Enhancing Spatial Transcriptomics Analysis by Integrating Image-Aware Deep Learning Methods

Jiarong Song[1,4], Josh Lamstein[2], Vivek Gopal Ramaswamy[2], Michelle Webb[4], Gabriel Zada[3], Steven Finkbeiner[3], and David W. Craig[1,4,5†]

[1]*Department of Integrated Translational Sciences; City of Hope, Duarte, CA 91010, USA*
[2]*Center for Systems and Therapeutics, Gladstone Institutes, San Francisco, CA 94158, USA*
[3]*USC Radiosurgery Center, Keck School of Medicine of USC, CA 91008, USA*
[4]*Dept of Translational Genomics, Keck School of Medicine of USC, CA 91008, USA*
[5]*Departments of Neurology and Physiology, UCSF, San Francisco CA 91458*
[†]*Email: dacraig@coh.org*

Spatial transcriptomics (ST) represents a pivotal advancement in biomedical research, enabling the transcriptional profiling of cells within their morphological context and providing a pivotal tool for understanding spatial heterogeneity in cancer tissues. However, current analytical approaches, akin to single-cell analysis, largely depend on gene expression, underutilizing the rich morphological information inherent in the tissue. We present a novel method integrating spatial transcriptomics and histopathological image data to better capture biologically meaningful patterns in patient data, focusing on aggressive cancer types such as glioblastoma and triple-negative breast cancer. We used a ResNet-based deep learning model to extract key morphological features from high-resolution whole-slide histology images. Spot-level PCA-reduced vectors of both the ResNet-50 analysis of the histological image and the spatial gene expression data were used in Louvain clustering to enable image-aware feature discovery. Assessment of features from image-aware clustering successfully pinpointed key biological features identified by manual histopathology, such as for regions of fibrosis and necrosis, as well as improved edge definition in EGFR-rich areas. Importantly, our combinatorial approach revealed crucial characteristics seen in histopathology that gene-expression-only analysis had missed.

Supplemental Material:
https://github.com/davcraig75/song_psb2014/blob/main/SupplementaryData.pdf

Keywords: Spatial transcriptomics; Deep learning; Image-aware clustering

1. Introduction

Mapping the spatial organization of genes and cells in tissues is the foundation for understanding higher-level molecular and cellular processes driving disease pathogenesis. In the past decade, paradigm-shifting approaches such as single-cell RNA-seq (scRNA) have provided unprecedented insights into cellular populations. More recently, Spatial Transcriptomics (ST) methods have emerged (e.g., Visium ST), providing a view of cellular RNA expression and disease pathology in ~~the context of neighboring cel~~ls and structures[1,2]. Instead of measuring a single bulk transcriptome from a tissue section, ST obtains thousands of transcriptomes across a tissue section at spatially distinct spots, where each spot covers a few cells with Visium[3], or provides sub-cellular data as with

MERFISH[4] and Xenium. These emerging ST technologies have unlocked unprecedented possibilities for exploring the transcriptomic architecture of multicellular organisms, revealing intricate cellular heterogeneity in diverse tissues and disease states[5].

However, analytical methodologies that do not take into account spatial context or the underlying histopathology frequently limit the full potential of these potent technologies. In the case of ST analysis methods, many are extensions of earlier strategies that lack direct incorporation of spatial information between spots or not to directly leverage the underlying imaging data of protein and cellular structure. For example, current clustering techniques, such as Louvain and k-means clustering provided by Seurat[6], primarily focus on gene expression[7], often neglecting spatial context and the potential complementary information that can be gleaned from tissue morphology. This incomplete fusion of transcriptomic and morphological data limits our ability to fully understand the cellular ecosystem within tissues, particularly in cancer related states. In particular groups, efforts are being made to integrate imaging data in order to better capture the richness of information embodied in high resolution H&E images[7–10].

In our exploration of the utility of advanced analytical methods for spatial transcriptomics, we developed a novel approach we termed "stMIC" (Spatial Transcriptomics and Morphological Integrated Clustering). Central to stMIC's design is the incorporation of a form of Convolutional Neural Network (CNN) deep learning, specifically the Residual Network-50 or ResNet-50, which is characterized by its versatility and effectiveness across a wide array of applications[3,11]. The Resnet-50 architecture uses the concept of residual connections, and, with its existing prior training, is highly effective for image classification, object detection, and image segmentation[12]. To underline the clinical implications of our study, we deployed our method on previously histopathological assessed disease specimens, with a particular focus on aggressive malignancies such as glioblastoma and triple-negative breast cancer.

2. Method

2.1. *Visium Spatial Gene Expression Assay, Sequencing, and Preprocessing*

Freshly frozen, OCT-embedded tissues were cryosectioned and mounted on Visium spatial gene expression slides (10x Genomics, #1000184), which contain four 6.5 mm * 6.5 mm capture areas comprising 5,000 barcoded spatial features each. Hematoxylin and eosin (H&E) staining was applied, and microscopic images were obtained subsequently with a Zeiss Axioscan2 microscope using a 10x objective. After staining, tissues underwent a permeabilization process to facilitate RNA binding to the slide surface, which was determined using the Spatial Tissue Optimization procedure (10x Genomics, #1000193). The on-slide cDNA synthesized from immobilized RNA was used to generate sequencing libraries, which were paired-end sequenced on an Illumina NovaSeq 6000 instrument to produce a minimum of ~250 million read pairs per sample[13]. The Triple-Negative Breast Cancer (TNBC) sample utilized in our study was originally characterized using Spatial Transcriptomics (ST) in the work of Bassiouni et al[13].

