employ the shared community score as an indirect measurement, which is incorporated as a weight for the cosine similarity of different common HVG pairs within the final loss function. Considering two graphs  $G_i = \{V_i, E_i\}$  and  $G_j = \{V_j, E_j\}$ , and one of their shared genes  $g$ , we identify the co-expressed genes of the given gene  $g$  in both datasets, denoted as  $N_{ig}$  and  $N_{jg}$ . Thus, the weight  $\lambda_{ijg}$  for gene  $g$  can be expressed as follows:

$$\lambda_{ijg} = \frac{|N_{ig} \cap N_{jg}|}{|N_{ig} \cup N_{jg}|}. \quad (5)$$

We can iterate over all shared genes from 1 to  $n$  between these two graphs, ultimately yielding a vector as  $\lambda_{ij} = [\lambda_{ij1}, \dots, \lambda_{ijn}]$ . This vector encapsulates the community similarity between the two graphs across all common HVGs. We can then modify the cosine similarity of various gene pairs by first multiplying this vector with cosine similarities and then summing the resultant values across all genes. The negation of the outcome is our final weighted similarity loss, denoted as  $\mathcal{L}_{\text{Sim}}$ . The ablation test for weighed similarity loss can be found in Appendix E.1. More detailed explanations and evidence supporting this design in multimodal conditions can be found in Appendix G.

### 3.6 Self-supervised Graph Contrastive Learning ( $\mathcal{L}_{\text{InfoNCE}}$ )

Specifically, when integrating multimodal biological data, we employ the contrastive learning strategy [72] to ensure that functionally similar genes are clustered together as closely as possible, while functionally different genes are separated apart from each other. We utilize Information Noise Contrastive Estimation (InfoNCE) as a part of our loss function to maximize the mutual information between the anchor genes and genes with the same functions. This loss is applicable to different genes in two arbitrary graphs during the training process. In general, if we represent the embeddings of  $N$  genes as  $\text{Gene}_N = \{e_1, \dots, e_N\}$ , InfoNCE is designed to minimize:

$$\mathcal{L}_{\text{InfoNCE}} = -\mathbb{E} \left[ \log \frac{\exp(e \cdot k_+ / \tau)}{\sum_{i=0}^K \exp(e \cdot k_i / \tau)} \right], \quad (6)$$

where samples  $\{k_0, k_1, k_2, \dots\}$  compose a set of gene embeddings known as keys of one dictionary and  $e$  is a query gene embedding.  $k_+$  represents a positive sample of  $e$  and  $k_i$  denotes a negative sample of  $e$ . Equation 6 can be interpreted as a log loss of a  $(K + 1)$ -way Softmax classifier, which attempts to classify  $e$  as  $k_+$ .  $\tau$  is a temperature parameter.  $\tau$  is set to 0.07 referenced in MoCo [41].

### 3.7 Final Loss Function

In summary, the training objective of MuSe-GNN for graph  $i$  comprises three components:

$$\begin{aligned} & \min_{e_i, e_j} \mathcal{L}_{\text{BCE}}(\text{DECODERMLP}_i(e_i e_i^T), E_i) - \mathbb{E} [\text{COSSIM}(e_i[\text{Common}_{ij}], e_j[\text{Common}_{ij}]) \lambda_{ij}^T] \\ & + \lambda_c \mathcal{L}_{\text{InfoNCE}}(e_i[\text{Diff}_i] \oplus e_j[\text{Diff}_j], e_i[\text{Diff}_{\mathcal{N}(i)}] \oplus e_j[\text{Diff}_{\mathcal{N}(j)}]), \end{aligned} \quad (7)$$

where  $\text{Common}_{ij}$  denotes the index set for common HVGs,  $\text{Diff}_i$  represents the index set for different HVGs in graph  $i$ , and  $\text{Diff}_{\mathcal{N}(i)}$  indicates the index set for neighbors of  $\text{Diff}_i$  in graph  $i$ . To expedite the training process and conserve memory usage, we sample graph  $j$  for each graph  $i$  during model training. We also employ multi-thread programming to accelerate the index set extraction process.  $\lambda_c$  is the weight for InfoNCE loss. All of the components in MuSe-GNN are supported by ablation experiments in Appendix E.1. Details of hyper-parameter tuning can be found in Appendix E.2.

## 4 Experiments

**Datasets & Embeddings generation.** Information on the different datasets used for different experiments is included at the beginning of each paragraph. The training algorithm of our model

---

**Algorithm 1** Multimodal Similarity Learning Graph Neural Network (MuSe-GNN)

---

**Input:** Model ENCODERGNN, DECODERMLP; Dataset  $\mathcal{D} = (\{V_i, E_i\})_{i=1}^T$ ; Number of epochs  $K$ ; Neighbor overlap list  $\lambda$ ; Weight coefficient for contrastive learning  $\lambda_c$ ;

**Helper Functions:** A function to find the genes with the same name FINDCOMMONGENES; A function to find the genes with different name FINDDIFFGENES; A function to calculate cosine similarity COSSIM; A function to sample data from a graph collection SAMPLE; A optimizer function to update weights ADAM.

