bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MARVEL: Microenvironment Annotation by Supervised Graph Contrastive Learning

Yan Cui[1], Hongzhi Wen[1,2], Robert Yang[1], Xi Luo[1], Hui Liu[2], and Yuyin Xie[2]

[1] CartaBio Inc
[2] Department of Computational Mathematics, Science and Engineering; Michigan State University
{yancui, hongzhiwen, rob, xiluo}@cartabio.ai
{liuhui7, xyy}@msu.edu

Abstract. Recent advancements in *in situ* molecular profiling technologies, including spatial proteomics and transcriptomics, have enabled detailed characterization of the microenvironment at cellular and subcellular levels. While these techniques provide rich information about individual cells' spatial coordinates and expression profiles, extracting biologically meaningful spatial structures from the data remains a significant challenge. Current methodologies often rely on unsupervised clustering followed by cell type annotation based on differentially expressed genes within each cluster and most of the time will require other information as the reference (e.g., HE-stained images). This is labor-intensive and demands extensive domain knowledge. To address these challenges, we propose a supervised graph contrastive learning framework, MARVEL. MARVEL is a supervised graph contrastive learning method that can effectively embed local microenvironments represented by cell neighbor graphs into a continuous representation space, facilitating various downstream microenvironment annotation scenarios. By leveraging partially annotated examples as strong positives, our approach mitigates the common issues of false positives encountered in conventional graph contrastive learning. Using real-world annotated data, we demonstrate that MARVEL outperforms existing methods in three key microenvironment-related tasks: transductive microenvironment annotation, inductive microenvironment querying, and the identification of novel microenvironments across different slices.

Keywords: Spatial Omics · Graph Contrasitve Learning

1 Introduction

The tissue microenvironment (ME), a dynamic mix of cellular and non-cellular components, forms a complex regulatory network, essential for maintaining the homeostasis of a tissue [2]. The rapid development of *in situ* molecular profiling techniques, such as spatial proteomics (e.g., CODEX [4] and SAFE [12]) and transcriptomics (e.g., Visium [13], Xenium [10], CosMX [7], GeoMX [16], MERFISH [14]), has enabled high-resolution characterization of tissue MEs at

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

both cellular and subcellular levels. These technologies generate detailed spatial coordinates and expression profiles for individual cells, which can be used to annotate cell types and study cell-cell interactions. However, capturing biologically meaningful and disease-relevant spatial structures from those data remains a significant challenge. Existing methods rely on unsupervised clustering combined with differential expression marker analysis, which requires expert interpretation—an approach that is both time-consuming and costly [3, 9, 17].

With the explosive growth of spatial omics data, traditional clustering methods followed by manual annotation becomes increasingly infeasible due to the extreme time and difficulty involved. It is essential to develop models that can capture spatial tissue architectures across slices/samples and automatically assign microenvironment (ME) labels to new slices with minimal prior annotation information. Several previous studies [17, 9] have attempted to tackle cross-slice annotation through unsupervised spatial domain co-clustering to achieve consistent annotations across multiple slices. While these methods address consistency and scalability issues in multi-slice applications, they still suffer from the shortcomings of unsupervised clustering: they cannot utilize previous annotations and require manual annotation to translate clustering into biologically meaningful categories.

To bridge the gap between the growing demand for cross-slice annotation and the lack of efficient methods, we propose a new framework, MARVEL (Microenvironment Annotation by supervised gRaph contrastiVE Learning), which facilitates label transfer and querying efficiently and flexibly in various scenarios. Inspired by the microenvironment modeling approach from SPACE-GM [18], we leverage cell-neighbor graphs to represent the MEs of each cell. Here, an ME is defined as neighboring cells surrounding a centric cell at various radii, which forms a spatial neighbor graph.

