This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Bulk RNA-seq Guided Multi-modal Detection of Anomalous Regions in Human Cancer via Spatial Transcriptomics

Hang Shi[1,2], Ruocheng Yang[1,2], Wenjie You[1,2], Zhilin Huang[1,2], Daoqiang Zhang[1,2], Wei Shao[1,2*]
[1]The College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics
[2]The Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education
shaowei20022005@nuaa.edu.cn

Abstract

Spatial transcriptomics (ST) has emerged as a revolutionary approach in the field of tissue analysis that can offer spatial resolved molecular insights for the identification of anomalous regions (AR) on human cancers. Current ST-based methods for detecting AR focus narrowly on the molecular features of local tissue spots, overlooking the matched bulk RNA-seq data that contains crucial diagnostic information. This oversight limits their effectiveness in identifying subtle or heterogeneous tumors, where accurate detection depends on broader genetic context. Besides the genomic signatures, the pathological images can also provide rich visual information to reflect the morphology of AR. To utilize the patient-level diagnostic knowledge and harness complementary information from both histology images and ST, we develop a Bulk RNA-seq Guided Multi-modal Anomalous Regions Detection method (BRGMAR) for the identification of AR from human tissues. Specifically, to effectively model the dependencies in ST, we introduce a Dynamic Multi-Relational Graph Learning (DMRGL) module to adaptively capture complex relationships in ST, including both spatial proximity and gene expression similarity. Then, we design an Optimal Transportation-based Gene Module Alignment (OTGMA) approach to align ST data with patient-level bulk RNA-seq data by matching the compositional and functional similarities of their corresponding gene modules. Finally, we combine the learned genomic features with pathological image representations for accurate AR detection. We evaluate our method on three public available ST datasets for the purpose of identifying cancerous regions from normal tissues, and the experimental results demonstrate the advantage of our method in comparison with the existing studies.

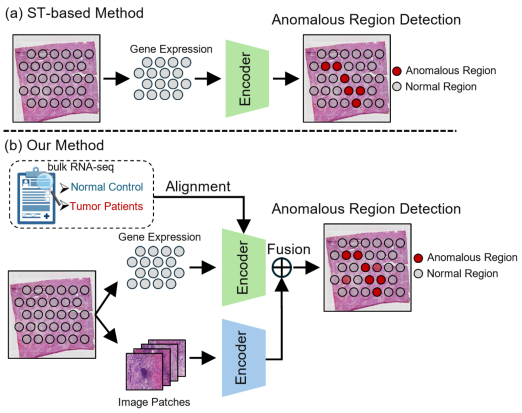


Figure 1. (a) Most of the existing studies rely on gene expression profiles derived from ST for anomalous region detection. (b) Our BRGMAR fuses ST and pathological images, as well as leveraging the corresponding bulk RNA-seq with diagnostic information for performance improvement.

1. Introduction

Anomalous regions (AR) in human tissues refer to spatially confined areas that deviate from normal histological or molecular patterns, and are closely associated with the onset and progression of various diseases particularly for human cancer [1, 5, 20]. In the study of tumor biology, AR correspond to malignant transformations, such as carcinoma in situ or dysplastic regions, which often precede the development of invasive tumors [58, 60]. The accurate detection of these regions is critical not only for early cancer detection and risk stratification, but also for tracing clonal evolution and pinpointing potential sites of recurrence or therapeutic resistance [16, 47].

Early studies on AR detection of human tissues primarily rely on H&E-stained pathological images due to their capacity to provide high-resolution visualizations of tissue architecture and cellular morphology [10, 61]. In these approaches, AR are typically identified by employing computer vision and machine learning algorithms to detect

*Corresponding author.
Code and Dataset is available at: github.com/shihangjs/BRGMAR

morphological deviations between normal and abnormal tissue regions. For instance, the authors in [61] have presented a training system consisting of a CNN-based feature generator and a one-class classifier for the detection of abnormal tiles from Whole Slide Images (WSI). Dippel *et al.* [10] have introduced a self-supervised learning algorithm to flag anomalous regions in histopathology images without the need for extensive manual annotation. However, these image-based studies often face limitations in accurately detecting subtle anomalies in human cancers, particularly within the context of heterogeneous tumors and complex tissue micro-environments [42, 56].

To overcome the above limitations, the spatial transcriptomics (ST) technique has been widely adopted to detect anomalous regions at molecular level [36, 44]. By preserving the spatial context of gene expression within tissue sections, ST facilitates a more comprehensive characterization of tissue heterogeneity and reveals AR that may not be apparent through tissue morphology. Based on ST, researchers have designed several AR detection algorithms. For example, Hu *et al.* [21] formulate the problem of anomaly detection on ST as comparing spatially localized probability densities of gene expression among different conditions. Xu *et al.* [50] use a deep autoencoder coupled with a masked self-supervised learning mechanism for the identification of AR regions in ST. By considering that both ST and their corresponding visual representations can provide complementary information, other studies including [51] further adds pathological image to the ST data for further performance improvement.

Although much progress has been achieved, the existing ST-based studies for AR detection still face several challenges. Firstly, these methods primarily rely on local molecular features derived from individual tissue spots in ST, without incorporating the broader patient-level bulk RNA-seq data that contain diagnostic information. Intuitively, the combination of the overall tumor genetic profile and diagnostic information (*i.e.,* cancer patients or normal control) can provide more guidance to identify AR. Secondly, for the purpose of aligning the genomic data with different domains or sequencing technologies, most of the existing studies primarily focus on matching individual gene expression values [62]. However, such strategy often overlooks the fact that the functional abnormalities in tissues tend to involve disruptions in gene co-expression patterns rather than individual gene expression changes [41]. Thirdly, most existing graph-based ST models assume adjacent spots share similar profiles and thus construct graphs based on Euclidean distance. This approach fails to capture non-local biological relationships, where distant spots can also exhibit similar molecular patterns. Last but not least, most of the existing studies have demonstrated the advantages of combining genomic signatures and image data for various clinical prediction tasks [28, 31], it can be expected that better results can be achieved if we combine ST and pathological images for AR detection.

Based on the above consideration, we propose a Bulk RNA-seq Guided Multi-modal Anomalous Regions (BRGMAR) detection method for the identification of cancerous regions in human tissues by the combination of ST and pathological images data. BRGMAR aligns the extra bulk RNA-seq data to its corresponding target ST data with the same cancer subtype that can leverage the diagnostic information derived from bulk RNA-seq to enhance the performance for cancerous region detection. We conduct experiments on three ST datasets and the experimental results demonstrate the superiority of our method when comparing with the existing studies. We summarize the main contributions of this study as follows:

1) We propose a novel multi-modal AR detection method by the combination of Spatial Transcriptomics (ST), pathological images and prior bulk RNA-seq data with diagnostic information.

2) For each cancer subtype, we propose an Optimal Transportation-based Gene Module Alignment (OTGMA) approach to align ST data with bulk RNA-seq data by matching the compositional and functional similarities of their corresponding gene modules, allowing for the integration of diagnostic information derived from bulk RNA-seq to enhance AR detection on ST data.