**Output:** Model ENCODERGNN

- 1: INIT: initialize all parameters.
- 2: **for**  $s$  in  $K$  steps **do**
- 3:   **for** id,  $\{V, E\}$  in enumerate( $\mathcal{D}$ ) **do**
- 4:      $e \leftarrow \text{ENCODERGNN}(\{V, E\}; \theta_{enc})$ ;  $E_{rec} \leftarrow \text{DECODERMLP}(ee^T; \theta_{dec})$
- 5:      $\mathcal{L} \leftarrow \mathcal{L}_{\text{BCE}}(E_{rec}, E)$
- 6:      $\text{id}_{\text{new}}, \{V_{\text{new}}, E_{\text{new}}\} \leftarrow \text{SAMPLE}(G \setminus \{V, E\})$
- 7:      $e_{\text{new}} \leftarrow \text{ENCODERGNN}(\{V_{\text{new}}, E_{\text{new}}\}; \theta_{enc})$
- 8:      $e_{\text{Common}}, e'_{\text{Common}} \leftarrow \text{FINDCOMMONGENES}(e, e_{\text{new}})$
- 9:      $\mathcal{L} \leftarrow \mathcal{L} - \text{COSSIM}(e_{\text{Common}}, e'_{\text{Common}}) \lambda_{\text{id}, \text{id}_{\text{new}}}^T$
- 10:     $e_{\text{Diff}}, e'_{\text{Diff}}, e_{\text{Diff}_N}, e'_{\text{Diff}_N} \leftarrow \text{FINDDIFFGENES}(e, e_{\text{new}})$
- 11:     $\mathcal{L} \leftarrow \mathcal{L} + \lambda_c \mathcal{L}_{\text{InfoNCE}}(e_{\text{Diff}} \oplus e_{\text{Diff}_N}, e'_{\text{Diff}} \oplus e'_{\text{Diff}_N})$
- 12:     $\theta_{enc} \leftarrow \text{ADAM}(\mathcal{L}, \theta_{enc})$ ;  $\theta_{dec} \leftarrow \text{ADAM}(\mathcal{L}, \theta_{dec})$
- 13: **Return** ENCODERGNN

---

is outlined in Algorithm 1. We stored the gene embeddings in the AnnData structure provided by Scanpy [104]. To identify groups of genes with similar functions, we applied the Leiden clustering algorithm [91] to the obtained gene embeddings. Details can be found in Appendix E.4.

**Evaluation metrics.** For the benchmarking process, we used six metrics: edge AUC (AUC) [74], common genes ASW (ASW) [59], common genes graph connectivity (GC) [59], common genes iLISI (iLISI) [59], common genes ratio (CGR), and neighbors overlap (NO) to provide a comprehensive comparison. Detailed descriptions of these metrics can be found in Appendix H. We computed the metrics for the different methods and calculated the average rank (Avg Rank  $\in [1, 9]$ ). Moreover, to evaluate the model performance improvement, we applied min-max scaling to every metric across different models and computed the average score (Avg Score  $\in [0, 1]$ ).

**Baselines.** We selected eight models as competitors for MuSe-GNN. The first group of methods stems from previous work on learning embeddings for biomedical data, including Principal Component Analysis (PCA) [74] used by GSS, Gene2vec, GIANT (the SOTA model for gene representation learning), Weight-sharing Multi-layer Auto-encoder (WSMAE) used by [109] and scBERT (the SOTA model with pre-training for cell type annotation). The second group of methods comprises common unsupervised learning baseline models, including GAE, VGAE [48] and Masked Auto-encoder (MAE) (the SOTA model for self-supervised learning) [42]. MuSe-GNN, GIANT, GAE, VGAE, WSMAE and MAE have training parameter sizes between 52.5 and 349 M, and all models are tuned to their best performance. Details are shown in Appendix E.2.

**Biological applications.** For the pathway analysis, we used Gene Ontology Enrichment Analysis (GOEA) [1] to identify specific biological pathways enriched in distinct gene clusters with common functions. Moreover, we used Ingenuity Pathway Analysis (IPA) [2] to extract biological information from the genes within various clusters, including causal networks [50] and sets of diseases and biological functions. Biological pathways refer to processes identified based on the co-occurrence of genes within a particular cluster. The causal network depicts the relationships between regulatory genes and their target genes. Disease and biological function sets facilitate the discovery of key processes and complications associated with specific diseases. Using gene embeddings also improves the performance of models for gene function prediction. To visualize gene embeddings in a low-dimensional space, we utilized Uniform Manifold Approximation and Projection (UMAP) [63].

Table 1: Avg Score across different tissues. Standard deviations are reported in Appendix E.3.

Methods	Heart	Lung	Liver	Kidney	Thymus	Spleen	Pancreas	Cerebrum	Cerebellum	PBMC
PCA	0.52	0.48	0.56	0.47	0.56	0.60	0.51	0.62	0.53	0.51
Gene2vec	0.40	0.37	0.33	0.29	0.21	0.31	0.24	0.27	0.31	0.19
GIANT	0.50	0.40	0.33	0.38	0.58	0.33	0.56	0.29	0.28	0.28
WSMAE	0.50	0.47	0.54	0.46	0.57	0.53	0.52	0.55	0.59	0.50
GAE	0.61	0.45	0.58	0.40	0.56	0.58	0.52	0.56	0.60	0.54
VGAE	0.64	0.32	0.33	0.38	0.56	0.31	0.33	0.41	0.33	0.47
MAE	0.36	0.47	0.50	0.45	0.41	0.52	0.39	0.50	0.49	0.50
scBERT	0.41	0.49	0.55	0.62	0.17	0.58	0.46	0.60	0.61	0.58
MuSeGNN	<b>0.77</b>	<b>0.96</b>	<b>0.92</b>	<b>0.89</b>	<b>0.89</b>	<b>0.94</b>	<b>0.80</b>	<b>0.95</b>	<b>0.90</b>	<b>0.92</b>