MARVEL is a supervised graph contrastive learning method, i.e., a combination of supervised contrastive learning [11] with graph contrastive learning [21, 20, 5] in a semi-supervised setup. This advanced framework can embed the input cell-neighbor graph into a continuous and representative space while maintaining ME-related information by making full use of limited annotated samples. Meanwhile, MARVEL leverages partially annotated samples as strong positives, improving the quality of representations and mitigating the negative effects of false positives often seen in standard graph contrastive learning. Unlike previous methods focused on extracting aggregated ME embeddings to perform slice-level predictions [3, 9], we propose concentrating on a fine-grained cellular level. The main difference between MARVEL and previous graph-based ME models, e.g., SPACE-GM, lies in their target applications: MARVEL defines each cell's ME through its cell-neighbor graph and reframes the ME annotation task as a graph classification problem for these cell-neighbor graphs, addressing the cell-level task. Meanwhile, SPACE-GM conducts slice-level predictions, treating ME embedding as an intermediate representation to achieve a final slice embedding.

MARVEL is the first method capable of semi-supervised representation learning for microenvironments in spatial omics data, achieving efficient and precise

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

label transfer for inter- and intra-slice microenvironment annotation with limited label information. Specifically, MARVEL requires only a small amount (0.05 %) of partially annotated ME data from some slices, along with the remaining unannotated slices, as input. It can transfer these limited annotations to both the partially annotated and unannotated slices, demonstrating its transductive label transfer capability (around 80% F1 score). Additionally, thanks to the high quality of learned representations, MARVEL can be used for inductive label annotation and novel ME discovery without retraining, unlike traditional label transfer methods that require co-training on both labeled and unlabeled data. More specifically, in an inductive setting, the microenvironment (ME) to be annotated can be directly assigned a label by querying the pretrained reference representations, while in a traditional transductive setting, the ME to be annotated must be co-trained with the annotated MEs in a semi-supervised manner. Extensive benchmarking tests demonstrate that MARVEL outperforms baseline techniques across three key ME-related tasks: transductive ME annotation, inductive ME query, and novel ME discovery across multiple datasets in all evaluation metrics, indicating the generalization and application potential of MARVEL.

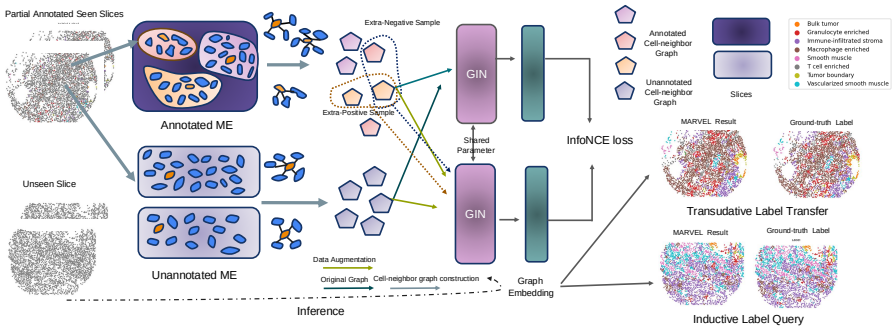

Fig. 1: Proposed pipeline for MARVEL: The input to MARVEL includes a small proportion of annotated microenvironments (for each centric cell) and the remaining unlabeled microenvironments from partial annotated seen slices as training slices. These MEs are represented using a 2-hop cell-neighbor graph. The ME representation is constructed through supervised graph contrastive learning with InfoNCE loss. The orange cell and its neighboring cells illustrate an example cell-neighbor graph. Different background colors in the annotated MEs represent corresponding labels. The learned GIN encoder can be utilized to both transducitive label transfer on partial annotated seen slices and inductive label query on the unseen slices.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2 Method

The input to MARVEL consists of multiple cell-neighbor graphs derived from N input spatial omics slices, collectively denoted as $G_{total} = [G_{total_1}, \ldots, G_{total_k}]$, where $k \in \{1, \ldots, N\}$. Each slice contains multiple cells represented as nodes in a graph, with a total of M cells in whole slices. Each spatial graph captures the spatial neighborhood relationships between cells in a slice using an adjacency matrix. We denote the spatial neighbor graph by an adjacency matrix $A_k \in \mathbb{R}^{n \times n}$, where $A_{kij} = 1$ indicates that cell i and cell j are adjacent in slice k, while $A_{kij} = 0$ indicates they are not neighbors.