The Space Ranger pipelines (version 1.1.0; 10x Genomics) were employed to preprocess the sequencing data. Demultiplexing of BCL data and conversion to FASTQ format were accomplished using the *spaceranger mkfastq* pipeline. Further, the spaceranger count pipeline enabled read

alignment to the human reference genome GRCh38, UMI counting, and the generation of feature-spot matrices corresponding to the microscopic tissue image. This pipeline also provided automatic tissue detection and fiducial alignment based on the image. Raw gene expression data underwent Counts Per Million (CPM) normalization, and subsequent log transformation, followed by scaling the data to zero mean and unit variance. All these steps were completed using Scanpy (version:1.6.0).[14]

2.2. *Histopathological Image Annotation and Evaluation*

Frozen tissues from each block selected for study were stained with H&E. Images derived from the Visium slides were examined and annotated by a pathologist utilizing Adobe Photoshop software[13]. Regions characterized by blood vessel, necrosis, dense immune cell infiltrates, or stromal fibrosis were indicated when applicable.

2.3. *Histological Image Segmentation and Patch Selection*

The primary stage of image preprocessing entailed segmenting whole-slide H&E histological images from patient samples into smaller patches. Patches were chosen such that each entirely encompassed corresponding spots under the tissue, as indicated in the second column of the "tissue_positions_list.csv" if it is 1. For each spot S_i, a corresponding patch was defined, with the geographic center of the patch aligned with the center of the spot, denoted by coordinates (x_i,y_i) from the last two columns in "tissue_positions_list.csv" file. The patch dimensions were such that both the height and width were equivalent to the diameter of the spot, d, from scalefactors_json.json. Thus, the boundaries of the patch were formally determined by the following coordinates: the upper boundary at $(x_i+d/2)$, the lower boundary at $(x_i-d/2)$, the left boundary at $(y_i-d/2)$, and the right boundary at $(y_i+d/2)$.

2.4. *Feature Extraction Using ResNet-50*

We implemented a convolutional neural network (CNN) model for feature extraction from each patch. We utilized a pretrained ResNet-50 model (Tensorflow version: 2.6.0) trained on the ImageNet dataset for optimal performance in our task. Specifically, it was employed with its top fully-connected layer excluded, selecting "avg" pooling mode for feature extraction (Tensorflow version: 2.6.0), and everything else was default setting. The segmented histology image patches, read in using OpenCV (version: 4.5.3) and resized to (224,224,3), served as inputs. The ResNet-50 model subsequently outputted a 2048-dimensional feature array that represented the patch's morphological features. Feature standardization was achieved using StandardScaler from sklearn (version 0.22.1), which removes the mean and scales to unit variance.

2.5. *Integration of Matrices and Clustering*

Both the normalized gene expression matrix and morphological feature matrix underwent Principal Component Analysis (PCA), separately processed through the top 10 principal components. The resulting matrices were concatenated based on the corresponding barcode of the spots. Clustering was accomplished using the Louvain modularity optimization algorithm with a resolution range of 1.5-1.9 and k-neighbors set at 39, implemented in Orange Data Mining (version

3.30.1). This configuration yielded a stable set of clusters, paralleling the cluster numbers obtained through the current analytical approach solely based on gene expression.

2.6. *Evaluation Measures*

The performance of our method was evaluated using multiple validation metrics. The Adjusted Rand Index (ARI) served as the initial metric, quantifying the similarity between clustering assignments relative to the pathologist's annotation[15]. Gene set enrichment analysis (GSEA) was then performed at the cluster level via the Broad Institute's GenePattern software (RRID:SCR_003199). Utilizing the FindAllMarkers feature of Seurat (version: 4.3.0.1), differentially expressed genes within chosen clusters were identified via Wilcoxon rank sum testing.

Gene lists from each selected cluster were subjected to Pre Ranked GSEA, contrasting against chosen gene sets (H: hallmark gene sets[16], C2:CP:KEGG[17], C4[18], C7: immunesigdb[19]), with permutations set at 1000 and the collapse dataset selected as "Remap_only". The final step involved visualizing spatial expression patterns for genes of interest using the SpatialPlot feature in Seurat.

2.7. *Implementation*

stMIC has been developed with Python 3.7 as a user-friendly pipeline. Setting up and tutorials are described in the stMIC GitHub page: https://github.com/USCDTG/stMIC.

2.8. *Supplemental Material*

Supplemental Material referred to in the paper may be found at the following URL: https://github.com/davcraig75/song_psb2014/blob/main/SupplementaryData.pdf

3. Results

3.1. *Pipeline*

Our primary goal was to enhance identification of biological features from 10X Visium ST by incorporating deep learning analysis. We first show the default approach in Figure 1A, where "spot-level" normalized gene expression obtained via the Space Ranger pipeline is first reduced by principal component analysis (PCA) from (spots x genes) to (spots x M) where M is 10. This is frequently followed by graph-based clustering using a sparse nearest neighbor graph, followed by Louvain Modularity Optimization to identify highly-connected modules in the graph. We note that other clustering methods, such as K-means, are used and are presenting the default clustering method of Space Ranger.

Our stMIC pipeline is shown in Figure 1B and includes partitioning or splitting the ST histological image into segmented tissue spots, followed by feature extraction in the ResNet-50 model on the underlying image for each spot. Thus, if there were 4,096 passing spots, we would have the same number of images. Computationally, this step took approximately 2.5-5 minutes per whole slide image on a machine with an NVIDIA Tesla T4 GPU with 16GB of VRAM, depending on the number of spots under tissue. The gene expression data was simultaneously processed with

the Space Ranger pipelines, turning raw sequencing data into normalized feature-spot matrices within 3 hours on a standard bioinformatics workstation with 32 CPU cores.