3) We introduce a Dynamic Multi-Relational Graph Learning (DMRGL) module to adaptively capture complex relationships in ST, including both spatial proximity and gene expression similarity.

2. Related Work

2.1. Anomalous Region Detection on Human Tissues

Anomalous Region (AR) detection on human tissues is an important task in personalized medicine aiming at identifying region of interests (ROIs) that deviate from normal biological states [1, 23]. Generally, the existing AR detection approaches on human tissues can be categorized into two folds *i.e.,* image and spatial transcriptomics (ST) based methods. The image based methods are primarily relying on Hematoxylin and Eosin (H&E) stained pathological images that can capture the cellular morphology [10, 61] or tissue architecture [26, 56] difference between AR and normal tissues. However, these image-based studies often struggle to detect subtle anomalies sharing similar visual appearance with normal tissues. To address the above limitations, the spatial transcriptomics (ST) technique has been utilized that can offer molecular-level perspective to identify AR [36]. Specifically, for the purpose of capturing the

spatial correlation among different spots, most of the AR prediction models are built upon the graph neural network (GNN) [32, 43, 50]. By considering that the pathological images can provide complementary visual information to ST for AR detection, recent studies in [51, 52] have combined them together for further performance improvement. However, all the ST-based AR detection studies are relying on spot-level genomic signatures, which overlook using the matched bulk RNA-seq data that can provide valuable patient-level diagnostic information to improve the performance for AR detection.

2.2. Cross-Domain Alignment for Genomic Data

As genomic profiling technologies continue to evolve [8], integrating genomic data across different cohorts or sequencing technologies has become increasingly important that allows researchers to leverage the complementary strengths of each platform to achieve a more comprehensive understanding of disease mechanisms [4, 19]. Specifically, by considering the challenges of acquiring sufficiently labeled survival data for every cancer type, the studies in [18, 55] have leveraged the knowledge from related cancer types by aligning their corresponding gene expression data. Bulk2Space [29] is a deep learning framework-based spatial deconvolution algorithm that can simultaneously disclose the spatial and cellular heterogeneity of bulk RNA-seq data via the alignment of single-cell and spatial transcriptomics references. Moreover, SpaGTL [22] is a ST-based transfer learning framework that can realize tissue alignment via gene-level transformers and spot-level manifold-dominated variational autoencoder. Other studies include [25] have presented a flexible deep transfer learning framework for prioritizing single-cell RNA data in relation to bulk-RNA data via a multi-task learning framework. However, most of the existing approaches primarily focus on matching the expression levels of individual genes, which overlook the fact that the functional abnormalities in tissues are often driven by disruptions in gene co-expression patterns, rather than isolated changes in individual genes [41].

3. Method

We show the flowchart of our proposed bulk RNA-seq guided multi-modal anomalous region detection method (BRGMAR) in Fig. 2, which is consisted of four steps. For the pathological images, the vision-language foundational model CONCH is applied for feature extraction [34]. As to the ST data, we firstly develop a dynamic multi-relational graph learning (DMRGL) module that can capture complex dependencies of spots including the spatial proximity and gene expression similarity. Then, we design an optimal transportation-based gene module alignment approach that can match ST and bulk RNA-seq data, by which we can leverage the diagnostic information in bulk RNA-seq data to help infer AR in ST. Finally, we fuse the representations from the ST and pathological images to facilitate accurate AR detection.

3.1. Data Preprocessing

For the ST data, we firstly exclude spots located outside the tissue region. Then, we follow the approach in [11] that applies the SCANPY package [49] to identify a common set of 5000 highly variable genes (HVGs). Next, for the matched pathological images of each ST slice, we crop them into patches according to the spatial context of the spots and further leverage the vision-language model CONCH [34] for feature extraction. On the other hand, as to the bulk RNA-seq data of each cancer cohort obtained from the The Cancer Genome Atlas (TCGA), we also select the top 5000 HVGs and intersect with the selected genes in ST for alignment. To mitigate the batch effect for both ST and bulk RNA-seq data, the SpaCross [13] algorithm is adopted.

3.2. Dynamic Multi-Relational Graph Learning

Recently, graph learning has emerged as a powerful tool for detecting cancerous regions in spatial transcriptomics (ST) [52]. However, most of the existing methods rely on the spatial neighborhood of spots to build graph that fails to take the gene similarity information into consideration. As a matter of fact, the spots with similar gene expression pattern also show close biological relationship [59]. To overcome the above limitations, we propose a dynamic multi-relational graph learning (DMRGL) module that jointly incorporates spatial proximity and gene expression similarity to model the relationship among different spots.

Mathematically, we represent each slice of ST as $X^s = [x_1^s, x_2^s, \ldots, x_n^s]^\top \in \mathbb{R}^{N_s \times D}$, where x_i^s is the gene expression vector of the i-th spot, and D refers to the number of genes. To capture multiple relational patterns among different spots, a dynamic graph is defined $\mathcal{G} = (\mathcal{V}, \mathcal{E}^{(t)})$, where $\mathcal{V}$ refers to the set of spots and $\mathcal{E}^{(t)}$ denotes the edge connections at stage t. In DMRGL, we initially (*i.e.,* $t = 0$) build the graph according to the spatial proximity of spots by applying the KNN algorithm. Then, by considering that the spots with long-range distance may share the same gene expression patterns, we later (*i.e.,* $t > 0$) integrate gene similarity to refine the graph structure by dynamically selecting k_s spots for each node with the largest cosine similarities in the gene embedding space. Here, we employ the Graph Attention Network (GAT) to learn genomic representations of different spots. The attention mechanism adaptively weights the influence of neighboring nodes as follows:

$$\alpha_{vu} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top[\mathbf{W}\vec{h}_v^s \| \mathbf{W}\vec{h}_u^s]\right)\right)}{\sum_{k\in\mathcal{N}(v)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top[\mathbf{W}\vec{h}_v^s \| \mathbf{W}\vec{h}_k^s]\right)\right)} \tag{1}$$

