

Optical phenotyping using label-free microscopy and deep learning

Shuyuan Guan,^a Thomas Knapp^b, Alba Alfonso-Garcia^{b,c}, Suzann Duan^{b,d}
and Travis W. Sawyer^{b,a,b,*}

^aThe University of Arizona, Wyant College of Optical Sciences, Tucson, Arizona, United States

^bThe University of Arizona, Department of Biomedical Engineering, Tucson, Arizona, United States

^cUniversity of California Davis, Biomedical Engineering, Davis, California, United States

^dUniversity of California Irvine, Department of Pharmaceutical Sciences, School of Pharmacy and Pharmaceutical Sciences, Irvine, California, United States

Statement of Discovery

We develop and demonstrate a method for optical phenotyping using deep learning to classify label-free microscopy images into tissue phenotypes defined by transcriptomic signatures.

ABSTRACT. **Significance:** Tissue phenotyping plays a critical role in biomedical research and clinical applications by providing insight into the structural and functional characteristics of tissues that can characterize clinical behavior and identify therapeutic targets. However, conventional phenotyping techniques are destructive, time-intensive, and expensive, posing challenges for both efficiency and widespread use.

Aim: We aim to develop an optical phenotyping approach in pancreatic cancer specimens using label-free multiphoton microscopy combined with spatial transcriptomics and deep learning.

Approach: We measure and co-register a dataset comprised of spatial transcriptomics, autofluorescence, and second harmonic generation microscopy. We then cluster tissue subregions into meaningful phenotypes using transcriptomic signatures. We evaluate three different classification models to predict phenotype based on label-free imaging data, and we assess generalizability and prediction accuracy.

Result: Our deep-learning classification model achieves over 89% accuracy in classifying six tissue types using label-free microscopy images. The one-versus-rest area under the curve (AUC) values for all classes approach 1, confirming the robustness of our model.

Conclusion: We demonstrate the feasibility of optical phenotyping in distinguishing the structural and functional characteristics of pancreatic cancer specimens. Integrating additional gene-expression data or complementary label-free imaging modalities, such as fluorescence lifetime imaging microscopy, holds the potential to further enhance its accuracy and expand its applications in clinical research and diagnostics.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.BIOS.2.3.035001](https://doi.org/10.1117/1.BIOS.2.3.035001)]

Keywords: optical phenotyping; spatial transcriptomics; label-free microscopy; autofluorescence; deep learning; convolutional neural networks

Paper 250037GR received Mar. 31, 2025; revised Jun. 20, 2025; accepted Jun. 26, 2025; published Jul. 11, 2025.

*Address all correspondence to Travis W. Sawyer, tsawyer9226@arizona.edu

1 Introduction

Tissue phenotyping plays a decisive role in profiling disease progression and onset, particularly in the context of cancer.¹ For example, many phenotyping efforts have focused on profiling the tumor microenvironment (TME), which plays an important role in disease progression and metastasis.² With the recent advancements in precision medicine, tumor phenotyping is a crucial step to identifying specific therapeutic targets for diseases that are potentially heterogeneous, for example, in pancreatic cancer.³ Toward this end, phenotyping of *ex vivo* tissues, including those based on sequencing approaches, can increase our understanding of TME features that determine disease prognosis.⁴ Although this provides incredible power for mapping tissue and cell genotype–phenotype relationships, an outstanding challenge is the lack of approaches that are compatible for *in vivo* tissue phenotyping. In addition, many such methods require significant processing time and can be costly—the ability to identify phenotypes in a minimally invasively fashion, and without lengthy and expensive processes, could offer opportunities to glean insight into fundamental processes of disease onset and evolution in basic science research, as well as lead to unique approaches for advancing precision medicine. In this paper, we primarily focus on tissue phenotypes that are defined by gene expression as measured through RNA sequencing, or transcriptomics, as this provides a description of underlying biological processes that dictate tissue function.⁵

Label-free biomedical imaging techniques could provide an avenue for rapid, scalable, and/or *in vivo* compatible “optical phenotyping” by sensing naturally occurring contrast to visualize fundamental tissue properties, which provide powerful insight into biological processes and pathological events.⁶ For example, autofluorescence and fluorescence lifetime imaging can probe endogenous markers of metabolism, immune response, and vasculature.^{7–10} Other methods, such as polarized light imaging, second harmonic generation, or optical coherence tomography, can visualize microstructural features such as collagen distribution and the extracellular matrix.^{11–13} Different disease phenotypes can exhibit clear alterations in these characteristics—for example, immunologically cold and hot tumors within a specific type of cancer respond differently to immunotherapies that rely on the presence of tumor immune infiltrate.¹⁴ Other examples include extracellular matrix alterations that promote tumor cell invasion and migration toward vasculature.¹⁵ As such, it follows that label-free imaging could be utilized for minimally invasive and rapid measurement of key tissue properties to monitor biological events such as disease onset or to identify specific tissue phenotypes. However, an outstanding challenge is that the relationships between genetic pathway alterations that dictate phenotypes and the resulting downstream functional changes can be highly non-linear, which has inhibited the ability to accurately map underlying mechanistic pathways related to label-free contrast. An added complexity is that tissue changes associated with different diseases and phenotypes can occur at multiple scales and are not represented fully by the abundance of a given molecule¹⁶—the spatial organization and distribution of these molecules across different spatial scales carry significant biological information. As such, fully realizing the potential of label-free imaging for optical phenotyping requires the integration of more sophisticated approaches for image feature extraction.