For each cell i within a spatial graph G_{total_k}, where $i \in \{1, \ldots, M\}$, we first extract a 2-hop cell-neighbor graph as the input to MARVEL. The adjacency matrix of the 2-hop neighbor graph is defined as follows:

$$A_{k2} = A_k + A_k^2.$$

The node features are represented by the matrix $X_k \in \mathbb{R}^{n \times d}$, where d is the number of features associated with each cell. Categorical features, such as cell types, can be encoded as a one-hot embedding matrix, while numerical features, such as profiled molecular expressions, are stored as is. The constructed graph representation for each 2-hop neighbor subgraph centered on cell i and its index on corresponding slice k as $k(i)$ is expressed as:

$$G_i = (V_i, A_{k2,k(i)}, X_{k(i)})$$

where V_i denotes the set of nodes in the subgraph centered on cell i with node number $|V_i| = n_i$, $A_{k2,k(i)}$ is the 2-hop adjacency matrix for the subgraph, and $X_{k(i)}$ is the corresponding node feature matrix for the subgraph. The final input dataset for MARVEL consists of the combined graph set $\{G_1, G_2, \ldots, G_i, \ldots, G_M\}$.

The GNN encoder processes the cell-neighbor graphs for input cells. For cell-neighbor graph G_i, we input it into the GNN encoder to produce the node embeddings z_j. This is formalized as follows:

$$\mathbf{Z_i} = \mathrm{GNN}(G_i, X_i), \quad \text{where} \quad \mathbf{Z_i} = (z_1, z_2, \ldots, z_{n_i}), \quad z_j \in \mathbb{R}^d.$$

Given the updated node embeddings z_j in G_i, we then compute the microenvironment (ME) embedding $\mathbf{h}_i$ for the i-th cell-neighbor graph by applying mean pooling over the node embeddings. This aggregates the individual node-level representations into a single microenvironment embedding as follows:

$$\mathbf{h}_i = \frac{1}{|V_i|} \sum_{j \in V_i} z_j,$$

where $|V_i|$ denotes the number of nodes in the graph, and $\mathbf{h}_i \in \mathbb{R}^d$ represents the ME embedding for the i-th cell incorporating its neighbor information. The mean pooling operation computes the average of all node embeddings to generate $\mathbf{h}_i$, capturing the characteristics of the entire microenvironment.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MARVEL is designed to learn a robust and representative cell-neighbor graph latent representation. Thus, we apply the InfoNCE [1, 6] loss to the ME embedding $\mathbf{h}_i$. This loss is designed to maximize the similarity between representations of positive pairs (such as augmented versions of the same ME) while minimizing the similarity between negative pairs, which has shown good performance in previous Graph Contrastive Learning work[21, 20].

The InfoNCE loss is defined as follows:

$$\mathcal{L}_{\text{InfoNCE}} = -\frac{1}{|\mathcal{I}|} \sum_{G_i \in \mathcal{I}} \left(\sum_{G_{aug} \in \mathcal{P}(G_i)} \log \frac{\exp\left(\frac{\text{sim}(\mathbf{h}_{G_i}, \mathbf{h}_{G_{aug}})}{\tau}\right)}{\sum_{G_a \in \mathcal{A}(G_i)} \exp\left(\frac{\text{sim}(\mathbf{h}_{G_i}, \mathbf{h}_{G_a})}{\tau}\right)} \right),$$

where $\mathcal{I}$ is the set of all ME graphs in the batch, the temperature parameter τ is a scaling factor that controls the sharpness of the similarity distribution, $\mathcal{P}(G_i)$ represents the set of positive ME graph for ME graph G_i (augmented versions), while $\mathcal{A}(G_i)$ denotes the input original ME without data augmentation in the batch, which may include both same-category and differet category samples. The ME representation for G_i is denoted as $\mathbf{h}_{G_i}$, and $\mathbf{h}_{G_{aug}}$ is the representation for the augmented microenvironment graph. The similarity between two ME representations is measured using a similarity function, such as cosine similarity:

$$\text{sim}(\mathbf{h}_{G_i}, \mathbf{h}_{G_{aug}}) = \frac{\mathbf{h}_{G_i} \cdot \mathbf{h}_{G_{aug}}}{\|\mathbf{h}_{G_i}\| \|\mathbf{h}_{G_{aug}}\|}.$$