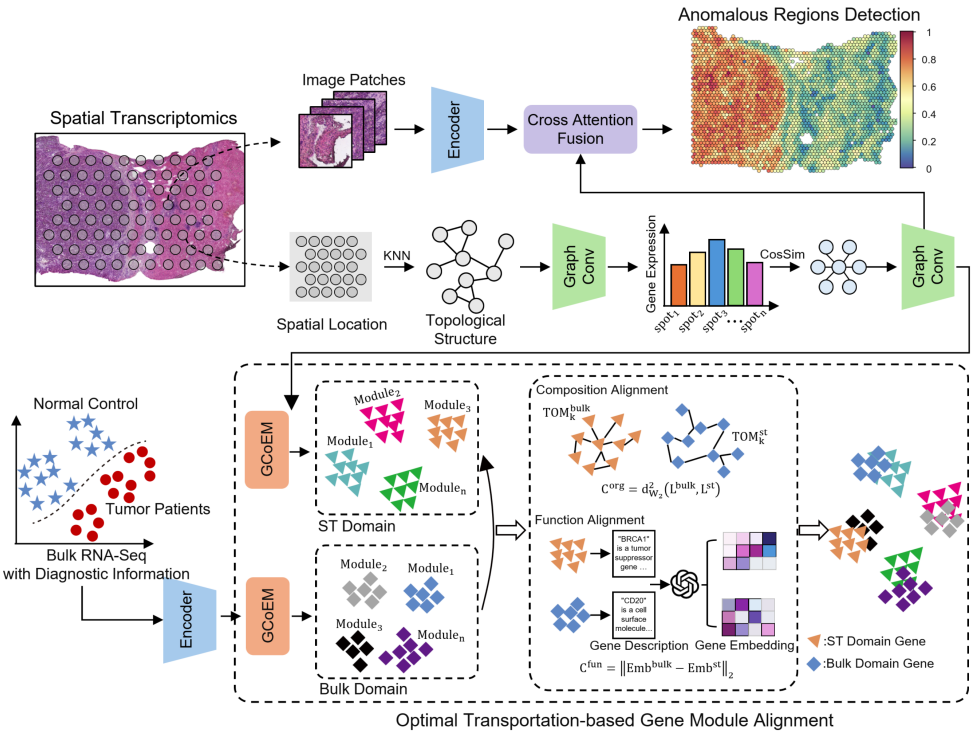


Figure 2. The flowchart of the proposed BRGMAR framework. BRGMAR learns the image representations via the vision-language foundation model CONCH [34]. For the ST data, BRGMAR utilizes a Dynamic Multi-Relational Graph Learning (DMRGL) module to delineate their complex dependencies. In addition, to leverage the patient-level diagnostic information from the bulk-RNA seq, an optimal transportation-based gene module alignment approach is proposed. Finally, we fuse the representations from both ST and pathological images to facilitate accurate AR detection.

where $\vec{h}_v^s$ denotes the node representation of spot v, $\mathbf{W}$ is a learnable weight matrix, $\mathbf{a}$ is the attention vector, and $\|$ represents the concatenation operator. The updated node representation is then computed as:

$$h_v^s = \sigma\Big(\sum_{u\in\mathcal{N}(v)} \alpha_{vu}\mathbf{W}\vec{h}_u^s\Big) \tag{2}$$

3.3. Optimal Transportation-based Gene Module Alignment

It is widely recognized that tissue functional abnormalities often involve disruptions in gene co-expression patterns rather than changes in individual gene expression levels [41]. To transfer diagnostic knowledge from source bulk RNA-seq data to the target ST domain, we propose a novel gene alignment framework by partitioning the shared D genes across ST and bulk RNA-seq data into latent functional modules and then aligning them by jointly considering their compositional and functional characteristics.

Specifically, we first compute domain-specific gene similarity matrices $A^m \in \mathbb{R}^{D\times D}$ from learned features, where the bulk features $Z^b = [z_1^b, z_2^b, \ldots, z_{N_b}^b]^\top \in \mathbb{R}^{N_b\times D}$ are extracted using a multi-layer MLP and the ST features $H^s = [h_1^s, h_2^s, \ldots, h_{N_s}^s]^\top \in \mathbb{R}^{N_s\times D}$ are taken from the previous subsection. Here, $m \in \{b, s\}$ denotes the bulk and ST domains, respectively. Then, given the learnable gene-module assignment matrix $R^m \in \mathbb{R}^{D\times K_m}$, where R_{ik}^m indicates the probability of gene i belonging to module k in domain m, and K_m denotes the number of latent gene modules specific to domain m, we construct a Topological Overlap Matrix (TOM) for each unit to capture its internal co-expression pattern:

$$\mathrm{TOM}_k^m[i,j] = \hat{A}_{ij}^m \cdot R_{ik}^m \cdot R_{jk}^m \tag{3}$$

where $\mathrm{TOM}_k^m[i,j]$ quantifies the co-expression relationship between gene i and gene j in module k, and $\hat{A}_{ij}$ represents the normalized gene similarity matrix. The following sections detail the composition and function alignment based on the TOM representation.

Composition Alignment. To align co-expression modules across the bulk and ST domains, we measure the divergence between their topological representations. Specifically, we treat TOM_k^m as a weighted graph and follow the study in [39] to compute its Laplacian matrix as $L_k^m = D_k^m - \mathrm{TOM}_k^m$, where D_k^m is the degree matrix of TOM_k^m. Then, the compositional dissimilarity between a source module (bulk) and a target module (ST) is quantified using

the squared Wasserstein-2 distance between their respective Laplacian matrices.

$$C_{ij}^{\text{org}} = \text{Tr}\big((L_i^b)^\dagger + (L_j^s)^\dagger\big) - 2\,\text{Tr}\Big(\sqrt{((L_i^b)^\dagger)^{1/2}(L_j^s)^\dagger((L_i^b)^\dagger)^{1/2}}\Big) \tag{4}$$

where $\text{Tr}(\cdot)$ represents the trace operator, and $\dagger$ indicates the pseudoinverse operator.

Function Alignment. Building upon compositional correspondence, we further incorporate functional resemblance to refine module alignment. Following prior work [9], we leverage pre-trained gene functional embeddings from a matrix $E = [e_1, e_2, \ldots, e_D]^\top \in \mathbb{R}^{D\times D_{\text{emb}}}$, where each row e_i corresponds to gene i and is encoded from its NCBI description using OpenAI's *text-embedding-3-large* model. Given the gene module assignment matrix $R^m \in \mathbb{R}^{D\times K_m}$ for domain m, the centroid of the k-th module is computed as:

$$p_k^m = \frac{\sum_{i=1}^{N} R_{ik}^m \cdot e_i}{\sum_{i=1}^{N} R_{ik}^m} \tag{5}$$

Based on p_k^m, we define the functional distance between cross-domain modules as $C_{ij}^{\text{fun}} = d(p_i^{\text{b}}, p_j^{\text{s}})$, where $d(\cdot,\cdot)$ is a distance function (e.g., Euclidean distance). This produces a functional distance matrix $C^{\text{fun}} \in \mathbb{R}^{K_{\text{b}}\times K_{\text{s}}}$, with K_{b} and K_{s} indicating the number of functional gene modules in bulk and ST data, respectively.

Joint Alignment via Optimal Transport. Finally, we formulate the module matching as an optimal transport (OT) problem. An integrated cost matrix is established by combining structural and functional divergences:

$$C_{ij}^{\text{total}} = C_{ij}^{\text{org}} + C_{ij}^{\text{fun}} \tag{6}$$

Our objective is to find an optimal transport plan $P \in \mathbb{R}^{K_{\text{b}}\times K_{\text{s}}}$ that minimizes the overall alignment cost:

$$\begin{aligned} \min_P \quad & \sum_{i=1}^{K_{\text{b}}}\sum_{j=1}^{K_{\text{s}}} P_{ij}\cdot C_{ij}^{\text{total}} \\ \text{s.t.} \quad & \sum_j P_{ij} = a_i, \\ & \sum_i P_{ij} = b_j. \end{aligned} \tag{7}$$

where $a_i = 1/K_{\text{b}}$ and $b_j = 1/K_{\text{s}}$ represent uniform marginal distributions over bulk and ST modules, respectively. To improve numerical stability and optimization efficiency, we introduce an entropic regularization term $H(P) = -\sum_{i,j} P_{ij}(\log P_{ij} - 1)$, resulting in the final Sinkhorn distance as the loss function:

$$\mathcal{L}_{\text{OT}}(C^{\text{total}}) = \sum_{i,j} P_{ij}\cdot C_{ij}^{\text{total}} - \lambda_{\text{reg}}\cdot H(P) \tag{8}$$

where $\lambda_{\text{reg}} = 0.1$ controls the strength of regularization. The optimal transport plan P is efficiently approximated using the Sinkhorn algorithm. The complete algorithmic workflow is provided in Algorithm 1 in the *Supplementary Material*.