Recent advancements in deep learning may offer a solution to mapping the nonlinear sequence–function relationship,¹⁷ offering a unique opportunity to bridge the gap and perform optical phenotyping of tissues. Since the explosion of artificial intelligence (AI) in biomedical imaging, a myriad of research efforts have aimed to integrate AI methods into different applications across the field. One of the most common applications is for image classification, where AI models such as deep neural networks (DNNs) are trained to classify images with extremely high accuracy.¹⁸ DNNs offer the ability to extract multiscale hierarchical spatial features in image data, offering exceedingly high performance for image analysis tasks. Binary, multiclass, and multilabel classification problems have all been investigated with much success; however, one limitation to these approaches is that the classes and/or labels are generally established based on general tissue categories, often with a significant subjective nature (e.g., diseased, normal, or pathological grades).¹⁹ Diseases such as cancer present with significant heterogeneity within tumor types and disease grades, and this intratumor heterogeneity contributes to drastically different phenotypes and clinical outcomes.^{20,21} These nuances are lost when grouped together

qualitatively, severely limiting both the performance and utility of these approaches. Tissue phenotypes can be reliably established quantitatively based on the transcriptomic signature, and developing technology to classify tissue into different phenotypes using label-free imaging would be a powerful tool in biomedical research as it could enable rapid, nondestructive, minimally invasive, and high-precision tissue assessment.²²

To date, several important works have implemented optical imaging alongside transcriptomic analyses. For example, Haase et al.²³ described the use of microscopy to isolate cells and preserve spatial information prior to single-cell sequencing. Separately, Zhang et al.²⁴ described the use of surface-enhanced Raman spectroscopy to obtain insights into the transcriptome, proteome, and metabolome at the single-cell level. Similarly, another work was recently published that introduced “Raman2RNA,” a method to estimate single-cell expression profiles using Raman spectroscopy and deep learning.²⁵ However, all of these works focus on single-cell omics analyses, and not at the tissue-level, where the surrounding microenvironment carries clinically relevant information, and may be composed of many different cell types.

In this work, we demonstrate the ability to perform optical phenotyping at the tissue level using label-free microscopy in a set of patient specimens of pancreatic cancer. A graphical abstract is included in Fig. 13 in the [Appendix](#). This work is among the first to use label-free imaging contrast to estimate omic characteristics at the tissue level. We perform autofluorescence and second-harmonic generation (SHG) imaging through multiphoton microscopy (MPM) on these specimens, as well as spatial transcriptomics. We co-register the dataset so that transcriptomic signatures are aligned with a region of interest of the label-free imaging modalities. Using this co-registered dataset, we first cluster tissue regions into different tissue types and phenotypes and define these by mapping gene expression to the Human Protein Atlas. We extract image features for each channel, including abundance and Haralick texture features. We performed statistical analysis on label-free imaging contrast among clusters and then built several classifiers to classify tissue using label-free images into phenotypes defined using transcriptomic signatures. Our results show that the abundance of different label-free imaging channels is not sufficient for differentiating tumor phenotypes. Image texture describing spatial distribution of abundance provides higher performance, and using unsupervised feature extraction coupled with DNN classification achieves over 90% accuracy in performing optical phenotyping. These results show the value of hierarchical multiscale feature extraction for capturing relevant biological characteristics. These promising results lay the foundation for our future work that will focus on expanding our patient cohort, optimizing DNN architecture, investigating the generalizability to other organs and diseases, and validating these findings in fresh tissues.

2 Methods

2.1 Sample Preparation

This study was performed on four specimens collected from four different patients who were diagnosed with pancreatic neuroendocrine tumors (NETs) via pathological examination. The formalin-fixed paraffin-embedded (FFPE) specimens were obtained from the University of Arizona Tissue Acquisition and Cellular/Molecular Analysis Shared Resource, originally collected under IRB 0600000609. All specimens were tested for RNA quality by evaluating the RIN and DV200 score. All specimens yielded a DV200 score of at least 50, and a RIN score of at least 2.3.^{26,27} The specimens were sectioned at a thickness of 5 microns onto standard microscope slides. Two serial sections for each specimen were used in this study: one was placed on a specific slide for spatial transcriptomics purchased from 10× Genomics (Visium slide; shown on the left in Fig. 1), and a second adjacent section was used for MPM. Additional clinical information about these patient specimens is included in Table 7 in the [Appendix](#).

2.2 Imaging Data Collection

MPM imaging was performed on a single unstained histological section with a thickness of 5 microns using a Zeiss LSM 780 NLO system with an automated 3D translation stage. A tunable MaiTai Ti:Sa laser and GaAsP detectors were used to image in five spectral channels, four channels capturing the signal from tissue autofluorescence, and a SHG channel. The excitation and

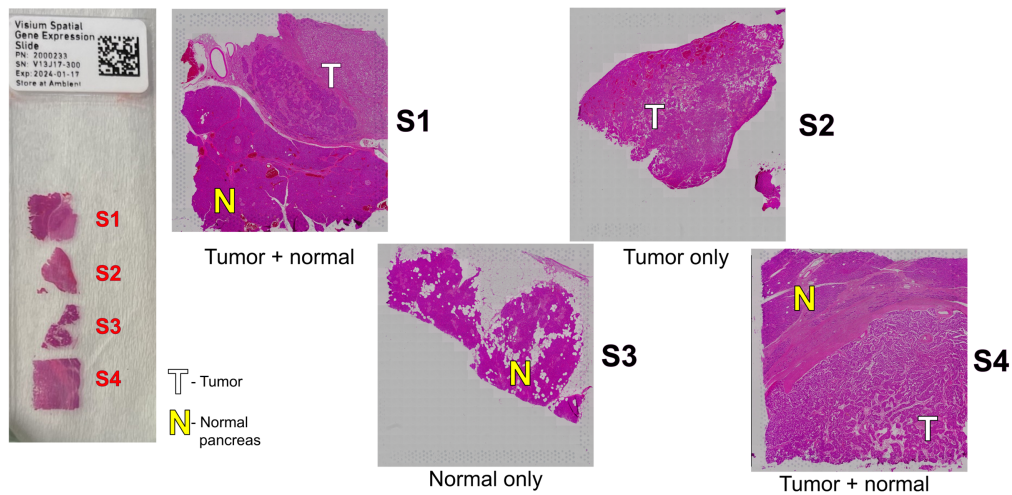


Fig. 1 Diagram of pancreatic cancer patient specimens (S1 to S4) used for the research study. Tumor and normal regions are denoted with T and N, respectively. The left side of the figure shows the specimens placed on the Visium slide for spatial transcriptomics.

Table 1 Imaging channels used for label-free microscopy. All images were acquired using two-photon excitation; thus, the excitation wavelength is half of the listed wavelength.