Following dimensionality reduction of both morphological and gene expression data via Principal Component Analysis (PCA), we concatenated and performed clustering using the Louvain algorithm, an operation that took roughly 1-2 minutes on the same workstation. To evaluate our method's performance, we used multiple validation metrics, including the Adjusted Rand Index (ARI) and Gene Set Enrichment Analysis (GSEA).

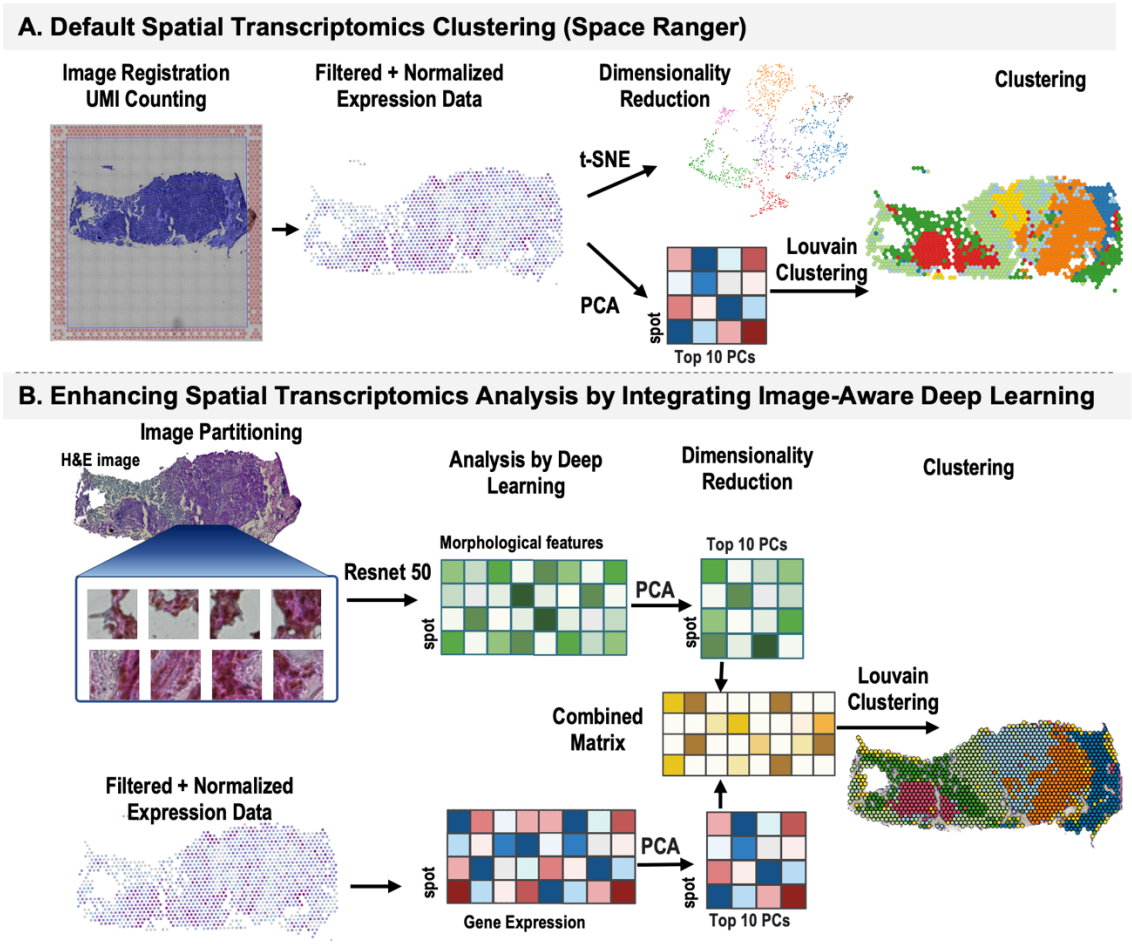

Fig. 1. (A) Default spatial transcriptomics clustering approach, e.g. used in 10X Space Ranger Pipeline. (B) stMIC: ST analysis integrating image-aware deep-learning analysis. High-resolution histology images undergo an initial cropping process into smaller patches, driven by the location and size of spots within the tissue. Subsequently, these patches are introduced to a deep learning model, ResNet-50, resulting in the production of a morphological feature matrix. Principal Component Analysis (PCA) is applied to this matrix, from which the top 10 principal components (PCs) are selected. A parallel procedure is enacted on the gene expression matrix derived from the spatial transcriptomics dataset. The reduced matrices from both the morphological and gene expression data are then concatenated at each spot to form a unified matrix.

3.2. *Application to human glioblastoma spatial transcriptomics data*

In the exploration of the pretrained ResNet-50 model's proficiency, we started with a representative glioblastoma sample, FFD1. This sample encompasses 3,594 spots and 33,538 genes, obtained from the 10x Genomics Visium platform. The analysis commences by contrasting the clustering outcomes from both the Louvain methodology, which is only dependent on gene expression (GEBC) (Fig. 2B), and the approach leveraging H&E histology image feature extraction via ResNet-50 (Fig. 2C). For enhanced visual interpretability, each cluster is assigned a unique color.