3.4. Multi-modal Fusion of ST and Image Data

To integrate representations from both ST and image modalities, we utilize a cross-attention mechanism. For each spot v, the gene expression feature h_v^{s} serves as the query, while the corresponding image feature x_v^{img} provides keys and values.

$$q_v = W_q h_v^{\text{s}}, \quad k_v = W_k x_v^{\text{img}}, \quad v_v = W_v x_v^{\text{img}} \tag{9}$$

where $W_q \in \mathbb{R}^{D\times D}$, $W_k \in \mathbb{R}^{D\times D_{img}}$, and $W_v \in \mathbb{R}^{D\times D_{img}}$ are learnable projection matrices. The cross-attention is computed as:

$$\alpha_v = \text{softmax}\left(\frac{q_v k_v^\top}{\sqrt{D}}\right), \quad \hat{h}_v^s = \alpha_v v_v \tag{10}$$

The attended image context $\hat{h}_v^s$ is then concatenated with the learned gene expression feature and projected to produce the final fused representation:

$$z_v^s = \text{MLP}([h_v^{\text{s}} \| \hat{h}_v^s]) \tag{11}$$

3.5. Anomalous Region Detection

We employ a unified classifier to supervise the shared latent representation space across domains. For each domain $m \in \{b, s\}$, we define learnable class-specific weight vectors organized into a matrix $W^m = [w_1^m, w_2^m, ..., w_{N_{cls}}^m]^\top \in \mathbb{R}^{N_{\text{cls}}\times D}$. Then, the classification probability for sample i in domain m is computed as:

$$P(y_i = c \mid z_i^m) = \frac{\exp((w_c^m)^\top z_i^m)}{\sum_{k=1}^{N_{\text{cls}}} \exp((w_k^m)^\top z_i^m)} \tag{12}$$

Finally, the overall training objective combines cross-entropy losses for both domains with optimal transport:

$$\mathcal{L} = \mathcal{L}_{\text{CE}}^b + \mathcal{L}_{\text{CE}}^s + \mathcal{L}_{\text{OT}} \tag{13}$$

where $\mathcal{L}_{\text{CE}}^m = -\frac{1}{N}\sum_{i=1}^{N} \log P(y_i = y_i^{\text{true}} \mid z_i^m)$.

4. Experimental Results

4.1. Experimental Settings

Datasets: In this study, we utilize three public available spatial transcriptomics (ST) datasets for experimental validation: BRCA[2], ccRCC[35], and HCC[24]. Specifically, the BRCA dataset, derived from breast cancer tissue samples, comprises eight slices (A1 to H1), each containing between 167 and 659 spatial spots. The number

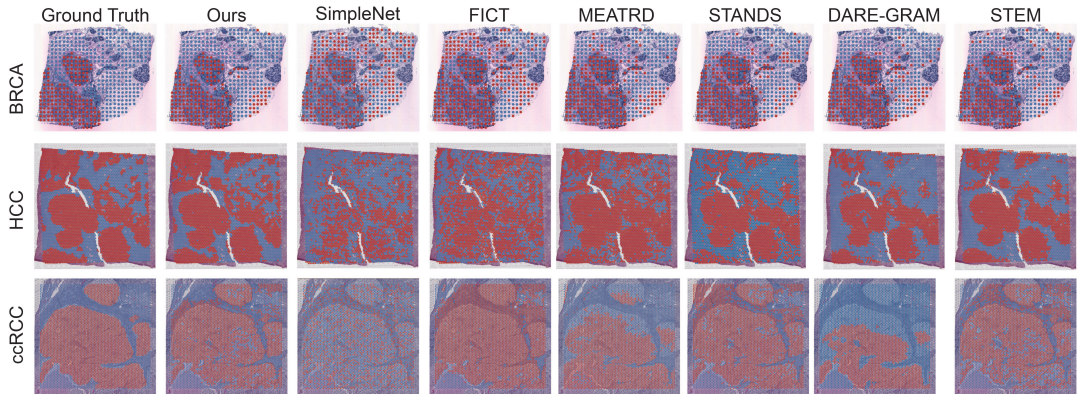


Figure 3. Comparison of the visualization results for anomalous region detection.

of detected genes per slice ranges from 14,861 to 15,671. The ccRCC dataset, obtained from clear cell renal cell carcinoma tissues, consists of eight samples (GSM5924032 to GSM5924041), with each sample comprising 3,585 to 4,975 spatial spots and a uniform gene count of 17,943. The HCC dataset, sourced from hepatocellular carcinoma specimens, includes ten tissue slices (HCC-1L to HCC-5D), with spot counts ranging from 2,755 to 4,750 and gene counts varying between approximately 16,010 and 17,283. Comprehensive statistics for spatial spots including both normal and cancerous spots as well as gene counts across all slices are provided in the *Supplementary Table 5*. For the referenced bulk RNA-seq data of each tissue type, we collect both normal and cancerous samples, which are obtained from The Cancer Genome Atlas (TCGA), a large-scale cancer genomics program supported by the NCI and NHGRI. Since the top 5000 HVGs obtained from bulk RNA-seq and ST data differ (see Section 3.1), we establish a common gene panel for cross-modal alignment by using the intersection of the HVGs identified in both datasets.

Implementation Details: In this paper, we adopt a leave-one-out cross-validation (LOOCV) protocol for performance evaluation across all datasets. Both our BRGMAR and its competitors are trained for 300 epochs using the Adam optimizer, with a learning rate of 1×10^{-4} and a batch size of 1. More specifically, the DMRGL module is implemented as a three-layer architecture with hidden dimensions are set as 256 and 512, respectively. In DMRGL, the spatial structure are formed based on the $k_n = 5$ nearest neighbors of each spot, while gene similarity graph is constructed by identifying the $k_s = 5$ most similar nodes for each spot. As to the OTGMA module, both ST and bulk RNA-seq data are partitioned into $N_m = 10$ functional gene modules. We evaluate the performance of our method by the metrics of the Area Under Curve (AUC) and the F1-score. All experiments are conducted on two NVIDIA RTX 3090 GPU with 24GB memory. We will analysis the time complexity compared with existing methods in *Supplementary Materials Section 7*