Channel	Two-photon excitation wavelength (nm)	Detection wavelength band (nm)	Potential fluorophores/sources of contrast
1	700	425 to 465	NADH/Collagen
2	920	475 to 600	FAD
3	830	550 to 600	Lipofuscin/lipopigments
4	800	590 to 625	Porphyrin
5	880	430 to 450	Organized collagen (SHG)

detection wavelengths are shown in Table 1 and have been used in previous studies.⁸ Each MPM channel is referred to in this paper by the endogenous molecule(s) believed to be the primary source of signal when imaging with the spectral parameters in Table 1. This terminology is meant to provide ease of reference as multiple fluorescent species can provide overlapping spectra, and tissue fixation may alter endogenous markers. Although these fluorescence properties could be elicited with single-photon fluorescence, we use multiphoton to also elicit SHG simultaneously. 256×256 pixel tile scans were collected in a grid array over each sample and combined with post-processing to form a composite image of the full specimen as it was imaged on the Visium slide. A $20\times$ objective was used, providing a square pixel resolution of ~ 1.28 micron. Z-stacks were acquired for each tile scan to ensure that an in-focus signal was collected over the full range of the uneven tissue surfaces, with the number of z-stacks adjusted as necessary. Despite z-stacking, some uneven brightness within the individual tile scans results in a grid-like artifact in the image mosaics. Postprocessing that has been previously described was used to balance image brightness between image tiles prior to quantification.²⁸

2.3 Spatial Sequencing

An adjacent histological section for each patient sample was cut onto the Visium slide before being deparaffinized and stained with hematoxylin and eosin (H&E). All procedures followed the standard $10\times$ Genomics protocols for FFPE tissue, and the sections were then slide-scanned using a Nikon BioPipeline SLIDE system (Nikon, Tokyo, Japan). Spatial transcriptomics was completed using the Visium Spatial Gene Expression system ($10\times$ Genomics, Pleasanton,

California, United States). Individual libraries were processed using 10× Genomics' Space Ranger count pipeline. The length of the R2 sequences was limited to the first 50 bps. Space Ranger v2.1.0 count pipeline was used to manually align the fiducial frames of the Visium slide H&E images to gene expression data. For easier export, the Visium Space Ranger software down-samples the original high-resolution H&E image such that its longest dimension is 2000 pixels, resulting in the other dimension being resized based on the image's original aspect ratio. The exported data frame of gene expression counts contains the Visium barcode coordinates from the original, full-resolution, H&E images. Each barcode region covered a circular area with a diameter of 55 microns and is spaced by 100 microns in a Cartesian grid. Due to the uniform down-sampling of the original image, a single scaling factor is provided by the software to match the original barcode coordinates to the dimensions of the exported, down-sampled, and H&E image, to which the MPM images are registered. This scale factor ranged between 0.040 and 0.044.

2.4 Data Processing

All image registration was performed using the open-source TrakEM2 ImageJ plugin.^{29,30} First, multichannel MPM images were registered to one another to eliminate image drift between acquisitions, which could arise from variations in spectral properties of the system's optics, physical movement of the tissue sample, and/or drift of the scanning mechanisms [Fig. 2(a)]. Following inter-channel registration, MPM images were then manually co-registered to the H&E reference image from the Visium dataset. The down-sampled H&E reference image matched to the Visium gene expression data was used as the final registration target due to the original slide-scanned images being prohibitively large to process (on the order of 50,000 pixels in each dimension). Because the down-sampled H&E images are limited to 2000 pixels in the largest dimension, they were first scaled to dimensions roughly matching that of the MPM images. In the case of registering the MPM images to the Visium data, the down-sampled H&E image was scaled by a factor of three, changing its largest dimension to 6000 pixels. With each MPM image having native dimensions of roughly 5500 × 5500 pixels, any differences in scale were further accounted for during the registration process.

For all subsequent registration steps, registration landmarks were manually placed on image regions that contained clear tissue landmarks, such as glands or cell clusters with unique morphology, as shown in Fig. 2(b). Landmarks were selected roughly uniformly over the full spatial area of each image, and at least 80 landmarks were selected for each round. Multiple rounds of manual landmark placement and registration were necessary to register MPM images to the

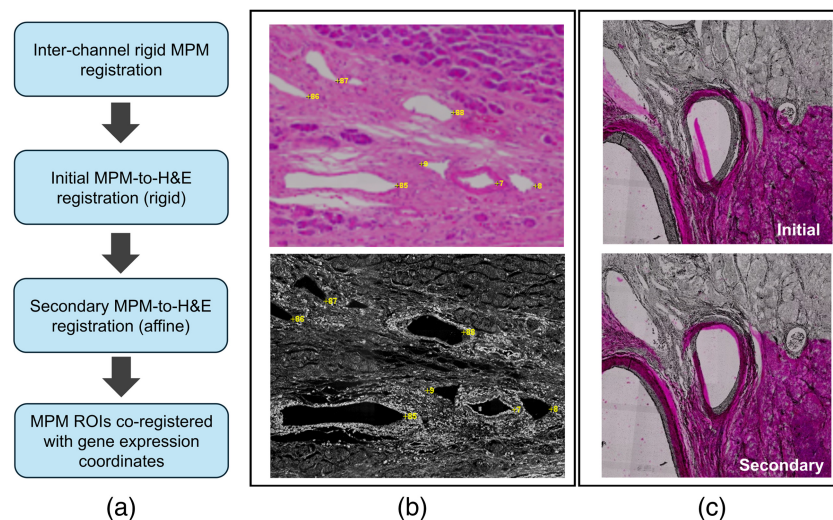


Fig. 2 (a) Diagram of the registration process to align the label-free image ROIs and transcriptomic signatures. (b) Large-scale morphological features were manually selected to register between H&E and MPM images. (c) Visual results of the two-stage registration process.

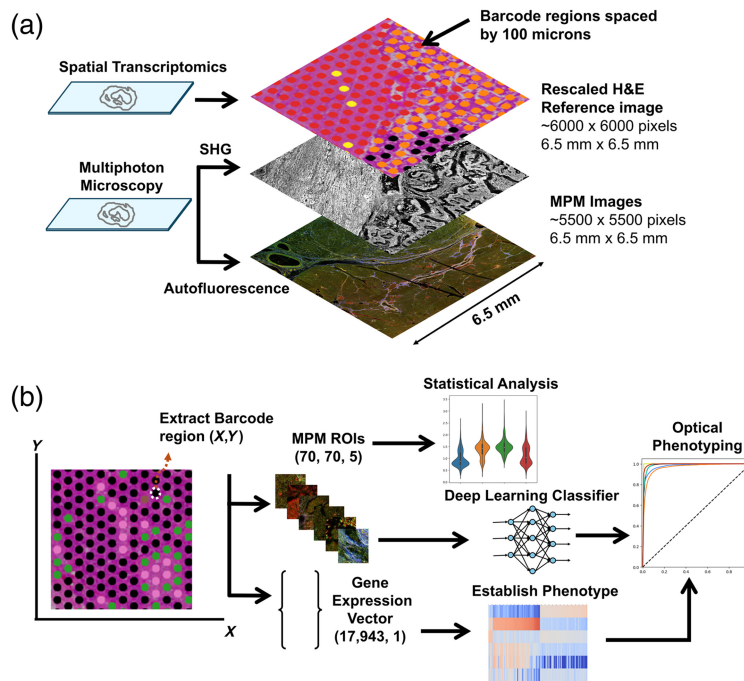


Fig. 3 Schematic diagram of the study. (a) Using two adjacent patient sections, a co-registered dataset composed of spatial transcriptomics, autofluorescence, and SHG images was generated. (b) This dataset was then used to perform statistical analysis of imaging features between clusters defined by transcriptomic signatures and to develop a deep learning model for optical phenotyping.