#### 4.1 Benchmarking Analysis

We executed each method 10 times by using the same setting of seeds to show the statistical significance based on datasets across different tissues. The performance comparison of nine gene embedding methods is presented in Tables 1 and 14. Based on these two tables, MuSe-GNN outperformed its competitors in terms of both average ranks and average scores across all the tissues. Based on Table 1, for major tissues, such as heart and lung, MuSe-GNN’s performance was 20.1% higher than the second-best method and 97.5% higher than the second-best method in heart and lung tissue, respectively. According to Appendix E.3, MuSe-GNN’s stability was also demonstrated through various metrics by comparing standard deviations, including AUC, GC, and NO. In contrast, methods such as Gene2vec, GAE, VGAE, MAE and scBERT exhibited significant instability in their evaluation results for kidney or thymus. Consequently, we concluded that MuSe-GNN is the best performing model for learning gene representation based on datasets from different tissues, making it applicable to learn gene embeddings from diverse multimodal biological data.

#### 4.2 Analysis of Gene Embeddings from Multimodal Biological Data

In Figure 3, we displayed the integration results for multimodal biological data from Humans. Figure 3 (a) and (b) demonstrated that MuSe-GNN could successfully integrate genes from different modalities into a co-embedded space, allowing us to identify functional groups using the Leiden algorithm shown in Figure 3 (c). Furthermore, Figure 3 (d) revealed that most of the clusters in (c) were shared across different modalities. We also identified three significant functional groups in Figure 3 (a): the nervous system (predominantly composed of the cerebrum and cerebellum [26]), the cardiovascular system (mainly composed of heart, lung, and kidney [29]), and the immunology system (primarily consisting of spleen, liver, and peripheral blood mononuclear cells (PBMC) [70]). All systems are important in regulating the life activities of the body. We also uncovered a pre-epigenetics group (mainly consisting of scATAC-seq data without imprinting, modification, or editing), emphasizing the biological gap existing in multi-omics and the importance of post-transcriptional regulation.

Using GOEA, we could identify significant pathways enriched by different co-embedded gene clusters. For instance, Figure 3 (d) displayed the top 5 pathways in an immunology system cluster. The rank was calculated based on the negative logarithm of the false discovery rate. Since all top pathways were related to immunological defense and response, it further supported the accuracy of our embeddings in representing gene functions. For our analysis of shared transcription factors and major pathways across all the tissues, please refer to Appendix I. For our analysis of multi-species gene embeddings, please refer to Appendix J.

#### 4.3 Analysis of Gene Embeddings for Diseases

We generated gene embeddings for human pancreas cells from samples with and without SARS-CoV-2 infection, as depicted in Figure 4 (a) and (b). We identified specific genes from COVID samples that did not align with control samples, which piqued our interest. These genes, highlighted by a red circle in Figure 4 (c), could be interpreted as differentially functional genes in diseased cells.

We conducted GOEA for the genes of interest and discovered a close relationship among these gene enrichment results and the top 5 pathways associated with immune activity. These results are displayed in Figure 4 (d). For the genes within our target cluster, we utilized IPA to identify the Entrez name of these genes, and 90.3% (122/135) genes in our cluster are related to immunoglobulin.

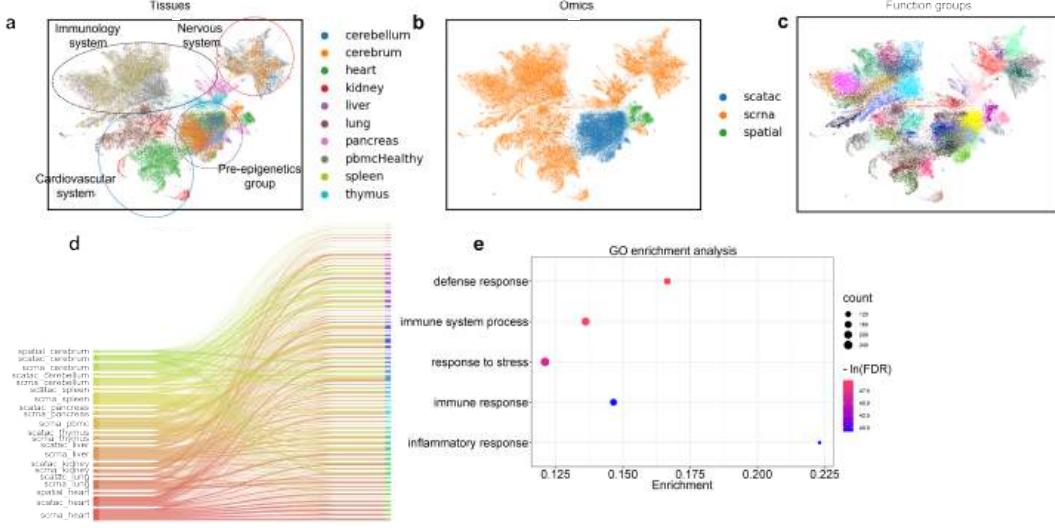


Figure 3: Gene representation learning results for multimodal biological data. (a) represents the UMAPs of gene embeddings colored by tissue type and highlighted by biological system. (b) represents the UMAPs of gene embeddings colored by omics type. (c) represents the UMAPs of gene embeddings colored by common function groups. (d) is a Sankey plot [80] to show the overlap of different modalities in the same clusters. (e) shows the top5 pathways related to the genes in the special cluster discovered by GOEA. The bubble plots in this paper were created based on ggplot2 [102].