Further exploration of the three most significantly differentially expressed genes (DEGs) within this cluster, *HBB*, *HBA1*, and *HBA2*, revealed their critical role in blood biochemistry. *HBB* encodes the beta-globin protein, while *HBA1* and *HBA2* code for the alpha-globin protein, forming essential components of hemoglobin[20]. Predominantly, the expression of these genes was concentrated within the region defined as cluster 10 (Fig. 2D), affirming the blood vessel identity of this cluster.

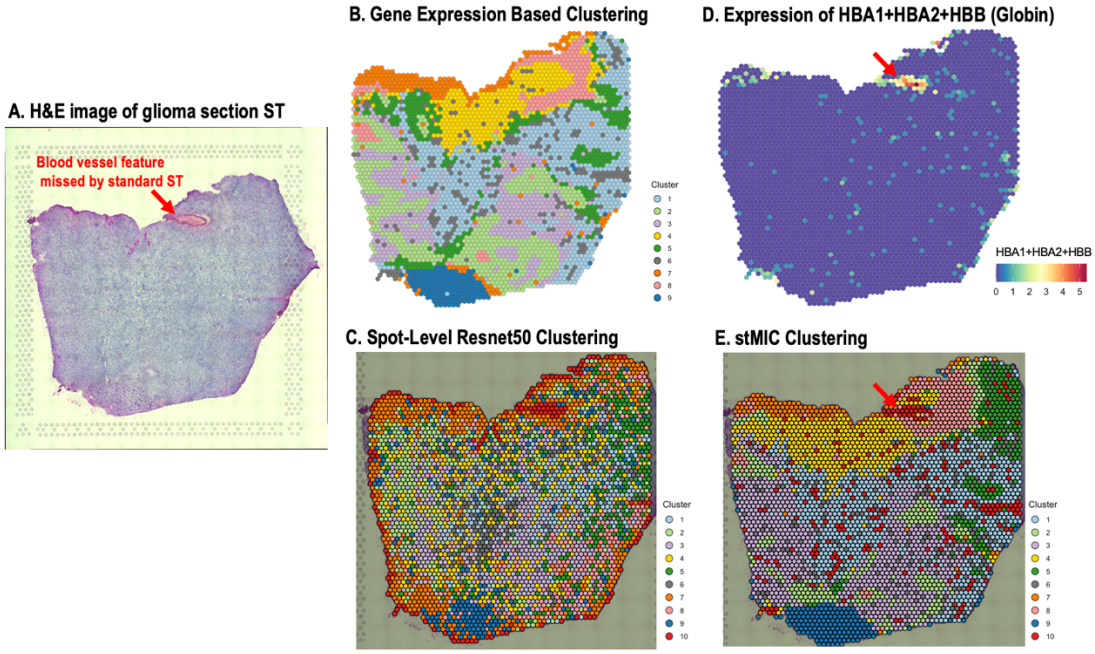

Fig. 2. Comparative analysis of clustering methods and spatial expression patterns of FFD1. (A) Haematoxylin and eosin (H&E) staining image of Glioblastoma sample. (B) Spatial domains identified by gene-expression-based Louvain clustering method. (C) Spatial domains identified by ResNet-50 feature extraction method. (D) Spatial distribution of the top upregulated genes in cluster 10 (HBA1, HBB, HBA2). (E) Spatial domains identified by stMIC method.

These findings attest to the remarkable capability of ResNet-50 in uncovering areas that remain undetected by the gene-expression-based methodology, which relies strictly on gene expression. These advantages provide a solid foundation for developing an even more robust and comprehensive analysis method, stimulating the formulation of an integrated approach. Motivated by this, we

proceeded to implement an advanced strategy that harnesses both histological and gene expression information. This integrative method not only enhances our ability to discern various tissue regions, but also seeks to offer a more nuanced, multidimensional view of the complex histopathological landscape. The harmonization of image-derived and transcriptomics data enables us to move beyond the limitations of each individual data type, allowing a more holistic exploration of biological phenomena at the tissue level. Importantly, stMIC strategy not only successfully pinpoints the blood vessel area, but also enhances edge definition in this key area (Fig. 2E), thus increasing precision in key region detection. These results highlight the potential of the integrated image-aware method methodology for providing comprehensive and accurate histopathological profiling.

We set the number of clusters at ten for all methods to compare the clustering results of SpaGCN, SpaCell, and stLearn. SpaCell, which employs an autoencoder for dimension reduction, failed to detect the EGFR-rich region (not detailed in the main results) and did not align closely with the pathologist-annotated blood vessel region as depicted in Supplementary Fig. 1D. Notably, both SpaGCN and stLearn in their default implementation were also unable to identify the blood vessel region, as shown in Supplementary Fig 1.E&F. Still any interpretation of features missed or seen should be taken with caution since these types of features were not part of their development.