Comparing Methods: To thoroughly assess the effectiveness of our proposed framework, we conduct comprehensive benchmarking against a broad spectrum of state-of-the-art methods. These baselines are categorized into four distinct groups: (i) Uni-modal AR detection methods relying on histopathology images, including SimpleNet[33], THOR[6], BIHN[61], AnoDDPM[30]; (ii) Uni-modal AR detection approaches relying on ST data, including FICT[43], Giotto[12], SEDR[50], SpaGIC[32] and BRGMAR-G. Here, BRGMAR-G is a variant of our BRGMAR solely relying on ST data; (iii) Multi-modal AR detection methods by the combination of ST and histopathological image data, including MEATRD[52], STAIG[53], STANDS[51], STGIC[57], stMMC[27]; (iv) Domain alignment methods for transfer learning, including Bulk2Space[29], DARE-GRAM[37], IRAD[54], LOCIT[46], STEM[19]. More detailed introductions of these comparing methods can be found in *Supplementary Materials.*

4.2. Comparison with Uni-modal Methods

We first evaluate the performance of BRGMAR against uni-modal baselines that utilize either histopathology images or gene expression data alone. As shown in Table 1, methods relying solely on histopathology images consistently underperformed. For example, on the BRCA dataset, the best image-based method achieves relatively lower AUCs of 0.601. In contrast, models relying on ST data exhibit substantially better performance, highlighting the advantage of ST that can effectively identify subtle molecular perturbations for AR detection. Notably, our ST-only variant, BRGMAR-G, outperforms all other ST-based baselines. These results underscore the effectiveness of our DMRGL module in capturing complex, multiple spatial dependencies, including non-local biological relationships that are often missed by conventional graph-based or transformer-based approaches.

Table 1. Comparison of BRGMAR with Uni-modal Methods on (a) BRCA, (b) HCC, and (c) ccRCC datasets using average AUC and F1-score as evaluation metrics. The best results are in **bold**, and the second-best results are underlined.

Dataset	Metric	Image-based			ST-based				
		SimpleNet	f-AnoGAN	THOR	BRGMAR-G	SEDR	Giotto	FICT	SpaGIC
BRCA	AUC	0.601 ± 0.094	0.558 ± 0.001	0.512 ± 0.031	$\mathbf{0.726 \pm 0.009}$	0.712 ± 0.046	0.675 ± 0.009	0.674 ± 0.093	0.701 ± 0.082
	F1	0.579 ± 0.005	0.546 ± 0.101	0.557 ± 0.084	$\mathbf{0.734 \pm 0.119}$	0.649 ± 0.078	0.710 ± 0.119	0.585 ± 0.030	0.629 ± 0.037
HCC	AUC	0.586 ± 0.036	0.673 ± 0.032	0.529 ± 0.045	$\mathbf{0.806 \pm 0.033}$	0.773 ± 0.135	0.717 ± 0.033	0.630 ± 0.063	0.633 ± 0.145
	F1	0.517 ± 0.115	0.581 ± 0.083	0.641 ± 0.070	$\mathbf{0.755 \pm 0.045}$	0.654 ± 0.117	0.636 ± 0.045	0.662 ± 0.127	0.752 ± 0.014
ccRCC	AUC	0.536 ± 0.053	0.592 ± 0.128	0.568 ± 0.035	$\mathbf{0.671 \pm 0.050}$	0.660 ± 0.181	0.653 ± 0.055	0.668 ± 0.074	0.642 ± 0.057
	F1	0.540 ± 0.002	0.611 ± 0.048	0.607 ± 0.100	$\mathbf{0.678 \pm 0.001}$	0.622 ± 0.005	0.661 ± 0.199	0.655 ± 0.051	0.647 ± 0.066

4.3. Comparison with Multi-modal Methods

We further compare our BRGMAR with multi-modal methods that combine ST data and pathological image for cancerous region detection. As summarized in Table 2, the experimental results show that BRGMAR consistently and significantly outperforms all competing methods across all the three datasets. For example, on the BRCA dataset, BRGMAR achieves an F1-score of 0.775 and an AUC of 0.801. A similar trend is observed on the HCC dataset, where our framework attains a leading AUC of 0.897 and an F1-score of 0.863, demonstrating a clear performance gain over the next-best approach. We attribute this substantial improvement to the incorporation of a novel transfer learning strategy, which effectively leverages the rich diagnostic signals embedded in patient-level bulk RNA-seq data for cancerous region detection.

4.4. Comparison with Transfer Learning Methods

To comprehensively assess the efficacy of the proposed OTGMA within our BRGMAR framework for aligning ST and bulk RNA-seq data, we conducted the comparison against several SOTA transfer learning approaches. The experimental outcomes, as detailed in Table 2, unequivocally demonstrate that our method achieves superior performance across multiple evaluation metrics. This consistent advantage stems from the inherent limitations of conventional transfer learning techniques, which primarily focus on aligning individual transcriptomic features in isolation. Such approaches often overlook the critical biological insight that functional abnormalities in tissues predominantly emerge from disruptions in gene co-expression networks, rather than from isolated alterations in single-gene expression profiles [14]. In contrast, our OTGMA module innovatively addresses this gap by aligning ST and bulk RNA-seq data at the granular level of gene co-expression modules, leveraging both functional and compositional similarities. This design not only preserves the underlying regulatory architecture but also facilitates a more biologically meaningful transfer of knowledge across data modalities.

4.5. Qualitative Analysis of Anomaly Detection:

To qualitatively assess the anomaly detection performance of our method, we visualize the predicted results across different datasets. As shown in Fig. 3, on the BRCA dataset, our model can accurately delineate anomalous regions, with boundaries closely matching ground truth annotations. Similar results can also be observed for HCC and ccRCC datasets, where our method consistently produces coherent and well-defined contours aligned with tumor regions. By contrast, image-only methods such as SimpleNet often yields blurred or diffuse predictions due to their limited contextual understanding from the genomic perspective, while the gene expression-based approaches like FICT show stronger AR detection ability but may produce fragmented or spatially inconsistent outputs. Moreover, by the combination of image and genomic data, the multi-modal approach including MEATRD, STANDS can offer more advanced cancerous region detection results but still exhibits less precise boundary in comparison with our approach.

4.6. Ablation Study

In Table 3, we conduct the ablation study to verify the effectiveness of different components in our proposed BRGMAR. Here, C^{org} and C^{fun} in OTGMA refer to considering the compositional and functional similarities of gene modules for domain alignment via optimal transportation, while $\mathcal{E}^{sp}$ and $\mathcal{E}^{sim}$ correspond to the incorporation of spatial units or gene expression profile similarity edge in the DMRGL module. From Table 3, we can derive the following three observations. Firstly, capturing both compositional and functional similarities for gene module alignment is very important since it can enhance the biological relevance and accuracy of the alignment between ST and bulk RNA-seq data. Secondly, the DMRGL module that can combine both spatial proximity and gene expression similarity for graph learning can yield more promis-

Table 2. Comparison of BRGMAR with Multi-modal and Transfer Learning Methods on (a) BRCA, (b) HCC, and (c) ccRCC datasets using average AUC and F1-score as evaluation metrics.