In this InfoNCE loss function, the augmented version G_{aug} is treated as a positive sample, while the remaining samples within the same batch are treated as negative samples. Furthermore, we apply a supervised contrastive loss[11] to the ME representations $\mathbf{h}_G$ for the labeled ME graph. This loss is designed to encourage the ME representations with the same label (referred to as positive pairs) to be more similar, while making the representations of microenvironments graph with different labels (negative pairs) more dissimilar. The supervised contrastive loss is defined as follows:

$$\mathcal{L}_{\text{SCL}} = \sum_{G_i \in \mathcal{I}_{\text{labeled}}} \frac{1}{|\mathcal{P}(G_i)|} \sum_{G_p \in \mathcal{P}(G_i)} -\log \frac{\exp\left(\frac{\text{sim}(\mathbf{h}_{G_i}, \mathbf{h}_{G_p})}{\tau}\right)}{\sum_{G_a \in \mathcal{A}(G_i)} \exp\left(\frac{\text{sim}(\mathbf{h}_{G_i}, \mathbf{h}_{G_a})}{\tau}\right)},$$

where $\mathcal{I}_{\text{labeled}}$ is the set of all labeled ME graphs in the batch, $\mathbf{h}_{G_p}$ and $\mathbf{h}_{G_a}$ are the representations for positive and input original MEs, respectively.

The final loss function for MARVEL combines both the InfoNCE loss and the supervised contrastive loss:

$$\mathcal{L}_{\text{SSGCL}} = \mathcal{L}_{\text{InfoNCE}} + \mathcal{L}_{\text{SCL}}.$$

This formulation allows the model to effectively utilize both unlabeled data through augmentations and labeled instances, thereby improving the overall robustness and generalization of the ME graph representations learned.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3 Experiment

3.1 Dataset preparation, preprocessing and Implementation details

We collected three publicly available large-scale CODEX datasets with pathologist-annotated ground truth MEs to evaluate our proposed method against existing approaches. The first dataset [4] consists of multiplexed immunofluorescence imaging data from 140 tissue regions collected from colorectal cancer (CRC) patients, yielding 140 slices. The second dataset is collected from a TNBC (triple-negative breast cancer) patient and contains 12 slices [15], while the third dataset is from a melanoma patient, comprising 24 slices [8]. For our experiments, we utilize half of the slices from each dataset for training, applicable to both transductive and inductive settings, and reserve the remaining slices for evaluation in the inductive query setting. Specifically, in the transductive learning setting, both labeled and unlabeled instances are trained together; however, label information is only available for the labeled instances, which is often referred to as 'label transfer.' In the inductive setting, the annotation of the unlabeled data is directly assigned by the trained model without any co-training procedure.

In all datasets, we stick to the same backbone GNN architecture, the Graph Isomorphism Network (GIN) [19], with identical hyperparameters across experiments. For the transductive label transfer task, we use label proportions of [0.01, 0.05, 0.1, 0.2, 0.3, 0.4] as the seed ME, and evaluate performance using both the cell type and panel expression value-based feature versions of MARVEL. The data augmentation techniques applied during contrastive learning include edge dropout and node feature masking, using a fixed ratio.

In the inductive ME label query and novel ME label identification tasks, we set the default label proportion parameter to 0.1 to construct the reference atlas. Due to the absence of prior methods tailored for these tasks, we use a conventional classifier with the same input features as a baseline. Specifically, we employ cell-type one-hot encoding along with the panel profile values of the central cell, the mean panel expression values, and the cell proportions of neighboring cells as input features for Random Forest and logistic regression models. To ensure a fair comparison, baseline methods and MARVEL use the same set of 2-hop neighboring cells as input.