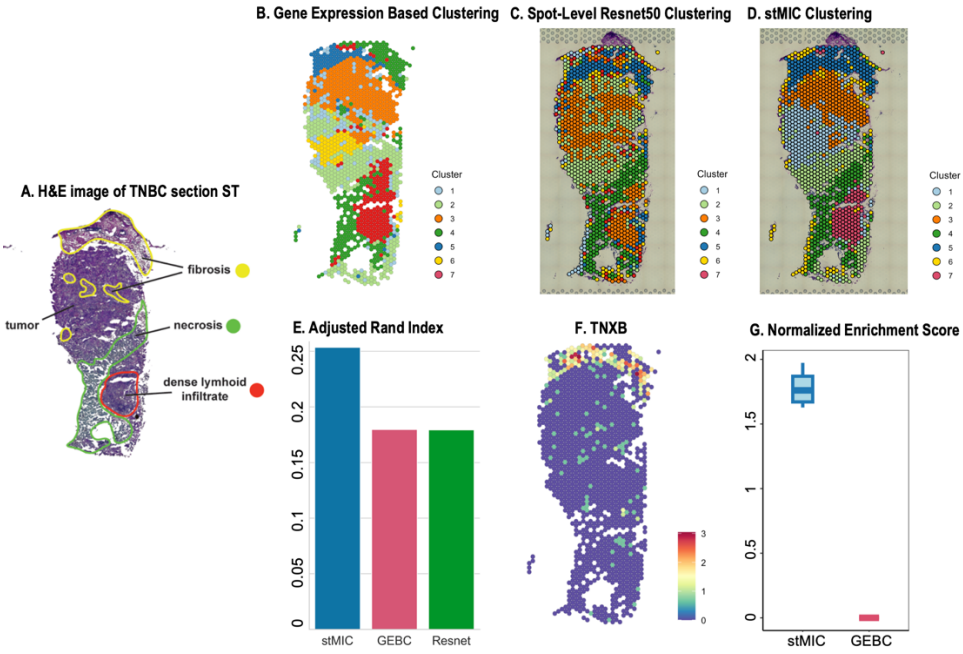

Fig. 3. Comparative analysis of clustering methods and spatial expression patterns of Slide 120D. (A) Histology and manually annotated structure for slide 120D. (B) Spatial domains identified by gene expression-based clustering method. (C) Spatial domains identified by ResNet-50 feature extraction method. (D) Spatial domains identified by stMIC method. (E) Adjusted Rand Index (ARI) in stMIC, gene expression based (GEBC), and ResNet clustering methods determined sections against the ground truth labels (pathologist annotation). (F) Spatial expression of the fibrosis marker gene. (G) boxplot of GSEA for cluster 4 of sample 120D against selected gene sets. NES = Normalized enrichment score.

3.3. *Application to human triple-negative breast cancer spatial transcriptomics data*

In our evaluation of the triple-negative breast cancer sample, termed Slide 120D, ResNet enhanced the performance in distinguishing between fibrotic and necrotic regions, a result that is corroborated by pathologist annotations (Fig. 3A,C). The gene expression-based method could not differentiate these distinct regions, splitting the fibrotic region into two clusters (Fig. 3B). Using an approach previously applied to a glioblastoma sample, we employed our integrated image-aware method, stMIC. Consequently, our stMIC method not only improved clustering of the fibrotic and necrotic regions but also significantly enhanced the accuracy in identifying the dense lymphoid infiltrate region (Fig. 3D). To assess and compare the performance of two clustering methods, the Adjusted Rand Index (ARI) was employed as an evaluation metric. The ARI is a widely used measure that quantifies the similarity between two clustering assignments by comparing their agreement with respect to the ground truth, in this case, the pathologist's annotation. Clustering based on the integration of morphological features with gene expression was more consistent with the regional annotations obtained by pathologists (ARI = 0.2536) compared to gene expression-based clustering method (ARI = 0.1796) and ResNet-50 feature extraction clustering (ARI = 0.1793) (Fig. 3E). To discern each method's proficiency in detecting intricate tissue structures, we observed stLearn's clustering closely mirrored that of the gene expression-based Louvain clustering. Both stLearn and SpaGCN does not segregate the fibrosis and necrosis regions (Supplementary Fig 2.E&F). Specifically, SpaCell managed to identify almost the entire fibrosis region but did not fully differentiate between these two crucial areas (Supplementary Fig. 2D). The Adjusted Rand Index (ARI) is 0.173 for SpaGCN, 0.151 for stLearn, and 0.125 for SpaCell (Supplementary Fig 2H). At some level, these results must be taken with caution since these tools had not been evaluated or designed for these types of pathologies.

From a biological point of view, improved clustering is supported by the fact that the key marker tenascin-XB (TNXB) of fibrosis shows high differential expression in cluster 4 (Fig. 3D&F). Tenascin-XB, a key component of the extracellular matrix, has been linked to tissue remodeling and fibrosis, and is often upregulated in fibrotic tissues[21] . Its elevated expression and localization within the fibrotic region not only strengthen the characterization of this region but also implies the ongoing process of tissue remodeling - a common event in fibrosis[22]. This discovery again emphasizes the capacity of our integrated method to reveal crucial biological elements and events, contributing to a more comprehensive understanding of tumor progression.

Furthermore, stMIC approach proved better in capturing biologically relevant features within the detected fibrotic region compared to the gene expression-based clustering method. Gene set enrichment analysis revealed significant distinctions between the two methods (Fig. 3G). Only results with a false discovery rate (FDR) < 0.05 are displayed. While no significant pathways were detected using the expression-based method, our integrated approach identified four significantly enriched pathways. These pathways were indicative of active immunity, antigen processing and presentation, as well as humoral immune and inflammatory responses, all with Normalized Enrichment Scores (NES) greater than 1.5 (Supplementary Fig. 2G). These findings demonstrate the profound immune involvement within the fibrotic area, further underscoring the added value of stMIC in capturing these nuanced dynamics.