reference H&E image due to differences in scale and distortion of the tissue during the process of mounting it onto the Visium slide and H&E staining. An initial round of rigid registration through translation, rotation, and isotropic scaling was done to match image dimensions and roughly overlay landmarked regions. An affine transformation was then used with newly placed landmarks, repeating as necessary to align image contents [Fig. 2(c)] to minimize mean-squared error. Following registration, tiles were visually compared between MPM and H&E to confirm accuracy. The registered MPM images were then exported in their original 16-bit image format. The final result is the multichannel MPM image being registered to a scaled version of the H&E image that defines the gene expression vector coordinates. The coordinates can be scaled by a factor of 3 (due to the rescaling during registration) to locate the center of the field of view for a region of interest corresponding to a given gene expression vector (Fig. 3).

2.5 Statistical Analysis and Clustering

Gene expression data were imported into a Python environment as a data frame containing the unique molecular identifier counts for each human mRNA target and the pixel coordinates for the center of each barcode region. The pixel coordinates (defined in the space of the down-sampled H&E image) were then scaled by a factor of 3 to match the space of the registered MPM image. Regions of interest (ROIs) of each MPM image corresponding to the barcode pixel coordinates were extracted. ROIs of 70×70 pixels were extracted to slightly oversample the barcode active region of 55 microns. A threshold was applied to the MPM ROIs to remove background (<2000) and saturated pixels ($>65,535$) from any subsequent analysis. Abundance was measured by taking the mean value of each ROI corresponding to each channel. Channels 1, 2, 3, and 5 were normalized by dividing by the fourth channel to account for sample-to-sample fluctuations in light source intensity, working distance, and other systematic effects.

Tissue phenotypes were established using *k*-means clustering of the transcriptomic signatures. Clustering of the spatial transcriptomic data was completed using the Loupe browser. *K*-means clustering was applied with $k = 6$, and all four patient samples were pooled together. The selection of $k = 6$ was chosen by assessing the Silhouette score and evaluating the spatial

heterogeneity of cluster assignment, as shown in Sec. 3.1. Cell clusters were assigned tissue types and phenotypes by mapping gene expression profiles to the Human Protein Atlas and assessing the top 50 differentially expressed genes for each cluster, as discussed in Sec. 3.1. Comparisons of MPM channels were performed for each cluster within a specific specimen, as well as for all tumor clusters separately. Data distributions are visualized using violin plots. The Shapiro–Wilk test was used to evaluate normality. All distributions were found to be significantly nonnormal, and therefore, statistical significance was evaluated using the Kruskal–Wallis test followed by post-hoc Mann–Whitney U -test with multiple testing comparison using the Bonferroni correction method by multiplying the p -value by 60. For these comparisons, each ROI/barcode region is treated as a separate independent sample, which is in alignment with statistical analysis approaches in multi-omic studies using spatial transcriptomics.³¹

2.6 Image Classification

Three different classification approaches were developed to test the ability to perform optical phenotyping (Fig. 4). We investigated these three different paradigms to understand what features within the image are most informative for predicting the tissue phenotype.

For the first two models, we utilized classical machine learning with handcrafted features and a support vector machine. Our first model used only the average brightness of each channel, which is representative of the abundance of different contrast sources, yielding a total of five features. Our second model was trained with the original abundance features as well as features extracted using Haralick texture feature extraction,³² which quantifies the spatial relationship between adjacent pixel values, yielding a total of 70 features. Evaluating the spatial distribution of brightness is a measure of how the local microenvironment may be altered in the organization of different contrast sources, which is representative of the tumor microenvironment. For our third model, we utilized a convolutional neural network to perform

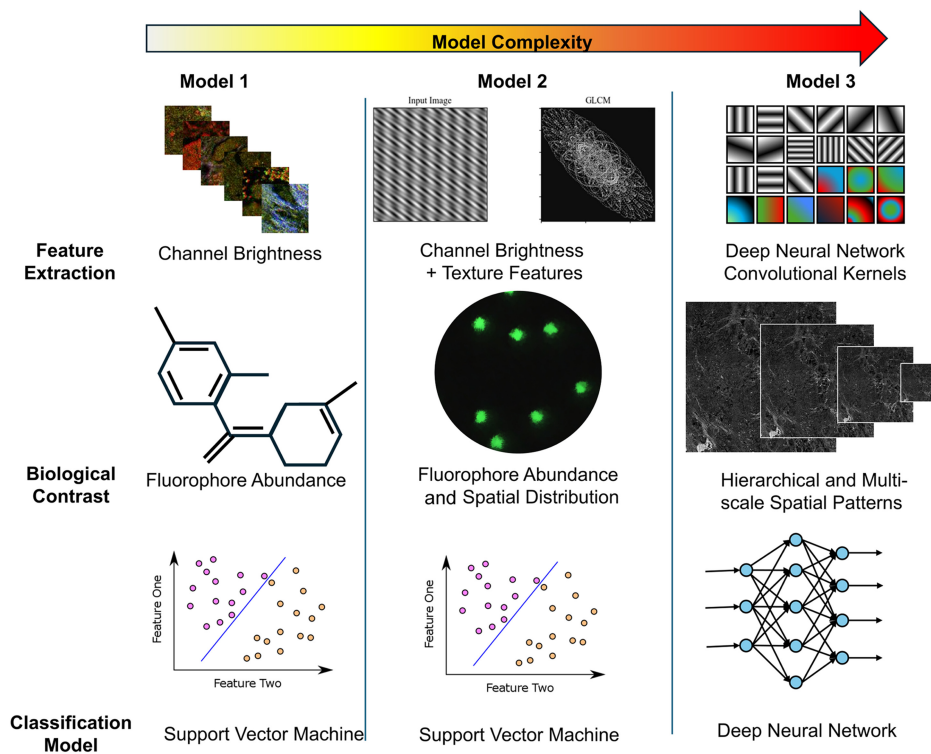


Fig. 4 Overview of classification models evaluated in this study for optical phenotyping. Three models were assessed with increasing complexity. First, we tested support vector machine classifiers with the average brightness of each channel used as input, followed by average brightness and image texture. Then, we trained a deep learning model based on the VGG16 architecture.