To ensure the fair benchmarking comparison and remove the variance introduced by model architecture, we apply the same GIN architecture, with supervised classification as the primary training objective. In the inductive ME label query and novel ME label identification tasks, we use reference embeddings based on neighbor cell proportions and panel mean values as the baseline.

3.2 Transudtive ME label transfer

Annotating MEs on reference spatial omics datasets often requires significant time and effort from pathologists. As a result, transferring labels from a small subset of annotated cells to unlabeled ones within a slide, also known as transductive label transfer, is a practical and efficient strategy. However, no existing

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

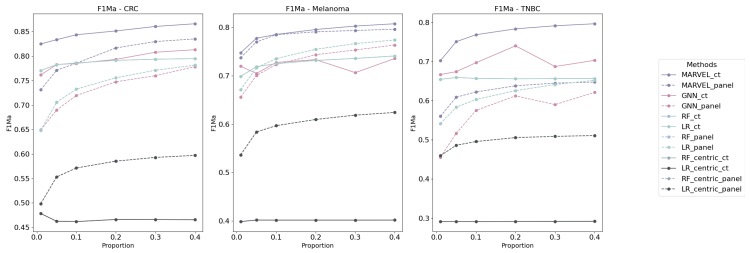

Fig. 2: Transductive label transfer results (F1-Macro) across multiple baselines for different label proportions [0.01, 0.05, 0.1, 0.2, 0.3, 0.4] on three different datasets.

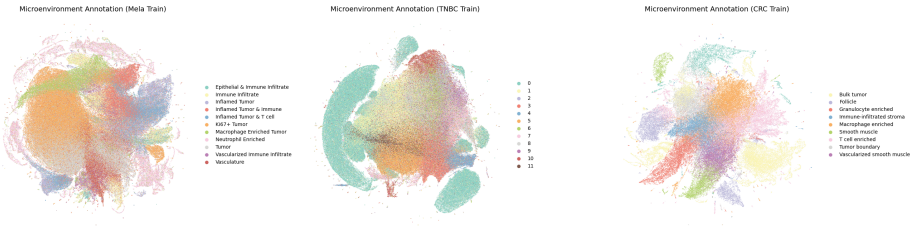

Fig. 3: UMAP visualization of the transductive label transfer learned representation with 0.1 labeled proportion.

ME annotation method has been designed for this transductive learning task. To tackles this challenge, we propose a graph-supervised contrastive learning framework. We first employ graph-supervised contrastive learning, followed by linear probing to obtain classification results. Specifically, we train a logistic regression model using the labeled embeddings and perform inference on the unseen samples. To evaluate performance, we choose the F1 score (macro F1 score) to address class imbalances and assign equal weights to each cell type in the overall score. The results in Fig 2 demonstrate the superior performance of MARVEL over other benchmark methods.

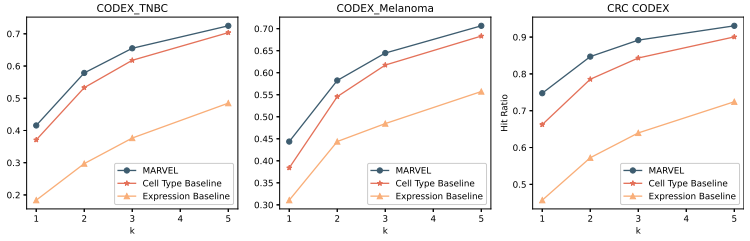

Fig. 4: Inductive label query benchmarking results against various baselines across different values of K in Hit Ratio@K on three different datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