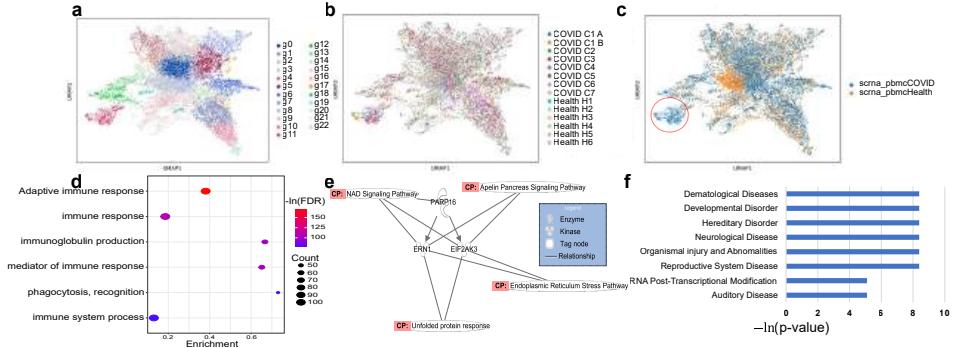


Figure 4: Gene embeddings from COVID samples and healthy samples. (a) represents the UMAPs of gene embeddings colored by functional groups. (b) represents the UMAPs of gene embeddings colored by datasets. (c) represents the gene embeddings colored by the conditions, and the red circle reflects the differential co-expression genes. (d) shows the top6 pathways related to the genes in the special cluster discovered by GOEA. (e) represents the causal network existing in the special cluster discovered by IPA. (f) represents the top diseases & biological functions discovered by IPA.

We could also infer the causal relationship existing in the gene regulatory activity of the immune system. For example, Figure 4 (e) showed a causal network inferred by IPA based on our genes cluster. PARP16, as an enzyme, can regulate ERN1 and EIF2AK3, and certain pathways are also related to this causal network. Moreover, we also showed the relation between the set of genes and Disease & Bio functions in Figure 4 (f). We identified top related Diseases & Bio functions ranked by negative logarithm of p-value, and all of these diseases could be interpreted as complications that may arise from new coronavirus infection [6, 64, 79, 69, 33, 7, 83, 84]. Our extra analyses for lung cancer data can be found in Appendix K.

#### 4.4 Analysis of Gene Embeddings for Gene Function Prediction.

Here we intend to predict the dosage-sensitivity of genes related to genetic diagnosis (as dosage-sensitive or not) [90]. We used MuSe-GNN to generate gene embeddings for different datasets based on an unsupervised learning framework and utilized the gene embeddings as training dataset to predict the function of genes based on k-NN classifier. k-NN classifier is a very naive model and can reflect the contribution of gene embeddings in the prediction task.

In this task, we evaluated the performance of MuSe-GNN based on the dataset used in Geneformer [90, 31], comparing it to the prediction results based on raw data or Geneformer (total supervised learning). As shown in Table 2, the prediction accuracy based on gene emebddings from MuSe-GNN is the highest one. Moreover, the performance of gene embeddings from MuSe-GNN is better than Geneformer, which is a totally supervised learning model. Such finding proves the advantages of MuSe-GNN in the application of gene function prediction task. Further application analysis can be found in Appendix L.

Table 2: Accuracy for dosage-sensitivity prediction

	<b>MuSe-GNN (unsup)</b>	<b>Geneformer (sup)</b>	<b>Raw</b>
Accuracy	0.77±0.01	0.74±0.06	0.75±0.01

## 5 Conclusion

In this paper, we introduce MuSe-GNN, a model based on Multimodal Machine Learning and Deep Graph Neural Networks, for learning gene embeddings from multi-context sequencing profiles. Through experiments on various multimodal biological datasets, we demonstrate that MuSe-GNN outperforms current gene embedding learning models across different metrics and can effectively learn the functional similarity of genes across tissues and techniques. Moreover, we performed various biological analyses using the learned gene embeddings, leveraging the capabilities of GOEA and IPA, such as identifying significant pathways, detecting diseases, and inferring causal networks. Our model can also contribute to the study of the pathogenic mechanisms of diseases like COVID and lung cancer, and improve the prediction performance for gene functions. Overall, the gene representations learned by MuSe-GNN are highly versatile and can be applied to different analysis frameworks.

At present, MuSe-GNN does not accept graphs with nodes other than genes as input. In the future, we plan to explore more efficient approaches for training large models related to Multimodal Machine Learning and extend MuSe-GNN to a more general version capable of handling a broader range of multimodal biological data.

## 6 Acknowledgements

This research was supported in part by NIH R01 GM134005, R56 AG074015, and NSF grant DMS 1902903 to H.Z. We appreciate the comments, feedback, and suggestions from Chang Su, Zichun Xu, Xinning Shan, Yuhang Xie, Mingze Dong, and Maria Brbic.