In another triple-negative breast cancer sample, 094D, in a horizontal comparison of results from the gene expression based, ResNet-50, and stMIC methods with annotation, stMIC approach unveiled critical biological phenomena (Fig. 4A-D). Notably, it achieved higher accuracy in capturing the middle right fibrotic region, a conclusion that was further corroborated by the higher Adjusted Rand Index (ARI) in the stMIC (ARI = 0.1479) vs. gene-expression alone (0.1111) or ResNet feature clustering (0.1051). In a comparison between the clustering results of stMIC and spaCell, Figure 4G indicates SpaCell did not differentiate between these sub-clusters (Supplementary Fig. 3D). While both spaGCN and stLearn could differentiate these sub-clusters((Supplementary Fig 3.E&F), their performance was not markedly superior to stMIC. When measured against pathologist annotations using the ARI), SpaGCN, stLearn and SpaCell recorded ARIs of 0.098, 0.127 and 0.091, respectively (Supplementary Fig 3H).

In-depth analysis of marker genes, facilitated by the stMIC approach, revealed notable features absent from gene-expression-only clustering. Specifically, our approach discerned that what was identified as cluster 1 in gene-expression-only clustering comprised two distinct, biologically relevant clusters. One of these exhibited increased expression of hypoxia markers CA9/NDRG1, while the other was characterized by the presence of IFIT1, a marker indicative of an active interferon response (Fig. 4 F-H).

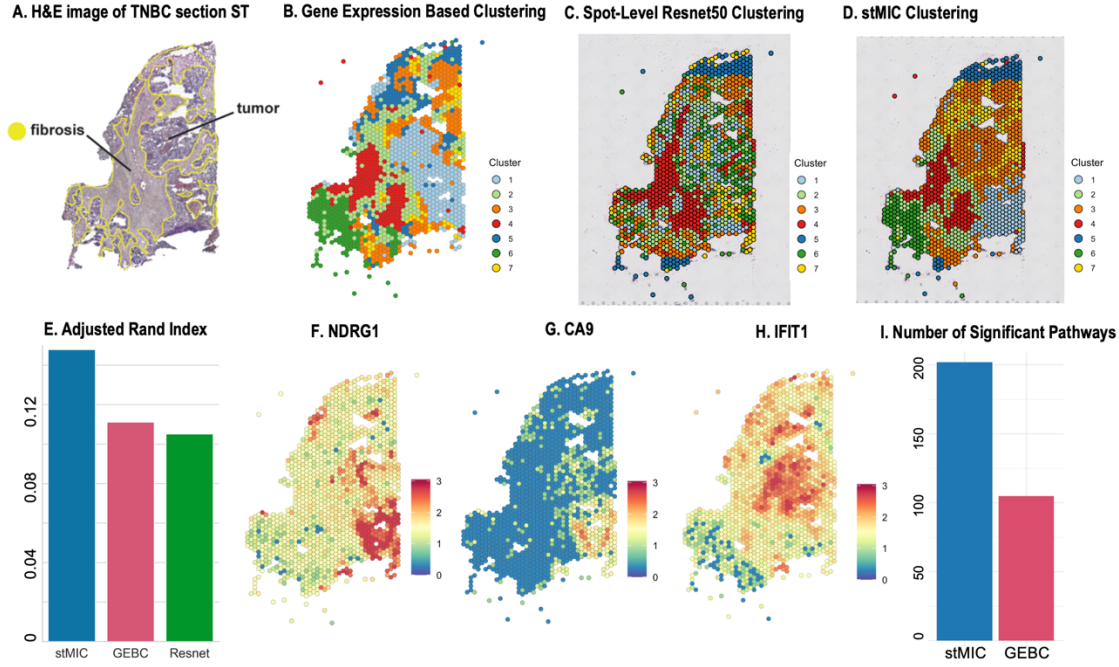

Fig. 4. Comparative analysis of clustering methods and spatial expression patterns of Slide 94D. (A) Histology and manually annotated structure for slide 94D. (B) Spatial domains identified by gene-expression-based Louvain clustering method. (C) Spatial domains identified by ResNet-50 feature extraction method. (D) Spatial domains identified by stMIC method. (E) Adjusted Rand Index (ARI) in stMIC, gene-expression (graph-based), and ResNet-50 clustering methods determined sections against the ground truth labels (pathologist annotation). (F-G) Spatial expression of hypoxia marker genes. (H) Spatial expression of hypoxia marker gene. (I) Boxplot of GSEA for cluster 4 of sample 94D against selected gene sets.

The genes CA9 (Carbonic Anhydrase IX) and NDRG1 (N-Myc Downstream Regulated 1) (Fig. 4F&G) are known to be upregulated under hypoxic conditions, which often occur in solid tumors such as breast cancer due to inadequate oxygen supply[23,24]. This upregulation is a response to the low oxygen tension in an attempt to adapt to the harsh microenvironment. Hypoxia within tumors is associated with increased invasiveness, resistance to therapy, and a poor prognosis, thus indicating a potentially more aggressive disease state within this cluster[25].

In addition to these hypoxia markers, stMIC discerns a separate region with high expression of IFIT1 (Fig. 4H). This interferon-induced protein has been associated with active interferon signaling and an ongoing immune response. Active interferon signaling can have complex implications in cancer, possessing both tumor-suppressive and tumor-promoting properties[26]. Meanwhile, an ongoing immune response may reflect the immune system's attempts to counteract tumor progression or could even suggest the shaping influence of the immune system on the tumor's behavior[27]. Beside this, *IFIT1* has been associated with chemotherapy response in breast cancer[28], suggesting that this region might be more susceptible to chemotherapy.