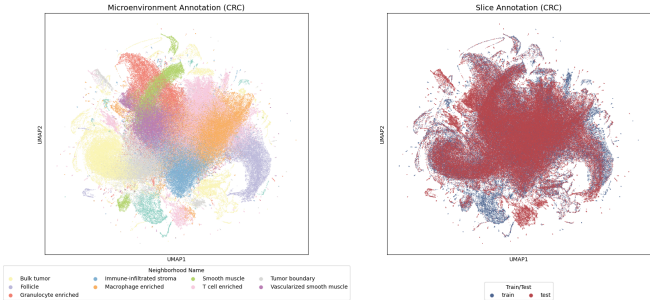

Fig. 5: UMAP visualization of the learned representation with 0.1 labeled proportion for both reference and query microenvironments

3.3 Inductive ME label query

When dealing with large-scale labeled datasets and an increasing number of new datasets, the transductive method becomes costly. This is because handling a rapidly updated query dataset requires the user to re-train the model each time a new query dataset is introduced. To address this issue, a direct *Inductive Label Query* method, which learns a general classification function on the pre-constructed embedding reference dataset, can alleviate the computational cost.

Formally, given a pre-constructed dataset X_{train}, with a new input query dataset X_{test}, the label for each sample in X_{test} is assigned based on the cosine similarity to the existing embeddings in X_{train}.

The cosine similarity between a query sample x^i_{test} and a reference sample x^j_{train} is computed as follows:

$$\text{Cosine Similarity}(x^i_{\text{test}}, x^j_{\text{train}}) = \frac{x^i_{\text{test}} \cdot x^j_{\text{train}}}{\|x^i_{\text{test}}\| \|x^j_{\text{train}}\|}.$$

The label for the query x^i_{test} is assigned to the label of the nearest sample in the reference dataset X_{train}. This inductive method reduces the computational cost by avoiding repeated training by using the pre-constructed reference embeddings.

For evaluation, we calculate the Top-K Hit Ratio, which evaluates whether the true label of a query sample is included among the top K most similar samples in the reference dataset. The Top-K Hit Ratio is defined as:

$$\text{Hit Ratio@K} = \frac{1}{|X_{\text{test}}|} \sum_{i=1}^{|X_{\text{test}}|} \mathbf{I}\left(y^i_{\text{test}} \in \{\hat{y}^i_1, \hat{y}^i_2, \ldots, \hat{y}^i_K\}\right),$$

where y^i_{test} is the true label of the i-th test sample, and $\hat{y}^i_1, \hat{y}^i_2, \ldots, \hat{y}^i_K$ represent the predicted labels of the top K most similar samples from the training dataset. The indicator function $\mathbf{I}(\cdot)$ returns 1 if the true label is within the top K predictions, and 0 otherwise. Higher Hit Ratio@K indicates higher capability of the

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

given embedding methods. We train MARVEL on three datasets on half of slices with the other half slices as the unseen query slices. MARVEL achieves the best Top-K Hit Ratio across different values of K when compared to two baselines, demonstrating its potential to construct the ME reference atlas effectively.

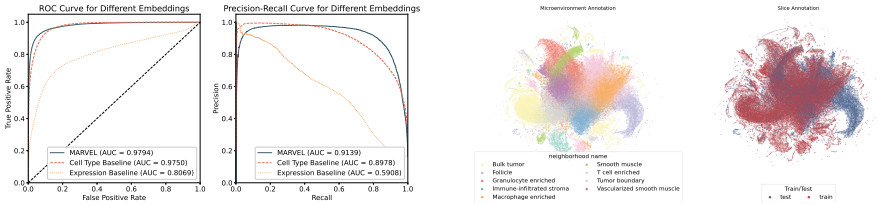

Fig. 6: Novel microenvironment detection.Left, AUC-ROC and AUPRC benchmarking results against baselines for novel microenvironment identification in the CODEX CRC dataset. Right, Visualization of the learned ME representation for both query and reference datasets, where the query dataset contains a unique novel ME (Follicle) identified in the CODEX CRC dataset.