## References

- [1] Gene ontology enrichment analysis. URL <http://geneontology.org>.
- [2] Ingenuity pathway analysis. URL <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>.
- [3] 10x Genomics. 10x genomics acquires spatial transcriptomics. *Science*, 2018. URL <https://www.10xgenomics.com/news/10x-genomics-acquires-spatial-transcriptomics>.
- [4] Adrian M Altenhoff, Clément-Marie Train, Kimberly J Gilbert, Ishita Mediratta, Tarcisio Mendes de Farias, David Moi, Yannis Nevers, Hale-Seda Radoykova, Victor Rossier, Alex

- Warwick Vesztrocy, et al. Oma orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic acids research*, 49(D1):D373–D379, 2021.
- [5] Tallulah S Andrews and Martin Hemberg. False signals induced by single-cell imputation. *F1000Research*, 7, 2018.
  - [6] Khashayar Aram, Anant Patil, Mohamad Goldust, and Fateme Rajabi. Covid-19 and exacerbation of dermatological diseases: a review of the available literature. *Dermatologic Therapy*, 34(6):e15113, 2021.
  - [7] Arash Ardestani Zadeh and Davood Arab. Covid-19 and male reproductive system: pathogenic features and possible mechanisms. *Journal of molecular histology*, 52:869–878, 2021.
  - [8] Ricard Argelaguet, Anna SE Cuomo, Oliver Stegle, and John C Marioni. Computational principles and challenges in single-cell data integration. *Nature biotechnology*, 39(10):1202–1215, 2021.
  - [9] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443, 2018.
  - [10] Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.
  - [11] A Sina Booeshaghi, Ingileif B Hallgrímsdóttir, Ángel Gálvez-Merchán, and Lior Pachter. Depth normalization for single-cell genomics count data. *bioRxiv*, pages 2022–05, 2022.
  - [12] Andrew Butler, Paul Hoffman, Peter Smibert, Efthymia Papalexis, and Rahul Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology*, 36(5):411–420, 2018.
  - [13] Tianle Cai, Shengjie Luo, Keyulu Xu, Di He, Tie-yan Liu, and Liwei Wang. Graphnorm: A principled approach to accelerating graph neural network training. In *International Conference on Machine Learning*, pages 1204–1215. PMLR, 2021.
  - [14] Junyue Cao, Diana R O’Day, Hannah A Pliner, Paul D Kingsley, Mei Deng, Riza M Daza, Michael A Zager, Kimberly A Aldinger, Ronnie Blecher-Gonen, Fan Zhang, et al. A human cell atlas of fetal gene expression. *Science*, 370(6518):eaba7721, 2020.
  - [15] R Caruana. Multitask learning: A knowledge-based source of inductive bias1. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 41–48. Citeseer, 1993.
  - [16] Hao Chen, Nam D Nguyen, Matthew Ruffalo, and Ziv Bar-Joseph. A unified analysis of atlas single cell data. *bioRxiv*, 2022.
  - [17] Xi Chen, Ricardo J Miragaia, Kedar Nath Natarajan, and Sarah A Teichmann. A rapid and robust method for single cell chromatin accessibility profiling. *Nature communications*, 9(1):1–9, 2018.
  - [18] Kenneth Ward Church. Word2vec. *Natural Language Engineering*, 23(1):155–162, 2017.
  - [19] Kelsy C Cotto, Alex H Wagner, Yang-Yang Feng, Susanna Kiwala, Adam C Coffman, Gregory Spies, Alex Wollam, Nicholas C Spies, Obi L Griffith, and Malachi Griffith. Dgidb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic acids research*, 46(D1):D1068–D1073, 2018.
  - [20] Darren A Cusanovich, Riza Daza, Andrew Adey, Hannah A Pliner, Lena Christiansen, Kevin L Gunderson, Frank J Steemers, Cole Trapnell, and Jay Shendure. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, 348(6237):910–914, 2015.
  - [21] C Domínguez Conde, C Xu, LB Jarvis, DB Rainbow, SB Wells, T Gomes, SK Howlett, O Suchanek, K Polanski, HW King, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594):eabI5197, 2022.