In fact, previous studies have validated the relevance of these two distinct clusters, but this was determined only through a comprehensive joint expression analysis spanning 28 diverse samples[13]. Thus, while expression-analysis does identify these sub-clusters, it is reliant on the comparative context gleaned from multiple, external samples. Furthermore, our stMIC approach has identified significantly more enriched pathways in cluster 4 than what the gene-expression-based method alone was able to capture (Fig.4 I). The gene set enrichment score of top five Hallmark pathways are further compared between gene expression based vs. stMIC clustering method (Supplementary Fig.3E). Only results with a false discovery rate (FDR) < 0.05 are displayed. This difference underscores the potential of the integrated image-aware approach to provide a richer, more nuanced analysis of gene expression within specific tissue regions. In conclusion, our integrated image-aware approach has provided new insights into the spatial heterogeneity of the 094D triple-negative breast cancer sample, unraveling the complexities of hypoxia and immune responses within the tumor environment. The incorporation of morphological information in gene expression analysis can enhance the resolution of tumor substructure identification and pave the way for a more nuanced understanding of tumor biology, with potential implications for treatment strategies.

4. Discussion

Our results showed how integrating a naïve pre-trained ResNet-50 into Spatial Transcriptomics workflows identified features missed from standard ST gene-expression analysis. Second, our results show an improved ability to recapitulate features identified by a pathologist.

An image-aware, integrated approach can identify previously overlooked or undervalued features. The most prominent feature identified from the combined approach was the glioma's blood vessel and vascularization features. These features are unmistakable in the histology image. However, only three genes (*HBA*, *HBA1*, and *HBA2*) within this biological feature have significant expression within the gene expression data. While this is expected from globin-producing erythrocytes, these three genes did not contribute significantly to the first 10 PCs used in gene expression clustering. In practice, given that hallmarks of the disease involve post-translational

modifications, many types of features would be less evident by gene expression alone. The ability to highlight these features underscores the power of stMIC approach that draws from the strengths of deep learning and the rich data afforded by spatial transcriptomics.

Clear identification of novel features and improved clustering is encouraging, given that there are several areas where the approaches used here could be improved or optimized. First, we utilized a naive, pre-trained ResNet-50 model. Training approaches could lead to improved clustering. However, histopathological image data often have well-described biases that limit transferability [29]. Indeed, some groups have identified approaches for internal training, as in the case of the tool stMIC, which shows promising early results in disease relevant contexts[30]. Thus, the value seen even in this naive model is notable, as improvement was seen without further training.

Our approach differs significantly from RESEPT[31], stLearn[32], and SpaCell[9], which are well established pipelines using the ResNet-50 model in spatial transcriptomics analysis. RESEPT uses a graph autoencoder to embed ST gene expression data into a three-dimensional representation and maps this to an RGB image for visual analysis using the ResNet101 deep learning model. Unlike RESEPT, which does not consider H&E image data, our approach integrates both spatial gene expression and H&E image data to capture comprehensive morphological and transcriptional details, offering the potential for richer biological insights. Like stLearn, our pipeline utilizes a pre-trained ResNet-50 model, but we diverge by integrating the image data with gene-expression clustering, rather than just normalization, offering a richer exploration of tissue features. By comparison, SpaCell also uses a ResNet-50 model for feature extraction and integrates imaging with transcriptomics; they differ by using autoencoders to reduce features and mainly focusing on classification/prediction.

Within this work, we leveraged pathology annotated samples to assess performance, and very importantly, while some methods employ models on non-pathological systems, such as Spatial Transcriptomics (ST) on a mouse cortex or similar non-disease settings, their utility can be limited in capturing the full spectrum of cellular interactions in aggressive human diseases like cancer. Instead, our approach prioritizes pathological samples from patients, given their inherent complexity and heterogeneity. By targeting unique pathological landmarks and intricacies, our method seeks to illuminate tumor progression and identify key biological elements and events. While we can see improved results using pathology annotation as a benchmark, it has limitations. First, these samples were predominantly done on fresh-frozen tissue, whereas pathologists typically prefer annotation on FFPE fixed tissues. With the emergence of spatial transcriptomics approaches that work with FFPE tissue, these ideally annotated histology images will become more available; however, FFPE usually comes with lower quality ST due to degraded RNA.

A deeper dive into understanding the exact mechanisms by which ResNet-50 outperforms solely gene-expression-based approaches would be beneficial. Unraveling the strengths and limitations of image-aware deep learning model in the context of spatial transcriptomics can provide a foundation for optimizing or even designing new architectures that can more comprehensively capture pathology-related features. We recognize this as a crucial next step and are planning further investigations to elucidate the specific advantages of these models and to guide the development of even more effective methodologies.

Some advanced deep learning architectures have emerged that could potentially enhance the methods presented in this study. The Vision Transformer (ViT)[33], for instance, processes images by segmenting them into fixed-size patches and then leverages the Transformer architecture, offering the potential to extract more nuanced spatial relationships crucial for spatial transcriptomics. Similarly, the MLP-Mixer[34], through its unique mixing of tokens with multilayer perceptrons, and the Swin Transformer[35], with its shifted windows approach, can be particularly advantageous for capturing intricate spatial hierarchies and features from histology images. Our study predominantly utilized the ResNet-50 model for its proven efficacy in image analysis. However, integrating recent architectures like ViT, MLP-Mixer, and Swin Transformer might allow for a more comprehensive feature extraction, bridging image-based nuances with spatial transcriptomic insights.