unsupervised feature extraction. Here, the feature complexity increased further at the cost of explainability. However, this model can extract information at multiple spatial scales and incorporates nonlinearities, which are likely necessary for modeling the complex sequence-function relationship in tissue.

For the first two models, the feature dataset was first standardized by centering at zero and scaling to unit variance for each channel. The support vector classifiers were trained with 80% of the dataset, and the remaining 20% was used for testing. Five-fold cross-validation with optimized weights was used for assessment for both scenarios, and all assessment metrics were averaged over all five folds.

For our third model, a customized convolutional neural network was designed to achieve high-accuracy multiclass image classification. Here, we applied transfer learning using a pre-trained VGG16,³³ an effective approach when limited labeled data is available, as it leverages features learned by the pretrained model. Figure 5(a) shows the customized VGG16 model architecture. The first four layers of the model were frozen to retain general feature representations extracted from the ImageNet dataset, whereas layers beyond the first maxpooling layer remained trainable to learn features specific to our MPM dataset. The batch size was set to 16. In addition, a dropout layer was incorporated to mitigate potential overfitting, and an early stopping monitor on the validation loss with patience of 6 epochs, along with a learning rate scheduler, was applied as part of the model's callbacks to optimize training.

Before the data were fed into the model, the four co-registered whole-slide MPM images were first converted into three-channel RGB images to ensure compatibility with the network model, as shown in Fig. 5(b). In this conversion, the red channel represented the NADH-dominant signal, the green channel corresponded to the lipofuscin-dominant signal, and the blue channel was assigned to the SHG signal. Based on gene expression data from each region of interest, tables were generated containing the coordinates of each region's center along with their corresponding cluster labels. These colored MPM whole slide images (WSIs) were then segmented into individual tiles using these tables. To effectively cover the entire barcode region, the ideal tile size for the MPM image was set to $70 \times 70 \times 3$, as shown in Fig. 5(b). However, to match the model's required input dimensions, each tile was resized to $224 \times 224 \times 3$, incorporating additional surrounding areas of each barcode region. This re-sizing resulted in $\sim 50\%$

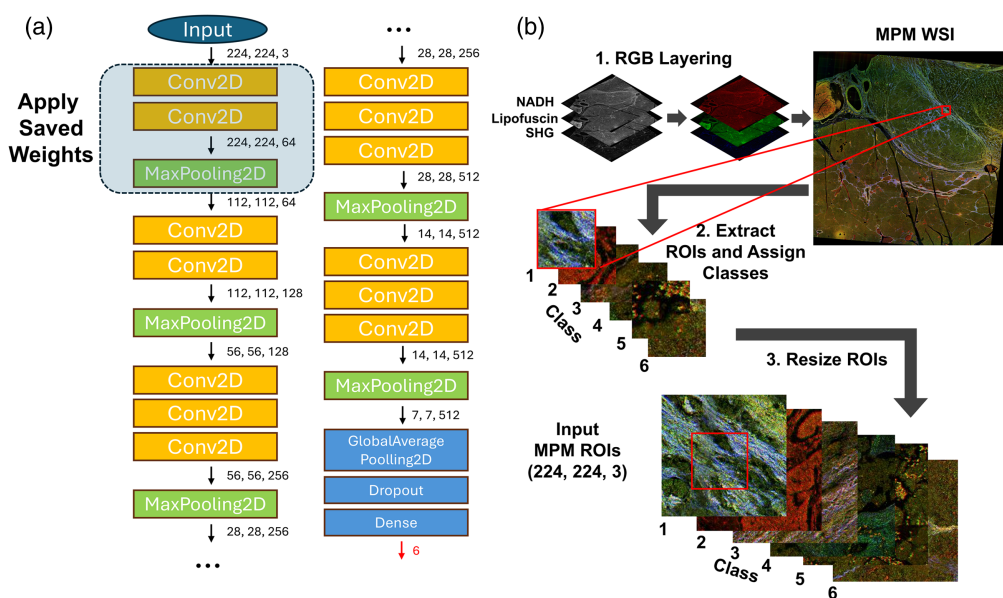


Fig. 5 (a) Diagram of the customized pre-trained VGG16 architecture. The numbers next to the arrows represent the output shape of the preceding layer. The layers enclosed by the dashed line were frozen to preserve the pre-trained weights from the VGG16 model, originally trained on the ImageNet dataset. (b) The data preprocessing flowchart shows the RGB layering, tile extraction, and resizing procedures for the input tile images.

overlapping with adjacent regions of interest. As neighboring areas generally share similar gene expression data due to high spatial correlation, this overlap did not introduce significant ambiguity to the model.

All images were then organized into separate folders, with 80% allocated for training, 10% for validation, and the remaining 10% reserved for testing to assess the model's generalization. In addition, subfolders were created to categorize images into different clusters. The “image dataset from directory” function from the utils library was used to efficiently load and preprocess the input images from the custom folder structure. All methods were implemented using Python, including scikit-learn, the TensorFlow deep learning framework, and the Keras library on a PC equipped with an NVIDIA GeForce A4500 GPU. For all three classifiers, the receiver operator characteristic curves (ROCs) were generated for each class by evaluating the ability to classify each group in a “one versus rest” paradigm. Confusion matrices were also generated to evaluate misclassification for each group in a true multiclass classification paradigm. Classification reports were generated that detail the precision, recall, and *F1* score, as is standard for evaluating classification models. Together, these metrics provided a comprehensive assessment of the model's performance. For this model, we utilized 10-fold cross-validation to ensure all metrics reflected testing on the entire dataset. Visualization of the model's classification performance was also visually evaluated by back-projecting incorrect classifications onto the location of the ROI on the original MPM images. All data and code, including required libraries, used for analysis can be found in the link provided in the Code and Data Availability Statement.