- [22] Mingze Dong and Yuval Kluger. Towards understanding and reducing graph structural noise for gnns. 2023.
- [23] Rui Dong and Guo-Cheng Yuan. Spatialdwls: accurate deconvolution of spatial transcriptomic data. *Genome biology*, 22(1):145, 2021.
- [24] Jingcheng Du, Peilin Jia, Yulin Dai, Cui Tao, Zhongming Zhao, and Degui Zhi. Gene2vec: distributed representation of genes based on co-expression. *BMC genomics*, 20(1):7–15, 2019.
- [25] Simon Shaolei Du, Wei Hu, Sham M. Kakade, Jason D. Lee, and Qi Lei. Few-shot learning via learning the representation, provably. 2021. URL <https://openreview.net/forum?id=pW2Q2xLwIMD>.
- [26] Ramazan Durmaz, Metin Ant Atasoy, Gü̈l Durmaz, Baki Adapinar, Ali Arslantaş, Aydin Aydinli, and Esref Tel. Multiple nocardial abscesses of cerebrum, cerebellum and spinal cord, causing quadriplegia. *Clinical neurology and neurosurgery*, 103(1):59–62, 2001.
- [27] Yasha Ektefaie, George Dasoulas, Ayush Noori, Maha Farhat, and Marinka Zitnik. Multimodal learning with graphs. *Nature Machine Intelligence*, pages 1–11, 2023.
- [28] HuBMAP Consortium et al. The human body at cellular resolution: the nih human biomolecular atlas program. *Nature*, 574(7777):187–192, 2019.
- [29] HE Fessler. Heart-lung interactions: applications in the critically ill. *European Respiratory Journal*, 10(1):226–237, 1997.
- [30] David S Fischer, Anna K Fiedler, Eric M Kernfeld, Ryan MJ Genga, Aimée Bastidas-Ponce, Mostafa Bakhti, Heiko Lickert, Jan Hasenauer, Rene Maehr, and Fabian J Theis. Inferring population dynamics from single-cell rna-sequencing time series data. *Nature biotechnology*, 37(4):461–468, 2019.
- [31] Oscar Franzén, Li-Ming Gan, and Johan LM Björkegren. Panglaodb: a web server for exploration of mouse and human single-cell rna sequencing data. *Database*, 2019:baz046, 2019.
- [32] Toni Gabaldón and Eugene V Koonin. Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, 14(5):360–366, 2013.
- [33] Suhita Gayen Nee’Betal, Pedro Urday, Huda B Al-Kouatly, Kolawole Solarin, Joanna SY Chan, Sankar Addya, Rupsa C Boelig, and Zubair H Aghai. Covid-19 infection during pregnancy induces differential gene expression in human cord blood cells from term neonates. *Frontiers in Pediatrics*, 10:547, 2022.
- [34] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.
- [35] Christoph Hafemeister and Rahul Satija. Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression. *Genome biology*, 20(1):1–15, 2019.
- [36] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.
- [37] Heonjong Han, Jae-Won Cho, Sangyoung Lee, Ayoung Yun, Hyojin Kim, Dasom Bae, Sunmo Yang, Chan Yeong Kim, Muyoung Lee, Eunbeen Kim, et al. Trrust v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research*, 46(D1):D380–D386, 2018.
- [38] Xiaoping Han, Ziming Zhou, Lijiang Fei, Huiyu Sun, Renying Wang, Yao Chen, Haide Chen, Jingjing Wang, Huanna Tang, Wenhao Ge, et al. Construction of a human cell landscape at single-cell level. *Nature*, 581(7808):303–309, 2020.

- [39] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.
- [40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [41] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.
- [42] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16000–16009, 2022.
- [43] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 594–604, 2022.
- [44] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.
- [45] Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. What makes multi-modal learning better than single (provably). *Advances in Neural Information Processing Systems*, 34:10944–10956, 2021.
- [46] Byungjin Hwang, Ji Hyun Lee, and Duhee Bang. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental & molecular medicine*, 50(8):1–14, 2018.
- [47] Grigory Khromov and Sidak Pal Singh. Some fundamental aspects about lipschitz continuity of neural network functions. *arXiv preprint arXiv:2302.10886*, 2023.
- [48] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *NIPS Workshop on Bayesian Deep Learning*, 2016.
- [49] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL <https://openreview.net/forum?id=SJU4ayYg1>.
- [50] Andreas Krämer, Jeff Green, Jack Pollard Jr, and Stuart Tugendreich. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*, 30(4):523–530, 2014.
- [51] M Krassowski. Complexupset: Create complex upset plots using ggplot2 components. *R package version 0.5*, 18, 2020.
- [52] Peter Langfelder and Steve Horvath. Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics*, 9(1):1–13, 2008.
- [53] Patty J Lee, Philip Blood, Katy Börner, Judith Campisi, Feng Chen, Heike Daldrup-Link, Phil De Jager, Li Ding, Francesca E Duncan, Oliver Eickelberg, et al. Nih sennet consortium: Mapping senescent cells in the human body to understand health and disease, 2022.
- [54] Jeffrey T Leek, Robert B Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739, 2010.
- [55] Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions. *arXiv preprint arXiv:2209.03430*, 2022.
- [56] Xiang Lin, Tian Tian, Zhi Wei, and Hakon Hakonarson. Clustering of single-cell multi-omics data with a multimodal deep learning method. *Nature communications*, 13(1):7705, 2022.