Several potential extensions and improvements could be made to the presented approaches. Consideration could be given to incorporating other types of omics data to enrich the data pool further. Refinement of the deep learning model could enhance performance by integrating new layers or algorithms. A promising avenue to explore is the incorporation of spatial distances between spots into the analysis, offering a more nuanced understanding of cellular organization.

A broader sample set and more detailed annotations would certainly strengthen the robustness of our model. The requirement for specialized technical expertise to operate and interpret results is another hurdle that needs to be overcome. Furthermore, validating our methodology using a more comprehensive range of patient-derived data is necessary to ascertain the model's clinical relevance and translational applicability.

Beyond cancer, the potential application of our approach to other diseases warrants further exploration. With the in-depth view of the tumor microenvironment that our method provides, we foresee a crucial role for it in the realm of personalized medicine, particularly given advances in immune-oncology treatments. As our understanding of cellular heterogeneity within tissues becomes increasingly nuanced, an image-informed ST approach will likely serve as a powerful tool in understanding the role of the immune microenvironment. This work underscores the exciting potential of spatial transcriptomics and deep learning in shaping the future of understanding disease heterogeneity.

Availability of data and materials.

The stMIC code has been developed with Python 3.7 as a user-friendly pipeline. Code, setting up and tutorials are described in the stMIC GitHub page: https://github.com/USCDTG/stMIC.

Supplemental Material referred to in the paper may be found at the following URL: https://github.com/davcraig75/song_psb2014/blob/main/SupplementaryData.pdf

References

1. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. Nature 596, 211–220 (2021).
2. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. Bioessays 42, e1900221 (2020).
3. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. Nat. Methods 16, 987–990 (2019).
4. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090 (2015).
5. Li, Q., Zhang, X. & Ke, R. Spatial Transcriptomics for Tumor Heterogeneity Analysis. Front. Genet. 13, 906158 (2022).
6. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21 (2019).
7. Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. Nat. Methods 15, 339–342 (2018).
8. Hu, J. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat. Methods 18, 1342–1351 (2021).
9. Tan, X., Su, A., Tran, M. & Nguyen, Q. SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells. Bioinformatics 36, 2293–2294 (2020).
10. Ifvarsson, Falk, Vidman & Thorén. [Social Welfare Department's general advice on prevention, diagnosis and treatment of eye infections in newborn infants]. Jordemodern 99, 398–403 (1986).
11. Website. https://doi.org/10.48550/arXiv.1512.03385 doi:10.48550/arXiv.1512.03385.
12. Kutluer, N., Solmaz, O. A., Yamacli, V., Eristi, B. & Eristi, H. Classification of breast tumors by using a novel approach based on deep learning methods and feature selection. Breast Cancer Res. Treat. 200, 183–192 (2023).
13. Bassiouni, R. et al. Spatial Transcriptomic Analysis of a Diverse Patient Cohort Reveals a Conserved Architecture in Triple-Negative Breast Cancer. Cancer Res. 83, 34–48 (2023).
14. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).
15. Steinley, D. Properties of the Hubert-Arabie adjusted Rand index. Psychol. Methods 9, 386–396 (2004).
16. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417–425 (2015).
17. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30 (2000).
18. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. Nat. Genet. 36, 1090–1098 (2004).
19. Godec, J. et al. Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. Immunity 44, 194–206 (2016).

20. Richter, F., Meurers, B. H., Zhu, C., Medvedeva, V. P. & Chesselet, M.-F. Neurons express hemoglobin alpha- and beta-chains in rat and human brains. J. Comp. Neurol. 515, 538–547 (2009).
21. Cohen, C. et al. The roles of Tenascin C and Fibronectin 1 in adhesive capsulitis: a pilot gene expression study. Clinics 71, 325–331 (2016).
22. Caja, L. et al. TGF-β and the Tissue Microenvironment: Relevance in Fibrosis and Cancer. Int. J. Mol. Sci. 19, (2018).
23. Kuhlenbäumer, G., Stögbauer, F., Ringelstein, E. B. & Young, P. Hereditary Peripheral Neuropathies. (Springer Science & Business Media, 2005).
24. Shamis, S. A. K., Edwards, J. & McMillan, D. C. The relationship between carbonic anhydrase IX (CAIX) and patient survival in breast cancer: systematic review and meta-analysis. Diagn. Pathol. 18, 46 (2023).
25. Jing, X. et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Mol. Cancer 18, 157 (2019).
26. Minn, A. J. Interferons and the Immunogenic Effects of Cancer Therapy. Trends Immunol. 36, 725–737 (2015).
27. Hiam-Galvez, K. J., Allen, B. M. & Spitzer, M. H. Systemic immunity in cancer. Nat. Rev. Cancer 21, 345–359 (2021).
28. Weichselbaum, R. R. et al. An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. Proc. Natl. Acad. Sci. U. S. A. 105, 18490–18495 (2008).
29. Hägele, M. et al. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. Sci. Rep. 10, 6423 (2020).
30. Zuo, C. et al. Elucidating tumor heterogeneity from spatially resolved transcriptomics data by multi-view graph collaborative learning. Nat. Commun. 13, 5962 (2022).
31. Chang, Y. et al. Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. Comput. Struct. Biotechnol. J. 20, 4600–4617 (2022).
32. Pham, D. et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. bioRxiv 2020.05.31.125658 (2020) doi:10.1101/2020.05.31.125658.
33. Dosovitskiy, Alexey, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).
34. Tolstikhin, Ilya O., et al. Mlp-mixer: An all-mlp architecture for vision. Advances in neural information processing systems 34 (2021): 24261-24272.
35. Liu, Ze, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision. 2021.