3 Results

3.1 Clustering and Definition of Phenotypes

Clusters were visualized by projecting each cluster onto the respective barcode region and superimposing this on an H&E brightfield image to visualize the clusters and their spatial relationship (Fig. 6). Table 2 shows the results of K-means clustering with $k = 6$ using the full transcriptomic data. Figure 7 shows the top 50 up- and down-regulated genes for each cluster. Of the six clusters, clusters 4 and 5 show the most similarity in gene expression, which is not surprising given that both are represented adjacent to one another in sample 2. Table 3 lists key up-regulated genes for each cluster that are used for establishing tissue type and tumor phenotypes. Pathology reports

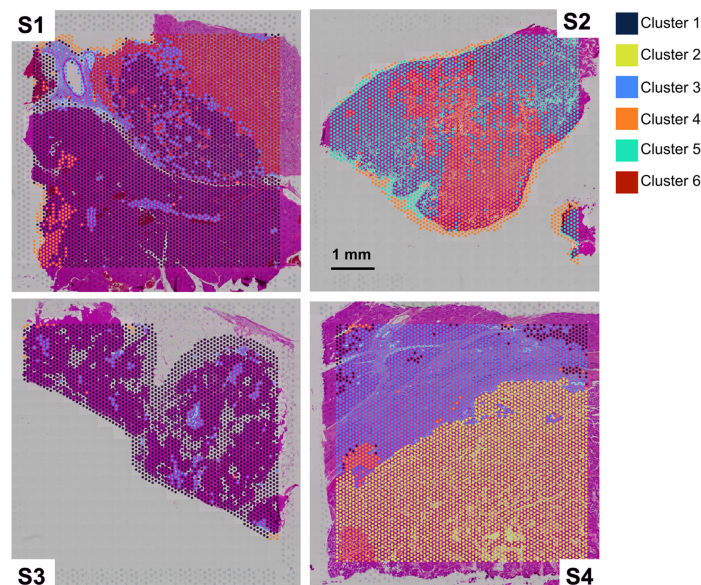


Fig. 6 K-means clustering with $k = 6$ performed for spatial transcriptomic data with all four patients pooled together. The clusters are superimposed on H&E images with each cluster positioned in the respective spatial barcode region and color coded depending on the cluster.

Table 2 Results of *K*-mean clustering with *k* = 6 using transcriptomic signatures. Numbers represent the total number of barcode regions assigned to each cluster across all samples.

	Total	Sample 1	Sample 2	Sample 3	Sample 4
Cluster 1	5748	3122	0	2452	174
Cluster 2	2851	0	0	0	2851
Cluster 3	2510	542	0	172	1796
Cluster 4	1789	204	1396	18	171
Cluster 5	1787	0	1787	0	0
Cluster 6	993	993	0	0	0

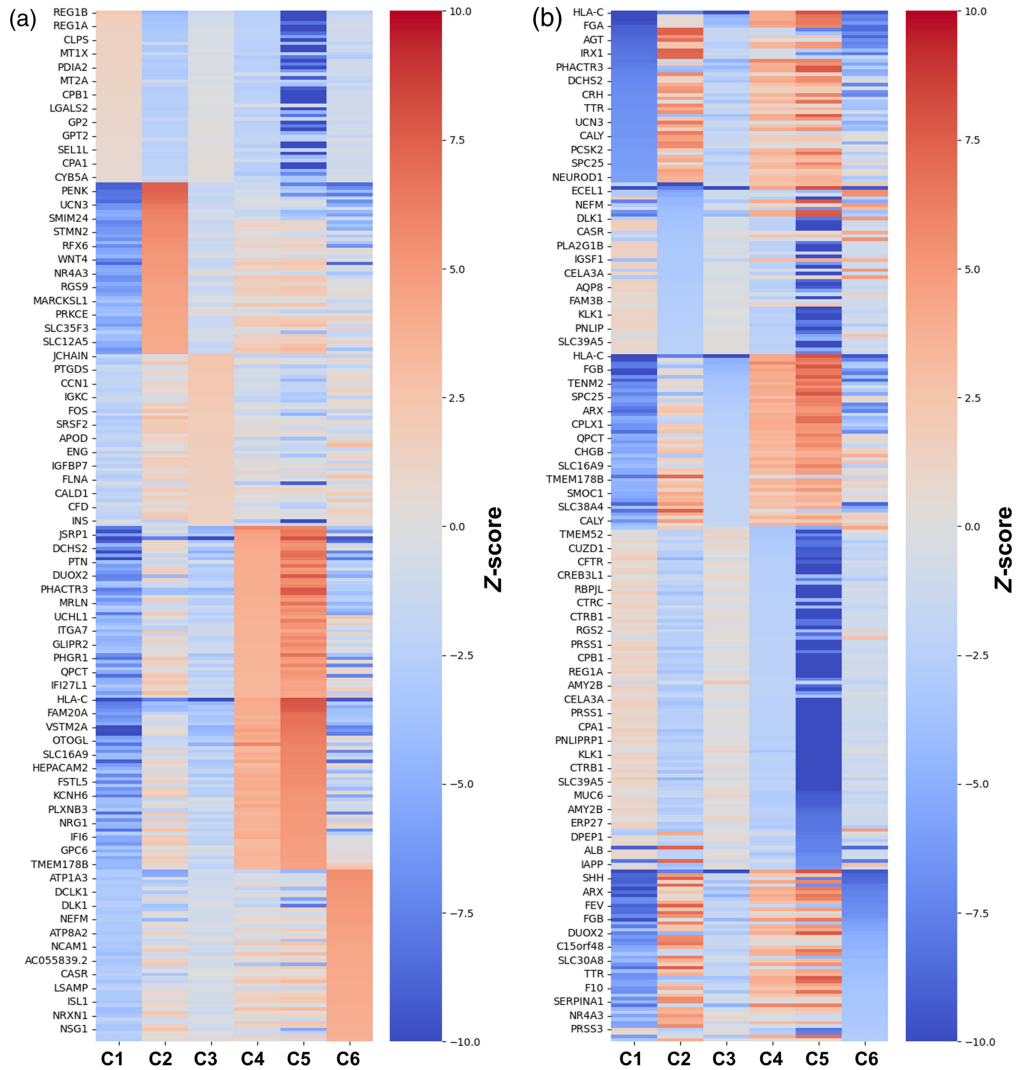


Fig. 7 Gene expression heatmaps of the (a) top 50 over-expressed and (b) top 50 under-expressed genes for each cluster. Heatmap colors are denoted as the z-score comparing the expression levels of each gene for each cluster among the entire dataset.