Dataset	Metric	BRGMAR	Multi-modal Methods					Transfer Learning Methods				
			MEATRD	STANDS	STAIG	STGIC	stMMC	STEM	DARE-GRAM	LOCIT	Bulk2Space	IRAD
BRCA	AUC	**0.801 ± 0.029**	0.764 ± 0.047	0.723 ± 0.164	0.756 ± 0.028	0.769 ± 0.069	0.737 ± 0.052	0.721 ± 0.017	0.738 ± 0.033	0.770 ± 0.028	0.740 ± 0.030	0.719 ± 0.039
	F1	**0.775 ± 0.046**	0.732 ± 0.079	0.742 ± 0.049	0.738 ± 0.013	0.744 ± 0.023	0.630 ± 0.023	0.742 ± 0.065	0.722 ± 0.084	0.732 ± 0.043	0.735 ± 0.018	0.701 ± 0.115
HCC	AUC	**0.897 ± 0.010**	0.861 ± 0.116	0.854 ± 0.015	0.828 ± 0.087	0.803 ± 0.035	0.748 ± 0.050	0.836 ± 0.036	0.732 ± 0.021	0.827 ± 0.095	0.831 ± 0.111	0.816 ± 0.024
	F1	**0.863 ± 0.037**	0.838 ± 0.004	0.835 ± 0.056	0.803 ± 0.063	0.749 ± 0.064	0.807 ± 0.104	0.811 ± 0.116	0.807 ± 0.104	0.842 ± 0.051	0.824 ± 0.141	0.807 ± 0.001
ccRCC	AUC	**0.724 ± 0.025**	0.681 ± 0.012	0.705 ± 0.143	0.651 ± 0.074	0.648 ± 0.091	0.691 ± 0.051	0.704 ± 0.024	0.685 ± 0.022	0.698 ± 0.122	0.663 ± 0.057	0.697 ± 0.084
	F1	**0.732 ± 0.041**	0.707 ± 0.039	0.669 ± 0.031	0.703 ± 0.009	0.676 ± 0.106	0.706 ± 0.009	0.705 ± 0.018	0.701 ± 0.057	0.684 ± 0.081	0.709 ± 0.007	0.684 ± 0.031

ing AR detection results. Finally, by leveraging dynamic graph construction, BRGMAR discovers latent tissue organization patterns during feature learning, which considerably enhances the model's detection capability for cancerous regions.

Table 3. Ablation study for different components in our proposed BRGMAR

OTGMA		DMRGL		BRCA		HCC		ccRCC	
C^{org}	C^{fun}	$\mathcal{E}^{sp}$	$\mathcal{E}^{sim}$	AUC	F1	AUC	F1	AUC	F1
×	✓	✓	✓	0.767	0.729	0.854	0.833	0.694	0.705
✓	×	✓	✓	0.752	0.743	0.877	0.842	0.690	0.710
×	×	✓	✓	0.748	0.714	0.830	0.819	0.643	0.669
✓	✓	×	✓	0.763	0.752	0.841	0.795	0.721	0.717
✓	✓	✓	×	0.785	0.746	0.860	0.816	0.716	0.723
✓	✓	✓	✓	**0.801**	**0.775**	**0.897**	**0.863**	**0.724**	**0.732**

4.7. Parameter Analysis

In this section, we discuss the effects of the following key parameters in our proposed BRGMAR method: k_n: Number of neighbors used to construct the spatial context graph. k_s: Number of neighbors when building the gene similarity edge. N_m: Number of generated gene modules in OTGMA. N_{hvg}:The number of top highly variable genes.

As shown in Fig. 4 in *Supplementary Materials*, our model is robust to variations in k_n and k_s, and with both evaluation metrics fluctuating within a small ranges. Furthermore, increasing the number of gene modules N_m initially improves performance, as the model captures finer-grained functional patterns for alignment between ST and bulk RNA-seq data. The performance stabilizes when N_m exceeds 10, suggesting that most biologically relevant patterns for AR detection have been adequately captured. For the number of HVGs, we observe that performance improves as the number increases from 3000 to 5000, then plateaus beyond 5000 genes, indicating that selecting around 5000 HVGs is sufficient to capture the key transcriptional features for optimal performance.

4.8. Results for Module Alignment

For each cancer cohort, We first identify the top three bulk-ST gene module pairs that exhibit the highest matching probabilities according to the optimal transport plan (*shown in Supplementary Materials*) derived from the OTGMA module. Then, for each pair, we perform KEGG pathway enrichment analysis and visualize the pathways that are commonly enriched in each paired module. As illustrated in Fig. 6 (*in Supplementary Materials*), for BRCA cohort, the selected bulk–ST pairs underscored core processes such as cellular signaling regulation, immune–stromal communication, and hormone response. These pathways are well-known to drive tumor-growth [17], influence the tumor microenvironment [45], and modulate hormone-dependent oncogenic programs [3] of breast cancer. Additional discussions for other cancer types can be found in the *Supplementary Materials*.

5. Conclusion

In this study, we introduces BRGMAR, a multi-modal approach for detecting anomalous tissue regions by jointly leveraging spatial transcriptomics, bulk RNA-seq, and histology images. Unlike existing methods that focus only on molecular signals in ST, BRGMAR incorporates broader patient-level information and visual features to enhance detection sensitivity. Through dynamic multi-relational graph learning, our model captures multiple relationships among spots in ST, while the proposed OTGMA module enables diagnostic knowledge transfer from bulk-RNA data. We evaluate our method on three public available ST datasets for the purpose of identifying cancerous regions from normal tissues, and the experimental results demonstrate the advantage of our method in comparison with the existing studies. Overall, our study underscores the value of multi-modal integration in spatial genomics and opens avenues for more precise and informative cancer diagnostics.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62136004, 62272226), National Key R&D Program of China (Grant No. 2023YFF1204803).

References

[1] Sana Ahuja and Sufian Zaheer. Advancements in pathology: Digital transformation, precision medicine, and beyond. *Journal of Pathology Informatics*, 16:100408, 2025. 1, 2

[2] Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, et al. Spatial deconvolution of her2-positive breast tumors reveals novel intercellular relationships. *bioRxiv*, pages 2020–07, 2020. 5

[3] Carlos L Arteaga and Jeffrey A Engelman. Erbb receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer cell*, 25(3):282–303, 2014. 8

[4] Hamidreza Ashayeri, Navid Sobhi, Paweł Pławiak, Siamak Pedrammehr, Roohallah Alizadehsani, and Ali Jafarizadeh. Transfer learning in cancer genetics, mutation detection, gene expression analysis, and syndrome recognition. *Cancers*, 16(11):2138, 2024. 3

[5] Cosmin I Bercea, Benedikt Wiestler, Daniel Rueckert, and Julia A Schnabel. Generalizing unsupervised anomaly detection: towards unbiased pathology screening. In *Medical Imaging with Deep Learning*, 2023. 1

[6] Cosmin I Bercea, Benedikt Wiestler, Daniel Rueckert, and Julia A Schnabel. Diffusion models with implicit guidance for medical anomaly detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 211–220. Springer, 2024. 6, 1