3.4 Novel ME identification

In real-world applications, newly profiled samples may contain novel microenvironments (MEs) that are absent from the reference atlas. Consequently, a robust method should have the capability to identify these previously unseen MEs. However, this task has also not been addressed before. MARVEL is the first methods to address this task. Intuitively, those novel MEs should show minimal similar to any samples in the reference atlas. Therefore, we propose to the novelty score $N(x^i_{\text{test}})$ for each test sample x^i_{test} as

$$N(x^i_{\text{test}}) = 1 - \text{Cosine Similarity}(x^i_{\text{test}}, x^{\hat{j}}_{\text{train}}),$$

where $x^{\hat{j}}_{\text{train}}$ is the sample in the reference dataset most similar to x^i_{test}. We define the ground-truth label y^i_{novel} for each test sample as 1 if the sample belongs to a novel microenvironment (ME) and 0 otherwise.

To simulate this scenario, we separate the CODEX dataset based on the presence of the Follicle Tumor ME. Slices lacking the Follicle ME are designated as reference slices and form the training dataset X_{train}, while slices containing Follicle ME are used as test slices, forming X_{test}. The Follicle ME is annotated as the novel ME. We choose the Follicle ME because it is naturally absent in some slices due to the CRC subtype differences in the original dataset, making this novel ME biologically meaningful.

We evaluate the performance of our method using AUPRC and AUC-ROC to assess how well the novelty scores N can distinguish between novel and non-novel MEs. Compared with baseline methods, MARVEL achieves the best performance as shown in Fig 6a. These results first demonstrate the feasibility of our proposed

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

new task (novel ME identification) in real biological applications. Additionally, they provide supporting evidence for MARVEL's ME embedding, as MARVEL must both capture the information of seen ME categories and maintain generalization to unseen ME categories. This highlights MARVEL's potential in complex real-world open set application scenarios.

4 Conclusion and discussion

In the current spatial omics research domain, most works focus on unsupervised spatial domain clustering within individual slices to annotate microenvironments. However, these approaches often struggle to extend to multi-slice settings, limiting their ability to efficiently leverage existing labels for robust label transfer across large-scale datasets. Our proposed method, MARVEL, aided by the graph formalization of microenvironments and our proposed graph-level supervised contrastive learning, achieves efficient and robust construction of reference microenvironments with accurate annotations.

5 Code Availability

MARVEL source code is available at https://github.com/C0nc/MARVEL

References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning (ICML). vol. 119, pp. 1597–1607. PMLR (2020), http://proceedings.mlr.press/v119/chen20j.html
2. Ding, J., Li, L., Lu, Q., Venegas, J., Wang, Y., Wu, L., Jin, W., Wen, H., Liu, R., Tang, W., et al.: Spatialctd: A large-scale tumor microenvironment spatial transcriptomic dataset to evaluate cell type deconvolution for immuno-oncology. Journal of Computational Biology **31**(9), 871–885 (2024)
3. Dong, K., Zhang, S.: Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nature Communications **13**(1), 1739 (2022). https://doi.org/10.1038/s41467-022-29439-6
4. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Nolan, G.P.: Deep profiling of mouse splenic architecture with codex multiplexed imaging. Cell **174**, 968–981 (2018). https://doi.org/10.1016/j.cell.2018.07.010
5. Hassani, K., Khasahmadi, A.H.: Contrastive multi-view representation learning on graphs. In: Proceedings of the 37th International Conference on Machine Learning, ICML 2020. vol. 119, pp. 4116–4126. PMLR (2020), https://proceedings.mlr.press/v119/hassani20a.html
6. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 9729–9738 (2020). https://doi.org/10.1109/CVPR42600.2020.00973