Table 3 Interpreted description of each cluster and top differentially expressed genes (DEGs) for each used to establish phenotypes. The second row shows the label used to refer to each cluster following Sec. 3.1.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Description	Pancreatic exocrine, acinar cells	Classical NET, pancreatic alpha cell lineage	Stroma	Tumor with immune, glial, and neuronal markers	Cluster 4 with added stromal enrichment	Tumor enriched with neuronal, entero-endocrine markers
Label	Exocrine	T1	Stroma	T2 (Core)	T2 (Rim)	T3
	<i>REG1A</i>	<i>IRX1</i>	<i>JCHAIN</i>	<i>UCHL1</i>	<i>PCSK2</i>	<i>NEFL</i>
	<i>REG1B</i>	<i>IRX2</i>	<i>IGHA1</i>	<i>SKAP1</i>	<i>ENPP2</i>	<i>NEFM</i>
	<i>CLPS</i>	<i>GCG</i>	<i>IGKC</i>	<i>ITGA7</i>	<i>NSG2</i>	<i>AMPH</i>
Top DEGs for defining tumor phenotypes and tissue types	<i>CELA2B</i>	<i>GHRL</i>	<i>HLA-DRA</i>		<i>NRG1</i>	<i>ECEL1</i>
	<i>CELA3A</i>	<i>CHGA</i>	<i>COL3A1</i>		<i>GLIPR2</i>	<i>NRXN1</i>
	<i>PRSS3</i>	<i>CHGB</i>	<i>COL1A2</i>		<i>HLA-C</i>	<i>AKAP12</i>
	<i>PRSS1</i>	<i>ENO2</i>	<i>COL1A1</i>		<i>IFI6</i>	<i>PCSK1</i>
	<i>PNLIPRP1</i>	<i>SHH</i>	<i>MYL9</i>		<i>ICAM2</i>	<i>HAP1</i>
	<i>PNLIP</i>	<i>SYP</i>	<i>DCN</i>		<i>FGA/FGB</i>	<i>DLK1</i>

were also evaluated for expression of specific hormones, including glucagon and gastrin, as well as Ki-67 proliferative index.

Cluster 1 is characterized by an enrichment in genes associated with acinar cells and the exocrine pancreas, including *REG1A/B*, *CLPS* and *CELA2B/3A* and *PNLIP*. This gene signature aligns with the expectation from pathological examination of the tissue.

Cluster 2 shows an enrichment in genes that are most classically associated with neuroendocrine tumors (e.g., *CHGA*, *CHGB*, *SYP*). Significantly upregulated genes in this cluster include *IRX1*, *IRX2*, and *GCG*, which are markers of the pancreatic alpha cell lineage. This is also consistent with the pathology report, showing positive tumor expression of glucagon. Interestingly, alterations of *CHGA* and *CHGB* at the protein level have been associated with different phenotypes in pancreatic neuroendocrine tumors; in this case, relating to aggressiveness.³⁴ Tumor cells in this cluster also express the neuroendocrine tumor markers *SYP* and *SHH*, a key developmental morphogen with a potential role in NET development.³⁵ Therefore, cluster 2 is defined as a classical neuroendocrine tumor.

Cluster 3 exhibits upregulated expression of stromal and immune-related genes. This includes an enrichment of plasma cell markers such as *JCHAIN*, *IGHA1*, and *IGKC*, immune cell markers including *HLA-DRA* and *COL3A1*, as well as markers of extracellular matrix and fibroblasts including *COL1A2*, *COL1A1*, *MYL9* (myofibroblasts and smooth muscle), and *DCN* (endothelial cells). These markers align with the H&E image and pathological examination of the tissues. Thus, this cluster is defined broadly as stromal tissue.

Clusters 4 and 5 occur within a single patient tumor and have similar transcriptomic profiles. However, clustering the dataset with $n = 5$ clusters did not merge the two clusters, suggesting important differences between the two from the transcriptomic standpoint. Both clusters exhibit downregulated *PRSS1* expression, which is associated with hereditary pancreatitis,³⁶ as well as *NRG1*, *GLIPR2*, and *HLA-C*, which are immune response-related markers, suggesting that the tumor may exhibit unique immune or inflammatory characteristics. Other shared features include expression of *NRG1* and *GLIPR2*, markers of glial cells, as well as *FGA* and *FGB*, which are related to the blood clotting response. Important differences between clusters 4 and 5 are the expression of *ENPP2* in cluster 5 (a stromal marker) in addition to a seemingly lower expression of genes in cluster 4 universally. This may suggest that cluster 4 represents some loss of gene expression, perhaps due to necrosis. As such, we define clusters 4 and 5 as a tumor with immune

features, with 5 being denoted “rim” and 4 being noted “core,” where some necrosis may be present. Clinically, this was the only specimen that was reported focally positive for gastrin expression, which is often related to a specific subtype of pancreatic NET.³⁷ Furthermore, the tumor specimen relating to clusters 4 and 5 yielded a clinical Ki-67 index of up to 8.2%, which was double the corresponding index of the specimens associated with clusters 2 and 6 (both of which were measured around 4%).

Finally, cluster 6 exhibits enrichment of enteroendocrine markers (*PCSK1*, *HAPI*, *DLS1*), as well as neuronal markers, including the expression of *NEFL*, *NEFM*, and *NCAM1*. Distinct differences are observed between clusters 6 and 2, and notably, this cluster is associated with a tumor that was reported as glucagon negative. In addition, there are a number of prognostic markers such as over-expression of *SYT17* and under-expression of *SHH*, which in other pancreatic cancers at the protein level have been associated with more favorable prognosis.³⁸ As such, we define cluster 6 as a tumor with enteroendocrine and neural features. Referencing clinical data, the patient in which this specimen was observed was the only patient who possessed lymph node metastases—which is highly correlated with poor survival.³⁹ For the remainder of this paper, we will refer to the different clusters by their label shown in Table 3 to assist with interpretation.