[7] Femke H Burgers, Johannes CK van der Mijn, Tom TP Seijkens, Inge Jedema, Axel Bex, and John BAG Haanen. Immunological features of clear-cell renal-cell carcinoma and resistance to immune checkpoint inhibitors. *Nature Reviews Nephrology*, pages 1–15, 2025. 2

[8] Debyani Chakravarty and David B Solit. Clinical cancer genomic profiling. *Nature Reviews Genetics*, 22(8):483–501, 2021. 3

[9] Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024. 5

[10] Jonas Dippel, Niklas Prenißl, Julius Hense, Philipp Liznerski, Tobias Winterhoff, Simon Schallenberg, Marius Kloft, Oliver Buchstab, David Horst, Maximilian Alber, et al. Ai-based anomaly detection for clinical-grade histopathological diagnostics. *Nejm Ai*, 1(11):AIoa2400468, 2024. 1, 2

[11] Kangning Dong and Shihua Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications*, 13(1):1739, 2022. 3

[12] Ruben Dries, Qian Zhu, Rui Dong, Chee-Huat Linus Eng, Huipeng Li, Kan Liu, Yuntian Fu, Tianxiao Zhao, Arpan Sarkar, Feng Bao, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology*, 22(1):78, 2021. 6, 1

[13] Donghai Fang and Wenwen Min. Spacross deciphers spatial structures and corrects batch effects in multi-slice spatially resolved transcriptomics. *Communications Biology*, 8 (1):1393, 2025. 3

[14] Chris Gaiteri, Ying Ding, Beverly French, George C Tseng, and Etienne Sibille. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes, brain and behavior*, 13(1):13–24, 2014. 7

[15] Bocheng Gao, Yan Lu, Xingyue Lai, Xi Xu, Shuhua Gou, Zhida Yang, Yanju Gong, and Hong Yang. Metabolic reprogramming in hepatocellular carcinoma: mechanisms of immune evasion and therapeutic implications. *Frontiers in Immunology*, 16:1592837, 2025. 2

[16] Qiangqiang Gu, Chady Meroueh, Jacob Levernier, Trynda Kroneman, Thomas Flotte, and Steven Hart. Using an anomaly detection approach for the segmentation of colorectal cancer tumors in whole slide images. *Journal of Pathology Informatics*, 14:100336, 2023. 1

[17] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011. 8

[18] Blaise Hanczar, Victoria Bourgeais, and Farida Zehraoui. Assessment of deep learning and transfer learning for cancer prediction based on gene expression data. *BMC Bioinformatics*, 23:262, 2022. 3

[19] Minsheng Hao, Erpai Luo, Yixin Chen, Yanhong Wu, Chen Li, Sijie Chen, Haoxiang Gao, Haiyang Bian, Jin Gu, Lei Wei, et al. Stem enables mapping of single-cell and spatial transcriptomics data with transfer learning. *Communications Biology*, 7(1):56, 2024. 3, 6, 1

[20] Rui Hou, Yifan Peng, Lars J Grimm, Yinhao Ren, Maciej A Mazurowski, Jeffrey R Marks, Lorraine M King, Carlo C Maley, E Shelley Hwang, and Joseph Y Lo. Anomaly detection of calcifications in mammography based on 11,000 negative cases. *IEEE Transactions on Biomedical Engineering*, 69(5):1639–1650, 2021. 1

[21] Gary Hu, Julian Gold, Uthsav Chitra, Sunay Joshi, and Benjamin J Raphael. Anomaly detection in spatial transcriptomics via spatially localized density comparison. *Bioinformatics*, 41(Supplement_1):i493–i501, 2025. 2

[22] Wendong Huang, Yaofeng Hu, Lequn Wang, Guangsheng Wu, Chuanchao Zhang, and Qianqian Shi. Spatially aligned graph transfer learning for characterizing spatial regulatory heterogeneity. *Briefings in Bioinformatics*, 26(1):bbaf021, 2025. 3

[23] Benjamin Jessie, Jingchen Liang, Yinan Li, and Babak Fahimi. High-frequency excitation and surface temperature analysis of breast tissue for detection of anomaly. *BioMed Research International*, 2023(1):4406235, 2023. 2

[24] Ziming Jiang, Yanhong Wu, Yuxin Miao, Kaige Deng, Fan Yang, Shuhuan Xu, Yupeng Wang, Renke You, Lei Zhang, Yuhan Fan, et al. Hccdb v2. 0: Decompose expression variations by single-cell rna-seq and spatial transcriptomics in hcc. *Genomics, Proteomics & Bioinformatics*, 22(1): qzae011, 2024. 5

[25] Travis S Johnson, Christina Y Yu, Zhi Huang, Siwen Xu, Tongxin Wang, Chuanpeng Dong, Wei Shao, Mohammad Abu Zaid, Xiaoqing Huang, Yijie Wang, et al. Diagnostic evidence gauge of single cells (degas): a flexible deep

transfer learning framework for prioritizing cells in relation to disease. *Genome medicine*, 14(1):11, 2022. 3

[26] Yu-Chen Lai and Wei-Ta Chu. Unsupervised anomaly detection on histopathology images using adversarial learning and simulated anomaly. In *Annual Conference on Medical Image Understanding and Analysis*, pages 357–371. Springer, 2024. 2

[27] Bingjun Li, Mostafa Karami, Masum Shah Junayed, and Sheida Nabavi. Multi-modal spatial clustering for spatial transcriptomics utilizing high-resolution histology images. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3469–3474. IEEE Computer Society, 2024. 6, 1

[28] Le Li, Hudan Pan, Yong Liang, Mingwen Shao, Shengli Xie, Shanghui Lu, and Shuilin Liao. Pmfn-ssl: Self-supervised learning-based progressive multimodal fusion network for cancer diagnosis and prognosis. *Knowledge-Based Systems*, 289:111502, 2024. 2

[29] Jie Liao, Jingyang Qian, Yin Fang, Zhuo Chen, Xiang Zhuang, Ningyu Zhang, Xin Shao, Yining Hu, Penghui Yang, Junyun Cheng, et al. De novo analysis of bulk rna-seq data at spatially resolved single-cell resolution. *Nature Communications*, 13(1):6498, 2022. 3, 6, 1

[30] Jasper Linmans, Gabriel Raya, Jeroen van der Laak, and Geert Litjens. Diffusion models for out-of-distribution detection in digital pathology. *Medical Image Analysis*, 93: 103088, 2024. 6, 1

[31] Jana Lipkova, Richard J Chen, Bowen Chen, Ming Y Lu, Matteo Barbieri, Daniel Shao, Anurag J Vaidya, Chengkuan Chen, Luoting Zhuang, Drew FK Williamson, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer cell*, 40(10):1095–1110, 2022. 2

[32] Wei Liu, Bo Wang, Yuting Bai, Xiao Liang, Li Xue, and Jiawei Luo. Spagic: graph-informed clustering in spatial transcriptomics via self-supervised contrastive learning. *Briefings in Bioinformatics*, 25(6):bbae578, 2024. 3, 6, 1

[33] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20402–20411, 2023. 6, 1

[34] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature medicine*, 30(3):863–874, 2024. 3, 4