- [57] Yinyin Liu, Janusz A Starzyk, and Zhen Zhu. Optimized approximation algorithm in neural networks without overfitting. *IEEE transactions on neural networks*, 19(6):983–995, 2008.
- [58] Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.
- [59] Malte D Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F Müller, Daniel C Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods*, 19(1):41–50, 2022.
- [60] Chongyuan Luo, Christopher L Keown, Laurie Kurihara, Jingtian Zhou, Yupeng He, Junhao Li, Rosa Castanon, Jacinta Lucero, Joseph R Nery, Justin P Sandoval, et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science*, 357(6351):600–604, 2017.
- [61] Kaili Ma, Haochen Yang, Han Yang, Tatiana Jin, Pengfei Chen, Yongqiang Chen, Barakeel Fanseu Kamhoua, and James Cheng. Improving graph representation learning by contrastive regularization. *arXiv preprint arXiv:2101.11525*, 2021.
- [62] Samuel Marguerat and Jürg Bähler. Rna-seq: from technology to biology. *Cellular and molecular life sciences*, 67:569–579, 2010.
- [63] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861. URL <https://doi.org/10.21105/joss.00861>.
- [64] Bekir Fatih Meral. Parental views of families of children with autism spectrum disorder and developmental disorders during the covid-19 pandemic. *Journal of Autism and Developmental Disorders*, 52(4):1712–1724, 2022.
- [65] Diganta Misra. Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681*, 2019.
- [66] Yujie Mo, Liang Peng, Jie Xu, Xiaoshuang Shi, and Xiaofeng Zhu. Simple unsupervised graph representation learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7):7797–7805, Jun. 2022. doi: 10.1609/aaai.v36i7.20748. URL <https://ojs.aaai.org/index.php/AAAI/article/view/20748>.
- [67] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.
- [68] Ryan M Mulqueen, Dmitry Pokholok, Steven J Norberg, Kristof A Torkenczy, Andrew J Fields, Duanchen Sun, John R Sinnamon, Jay Shendure, Cole Trapnell, Brian J O’Roak, et al. Highly scalable generation of dna methylation profiles in single cells. *Nature biotechnology*, 36(5):428–431, 2018.
- [69] Edward J Needham, Sherry H-Y Chou, Alasdair J Coles, and David K Menon. Neurological implications of covid-19 infections. *Neurocritical care*, 32:667–671, 2020.
- [70] Rana Nikzad, Laura S Angelo, Kevin Aviles-Padilla, Duy T Le, Vipul K Singh, Lynn Bimler, Milica Vukmanovic-Stejic, Elena Vendrame, Thanmayi Ranganath, Laura Simpson, et al. Human natural killer cells mediate adaptive immunity to viral antigens. *Science immunology*, 4(35):eaat8116, 2019.
- [71] Sehyun Oh, Ludwig Geistlinger, Marcel Ramos, Daniel Blankenberg, Marius van den Beek, Jaclyn N Taroni, Vincent J Carey, Casey S Greene, Levi Waldron, and Sean Davis. Genomicsupersignature facilitates interpretation of rna-seq experiments through robust, efficient comparison to public databases. *Nature Communications*, 13(1):3695, 2022.
- [72] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

- [73] Subarna Palit, Christoph Heuser, Gustavo P De Almeida, Fabian J Theis, and Christina E Zielinski. Meeting the challenges of high-dimensional single-cell data analysis in immunology. *Frontiers in immunology*, 10:1515, 2019.
- [74] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [75] Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35:14501–14515, 2022.
- [76] Yanay Rosen, Maria Brbić, Yusuf Roohani, Kyle Swanson, Ziang Li, and Jure Leskovec. Towards universal cell embeddings: Integrating single-cell rna-seq datasets across species with saturn. *Biorxiv: the Preprint Server for Biology*, 2023.
- [77] Orit Rozenblatt-Rosen, Michael JT Stubbington, Aviv Regev, and Sarah A Teichmann. The human cell atlas: from vision to reality. *Nature*, 550(7677):451–453, 2017.
- [78] Antoine-Emmanuel Saliba, Alexander J Westermann, Stanislaw A Gorski, and Jörg Vogel. Single-cell rna-seq: advances and future challenges. *Nucleic acids research*, 42(14):8845–8860, 2014.
- [79] Amanda Salih, Aaron Chin, Manisha Gandhi, Amir Shamshirsaz, Hennie Lombaard, and Joud Hajjar. Hereditary angioedema and covid-19 during pregnancy: Two case reports. *The Journal of Allergy and Clinical Immunology: In Practice*, 11(3):961–962, 2023.
- [80] Pierre Sassoulas, Anazalea, and Marcomanz. pySankey, 2018. URL <https://github.com/anazalea/pySankey>.
- [81] Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjing Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1548–1554. International Joint Conferences on Artificial Intelligence Organization, 8 2021. doi: 10.24963/ijcai.2021/214. URL <https://doi.org/10.24963/ijcai.2021/214>. Main Track.
- [82] Michael A Skinnider, Jordan W Squair, and Leonard J Foster. Evaluating measures of association for single-cell transcriptomics. *Nature methods*, 16(5):381–386, 2019.
- [83] Rajneesh Srivastava, Swapna Vidhur Daulatabad, Mansi Srivastava, and Sarath Chandra Janga. Role of sars-cov-2 in altering the rna-binding protein and mirna-directed post-transcriptional regulatory networks in humans. *International journal of molecular sciences*, 21(19):7090, 2020.
- [84] Won Sriwijitalai and Viroj Wiwanitkit. Hearing loss and covid-19: a note. *American journal of otolaryngology*, 2020.
- [85] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 14(9):865–868, 2017.
- [86] BJ Strober, Reem Elorbany, K Rhodes, Nirmal Krishnan, Karl Tayeb, Alexis Battle, and Yoav Gilad. Dynamic genetic regulation of gene expression during cellular differentiation. *Science*, 364(6447):1287–1290, 2019.
- [87] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexis, William M Mauck, Yuhao Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902, 2019.
- [88] Chang Su, Zichun Xu, Xinning Shan, Biao Cai, Hongyu Zhao, and Jingfei Zhang. Cell-type-specific co-expression inference from single cell rna-sequencing data. *Nature Communications*, 14(1):4846, 2023.