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7. He, S., Bhatt, R., Beechem, J.M., et al.: High-plex imaging of rna and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nature Biotechnology **40**, 1374–1381 (2022). https://doi.org/10.1038/s41587-022-01483-z
8. Hickey, J.W., Haist, M., Horowitz, N., Caraccio, C., Tan, Y., Rech, A.J., Baertsch, M.A., Rovira-Clavé, X., Zhu, B., Vazquez, G., Barlow, G., Agmon, E., Goltsev, Y., Sunwoo, J.B., Covert, M., Nolan, G.P.: T cell-mediated curation and restructuring of tumor tissue coordinates an effective immune response. Cell Reports **42**, 113494 (2023). https://doi.org/10.1016/j.celrep.2023.113494
9. Hu, Y., Rong, J., Xie, R., Xu, Y., Peng, J., Gao, L., Tan, K.: Learning predictive models of tissue cellular neighborhoods from cell phenotypes with graph pooling. bioRxiv (2022). https://doi.org/10.1101/2022.11.06.515344
10. Janesick, A., Shelansky, R., Gottscho, A.D., Wagner, F., Williams, S.R., Rouault, M., Beliakoff, G., Morrison, C.A., Oliveira, M.F., Sicherman, J.T., Kohlway, A., Abousoud, J., Drennon, T.Y., Mohabbat, S.H., Taylor, S.E.B., 10x Development Teams: High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. Nature Communications **14**, 1–15 (2023). https://doi.org/10.1038/s41467-023-43458-x
11. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. In: Advances in Neural Information Processing Systems. vol. 33, pp. 18661–18673 (2020), https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html
12. Ko, J., Wilkovitsch, M., Oh, J., Kohler, R.H., Bolli, E., Pittet, M.J., Vinegoni, C., Sykes, D.B., Mikula, H., Weissleder, R., et al.: Spatiotemporal multiplexed immunofluorescence imaging of living cells and tissues with bioorthogonal cycling of fluorescent probes. Nature biotechnology **40**(11), 1654–1662 (2022)
13. Lewis, S.M., Asselin-Labat, M.L., Nguyen, Q., Berthelet, J., Tan, X., Wimmer, V.C., Merino, D., Rogers, K.L., Naik, S.H.: Spatial omics and multiplexed imaging to explore cancer biology. Nature Methods **18**, 997–1012 (2021). https://doi.org/10.1038/s41592-021-01203-6
14. Moffitt, J.R., Hao, J., Wang, G., Chen, K., Babcock, H.P., Zhuang, X.: High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proceedings of the National Academy of Sciences (PNAS) **113**, 11046–11051 (2016). https://doi.org/10.1073/pnas.1612826113
15. Shiao, S.L., Gouin, K.H., Ing, N., Ho, A., Basho, R., et al.: Single-cell and spatial profiling identify three response trajectories to pembrolizumab and radiation therapy in triple negative breast cancer. Cancer Cell **42**, 70–84.e8 (2024). https://doi.org/10.1016/j.ccell.2023.12.012
16. Sullivan, S., Lawson, M., Stover, J., et al.: Multiplex digital spatial profiling of proteins and rna in fixed tissue. Cancer Research **82**, 1001–1009 (2022). https://doi.org/10.1158/0008-5472.CAN-21-2648
17. Varrone, M., Tavernari, D., Santamaria-Martínez, A., Walsh, L.A., Ciriello, G.: Cellcharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. bioRxiv (2023). https://doi.org/10.1101/2023.01.10.523386
18. Wu, Z., Trevino, A.E., Wu, E., Swanson, K., Kim, H.J., D'Angio, H.B., Preska, R., Charville, G.W., Dalerba, P., Egloff, A.M., Uppaluri, R., Duvvuri, U., Mayer, A.T., Zou, J.: Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. Nature Biomedical Engineering **6**, 1435–1448 (2022). https://doi.org/10.1038/s41551-022-00951-w

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.29.620828; this version posted November 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019), https://arxiv.org/abs/1810.00826
20. You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., Shen, Y.: Graph contrastive learning with augmentations. In: Advances in Neural Information Processing Systems. vol. 33, pp. 5812–5823 (2020), https://proceedings.neurips.cc/paper/2020/hash/3fe230348e9a12c13120749e3f9fa4cd-Abstract.html
21. Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., Wang, L.: Deep graph contrastive representation learning. arXiv:2006.04131 [cs, stat] (July 2020), http://arxiv.org/abs/2006.04131, arXiv: 2006.04131