[35] Maxime Meylan, Florent Petitprez, Etienne Becht, Antoine Bougoüin, Guilhem Pupier, Anne Calvez, Ilenia Giglioli, Virginie Verkarre, Guillaume Lacroix, Johanna Verneau, et al. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity*, 55(3):527–541, 2022. 5

[36] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5):534–546, 2022. 2

[37] Ismail Nejjar, Qin Wang, and Olga Fink. Dare-gram: Unsupervised domain adaptation regression by aligning inverse gram matrices. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11744–11754, 2023. 6, 1

[38] Jana Nekvindova, Alena Mrkvicova, Veronika Zubanova, Alena Hyrslova Vaculova, Pavel Anzenbacher, Pavel Soucek, Lenka Radova, Ondrej Slaby, Igor Kiss, Jan Vondracek, et al. Hepatocellular carcinoma: gene expression profiling and regulation of xenobiotic-metabolizing cytochromes p450. *Biochemical pharmacology*, 177:113912, 2020. 2

[39] Hermina Petric Maretic, Mireille El Gheche, Giovanni Chierchia, and Pascal Frossard. Got: an optimal transport framework for graph comparison. *Advances in Neural Information Processing Systems*, 32, 2019. 4

[40] Arya Mariam Roy, Renuka Iyer, and Sayan Chakraborty. The extracellular matrix in hepatocellular carcinoma: Mechanisms and therapeutic vulnerability. *Cell Reports Medicine*, 4(9), 2023. 2

[41] Ashis Saha, Yungil Kim, Ariel DH Gewirtz, Brian Jo, Chuan Gao, Ian C McDowell, Barbara E Engelhardt, Alexis Battle, François Aguet, Kristin G Ardlie, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome research*, 27(11):1843–1858, 2017. 2, 3, 4

[42] Tom Shenkar and Lior Wolf. Anomaly detection for tabular data with internal contrastive learning. In *International conference on learning representations*, 2022. 2

[43] Haotian Teng, Ye Yuan, and Ziv Bar-Joseph. Clustering spatial transcriptomics data. *Bioinformatics*, 38(4):997–1004, 2022. 3, 6, 1

[44] Luyi Tian, Fei Chen, and Evan Z Macosko. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41 (6):773–782, 2023. 2

[45] Claudia Urueña, Paola Lasso, David Bernal-Estevez, Diego Rubio, Ana Janeth Salazar, Mercedes Olaya, Alfonso Barreto, Mauricio Tawil, Lilian Torregrosa, and Susana Fiorentino. The breast cancer immune microenvironment is modified by neoadjuvant chemotherapy. *Scientific reports*, 12(1):7981, 2022. 8

[46] Vercruyssen Vincent, Meert Wannes, and Davis Jesse. Transfer learning for anomaly detection through localized and unsupervised instance selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6054–6061, 2020. 6, 1

[47] Jue Wang, Yunfang Yu, Yujie Tan, Huan Wan, Nafen Zheng, Zifan He, Luhui Mao, Wei Ren, Kai Chen, Zhen Lin, et al. Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. *Nature Communications*, 15(1):4369, 2024. 1

[48] Yongfeng Wang, Ci Yin, Lele Geng, and Weiyang Cai. Immune infiltration landscape in clear cell renal cell carcinoma implications. *Frontiers in oncology*, 10:491621, 2021. 2

[49] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):15, 2018. 3

[50] Hang Xu, Huazhu Fu, Yahui Long, Kok Siong Ang, Raman Sethi, Kelvin Chong, Mengwei Li, Rom Uddamvathanak, Hong Kai Lee, Jingjing Ling, et al. Unsupervised spatially

embedded deep representation of spatial transcriptomics. *Genome Medicine*, 16(1):12, 2024. 2, 3, 6, 1

[51] Kaichen Xu, Yan Lu, Suyang Hou, Kainan Liu, Yihang Du, Mengqian Huang, Hao Feng, Hao Wu, and Xiaobo Sun. Detecting anomalous anatomic regions in spatial transcriptomics with stands. *Nature Communications*, 15(1):8223, 2024. 2, 3, 6, 1

[52] Kaichen Xu, Qilong Wu, Yan Lu, Yinan Zheng, Wenlin Li, Xingjie Tang, Jun Wang, and Xiaobo Sun. Meatrd: Multimodal anomalous tissue region detection enhanced with spatial transcriptomics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12918–12926, 2025. 3, 6, 1

[53] Yitao Yang, Yang Cui, Xin Zeng, Yubo Zhang, Martin Loza, Sung-Joon Park, and Kenta Nakai. Staig: Spatial transcriptomics analysis via image-aided graph contrastive learning for domain exploration and alignment-free integration. *Nature Communications*, 16(1):1067, 2025. 6, 1

[54] Ziyi Yang, Iman Soltani, and Eric Darve. Anomaly detection with domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2958–2967, 2023. 6, 1

[55] Qingyan Yin, Wangwang Chen, Chunxia Zhang, and Zhi Wei. A convolutional neural network model for survival prediction based on prognosis-related cascaded wx feature selection. *Laboratory Investigation*, 102(10):1064–1074, 2022. 3

[56] Philip Zehnder, Jeffrey Feng, Reina N Fuji, Ruth Sullivan, and Fangyao Hu. Multiscale generative model using regularized skip-connections and perceptual loss for anomaly detection in toxicologic histopathology. *Journal of Pathology Informatics*, 13:100102, 2022. 2

[57] Chen Zhang, Junhui Gao, Hong-Yu Chen, Lingxin Kong, Guangshuo Cao, Xiangyu Guo, Wei Liu, Bin Ren, and Dong-Qing Wei. Stgic: A graph and image convolution-based method for spatial transcriptomic clustering. *PLOS Computational Biology*, 20(2):e1011935, 2024. 6, 1

[58] Shaosen Zhang, Xinyi Xiao, Yonglin Yi, Xinyu Wang, Lingxuan Zhu, Yanrong Shen, Dongxin Lin, and Chen Wu. Tumor initiation and early tumorigenesis: molecular mechanisms and interventional targets. *Signal transduction and targeted therapy*, 9(1):149, 2024. 1

[59] Ran Zhou, Gaoxia Yang, Yan Zhang, and Yuan Wang. Spatial transcriptomics in development and disease. *Molecular Biomedicine*, 4(1):32, 2023. 3

[60] Ran Zhou, Xiwen Tang, and Yuan Wang. Emerging strategies to investigate the biology of early cancer. *Nature Reviews Cancer*, 24(12):850–866, 2024. 1

[61] Igor Zingman, Birgit Stierstorfer, Charlotte Lempp, and Fabian Heinemann. Learning image representations for anomaly detection: application to discovery of histological alterations in drug development. *Medical Image Analysis*, 92:103067, 2024. 1, 2, 6

[62] Yun Zuo, Bangyi Zhang, Wenying He, Yue Bi, Xiangrong Liu, Xiangxiang Zeng, and Zhaohong Deng. Mslocpred: deep transfer learning-based identification of multi-label mrna subcellular localization. *Briefings in Bioinformatics*, 25(6):bbae504, 2024. 2