- [89] Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of machine learning research*, 8(5), 2007.
- [90] Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. Transfer learning enables predictions in network biology. *Nature*, pages 1–9, 2023.
- [91] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.
- [92] Nilesh Tripuraneni, Michael Jordan, and Chi Jin. On the theory of transfer learning: The importance of task diversity. *Advances in neural information processing systems*, 33:7852–7862, 2020.
- [93] Nilesh Tripuraneni, Chi Jin, and Michael Jordan. Provable meta-learning of linear representations. In *International Conference on Machine Learning*, pages 10434–10443. PMLR, 2021.
- [94] David Van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J Carr, Cassandra Burdziak, Kevin R Moon, Christine L Chaffer, Diwakar Pattabiraman, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3):716–729, 2018.
- [95] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [96] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=rJXMpikCZ>.
- [97] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.
- [98] Guangtao Wang, Rex Ying, Jing Huang, and Jure Leskovec. Multi-hop attention graph neural networks. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3089–3096. International Joint Conferences on Artificial Intelligence Organization, 8 2021. doi: 10.24963/ijcai.2021/425. URL <https://doi.org/10.24963/ijcai.2021/425>. Main Track.
- [99] Juexin Wang, Anjun Ma, Yuzhou Chang, Jianting Gong, Yuexu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, and Dong Xu. scgnn is a novel graph neural network framework for single-cell rna-seq analyses. *Nature communications*, 12(1):1–11, 2021.
- [100] Yue Wang, Shafiq Joty, Michael Lyu, Irwin King, Caiming Xiong, and Steven C.H. Hoi. VD-BERT: A Unified Vision and Dialog Transformer with BERT. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3325–3338, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.269. URL <https://aclanthology.org/2020.emnlp-main.269>.
- [101] Yuge Wang, Tianyu Liu, and Hongyu Zhao. Respan: a powerful batch correction model for scRNA-seq data through residual adversarial networks. *Bioinformatics*, 38(16):3942–3949, 2022.
- [102] Hadley Wickham. ggplot2. *Wiley interdisciplinary reviews: computational statistics*, 3(2):180–185, 2011.
- [103] Addie Woicik, Mingxin Zhang, Hanwen Xu, Sara Mostafavi, and Sheng Wang. Gemini: memory-efficient integration of hundreds of gene networks with high-order pooling. *Bioinformatics*, 39, 06 2023. doi: 10.1093/bioinformatics/btad247.

- [104] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):1–5, 2018.
- [105] Renjie Xie, Wei Xu, Yanzhi Chen, Jiabao Yu, Aiqun Hu, Derrick Wing Kwan Ng, and A. Lee Swindlehurst. A generalizable model-and-data driven approach for open-set rff authentication. *IEEE Transactions on Information Forensics and Security*, 16:4435–4450, 2021. doi: 10.1109/TIFS.2021.3106166.
- [106] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=ryGs6iA5Km>.
- [107] Mingcong Xu, Xuefeng Bai, Bo Ai, Guorui Zhang, Chao Song, Jun Zhao, Yuezhu Wang, Ling Wei, Fengcui Qian, Yanyu Li, et al. Tf-marker: a comprehensive manually curated database for transcription factors and related markers in specific cell and tissue types in human. *Nucleic acids research*, 50(D1):D402–D412, 2022.
- [108] Fan Yang, Wenchuan Wang, Fang Wang, Yuan Fang, Duyu Tang, Junzhou Huang, Hui Lu, and Jianhua Yao. scbert as a large-scale pretrained deep language model for cell type annotation of single-cell rna-seq data. *Nature Machine Intelligence*, 4(10):852–866, 2022.
- [109] Karren Dai Yang, Anastasiya Belyaeva, Saradha Venkatachalam, Karthik Damodaran, Abigail Katcoff, Adityanarayanan Radhakrishnan, GV Shivashankar, and Caroline Uhler. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nature communications*, 12(1):1–10, 2021.
- [110] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems*, 34:28877–28888, 2021.
- [111] Bohang Zhang, Shengjie Luo, Liwei Wang, and Di He. Rethinking the expressive power of GNNs via graph biconnectivity. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=r9hNv76KoT3>.
- [112] Di Zhang, Yanxiang Deng, Petra Kukanja, Eneritz Agirre, Marek Bartosovic, Mingze Dong, Cong Ma, Sai Ma, Graham Su, Shuzhen Bao, et al. Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature*, pages 1–10, 2023.
- [113] Shuang Zhang, Rui Fan, Yuti Liu, Shuang Chen, Qiao Liu, and Wanwen Zeng. Applications of transformer-based language models in bioinformatics: A survey. *Bioinformatics Advances*, 2023.
- [114] Grace XY Zheng, Jessica M Terry, Phillip Belgrader, Paul Ryvkin, Zachary W Bent, Ryan Wilson, Solongo B Ziraldo, Tobias D Wheeler, Geoff P McDermott, Junjie Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8(1):1–12, 2017.
- [115] Kaixiong Zhou, Xiao Huang, Qingquan Song, Rui Chen, and Xia Hu. Auto-gnn: Neural architecture search of graph neural networks. *Frontiers in Big Data*, 5, 2022. ISSN 2624-909X. doi: 10.3389/fdata.2022.1029307. URL <https://www.frontiersin.org/articles/10.3389/fdata.2022.1029307>.
- [116] Jiaqiang Zhu, Shiquan Sun, and Xiang Zhou. Spark-x: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biology*, 22(1):1–25, 2021.
- [117] Jiong Zhu, Ryan A. Rossi, Anup Rao, Tung Mai, Nedim Lipka, Nesreen K. Ahmed, and Danai Koutra. Graph neural networks with heterophily. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):11168–11176, May 2021. doi: 10.1609/aaai.v35i12.17332. URL <https://ojs.aaai.org/index.php/AAAI/article/view/17332>.
- [118] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*, 2020.
- [119] Marinka Zitnik and Jure Leskovec. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*, 33(14):i190–i198, 2017.