3.1.1 Statistical analysis of image features

Figure 8 shows the correlation between imaging channels before and after normalization. Prior to normalization by the porphyrin channel, interchannel correlation was high, over 0.65 on average. The SHG is lower in general, which is expected given that the optical process is not a fluorescence process. After correlation, the correlation is significantly reduced, being below 0.3 on average for all pairs of channels.

Figure 9 shows the violin plot comparing the normalized average normalized brightness of each channel for ROIs assigned to different clusters. Significance is denoted above each cluster as *** for $p < 0.001$ for all pair-wise comparisons except the pairs denoted with a line showing “n.s.” for not significant. Several observations can be made. First, each cluster exhibits a unique distribution across the four channels. For example, although channel 1 distribution is similar for the first three clusters (exocrine, T1, and stroma), the distributions differ significantly in channels 2 and 3. This suggests that the multichannel information is important for successfully differentiating tissue types. Among all clusters, the two clusters from T2—the rim and the core—show the most similar distributions, which is supported by the lack of statistical significance in two of the channels. There are subtle differences between the two, for example, a proportion of the T2 core cluster is contained in a lower tail of the distribution (indicating lower brightness, potentially related to necrosis). Given that both clusters are closely related transcriptomically, the similarity of the imaging characteristics is not surprising. This also indicates that the two clusters may be difficult to differentiate during classification. Considering this data overall, although most

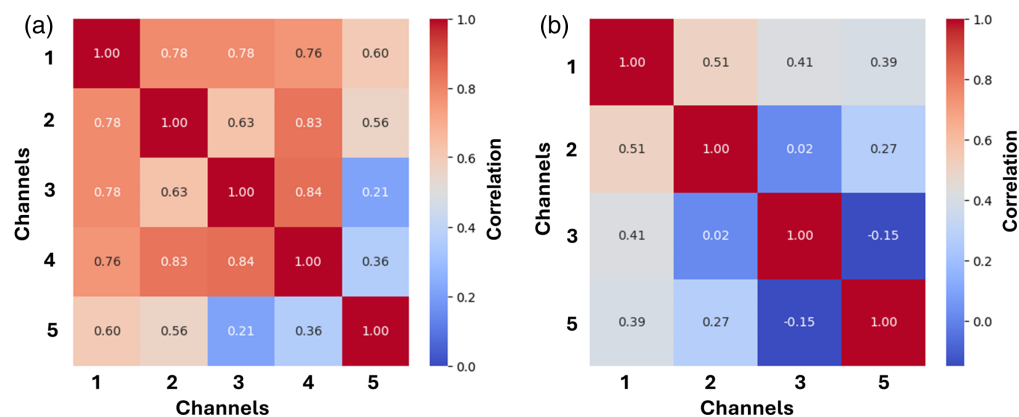


Fig. 8 Inter-channel correlation of MPM imaging channels (a) before and (b) after normalization by channel 4.

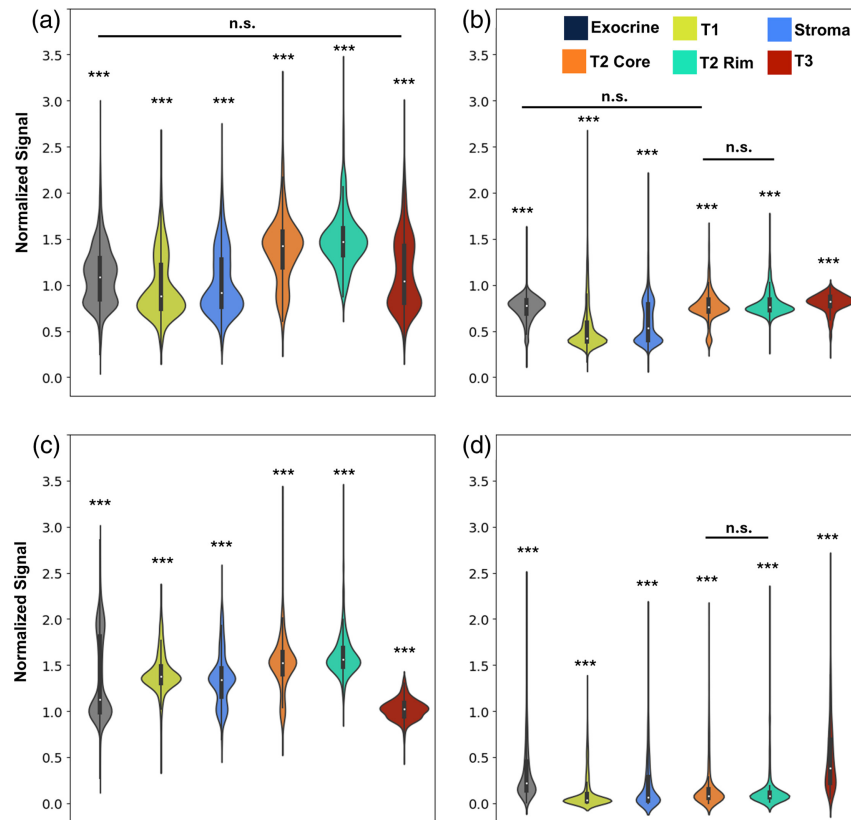


Fig. 9 Violin plots comparing average brightness for ROIs assigned to different tumor clusters for normalized (a) channel 1, (b) channel 2, (c) channel 3, and (d) channel 5. Significance is denoted *** for $p < 0.001$ for all pair-wise comparisons. Bars denoting no significant difference are used to show pairs that are not significantly different.

clusters have clear differences in the distributions among one another (supported by the strong statistical significance), there are still significant overlaps in the image brightness distributions. The consequence of this is that there likely will be significant challenges in classifying clusters based on this information alone, motivating the need for more advanced image features such as texture and other spatially derived features.

3.2 Image Classification

The classification results for each of the three models are shown in Tables 4–6, which present the recall, precision, and $F1$ score for each class and classifier, respectively, averaged over all

Table 4 Precision, recall, and $F1$ score for support vector classifier trained on abundance, averaged across all fivefolds.

Class	Precision	Recall	$F1$ score	AUC (one versus rest)
Exocrine	0.71	0.42	0.53	0.77
T1	0.53	0.76	0.62	0.82
Stroma	0.10	0.00	0.00	0.67
T2 core	0.09	0.03	0.04	0.77
T2 rim	0.35	0.90	0.50	0.88
T3	0.21	0.62	0.31	0.86

Table 5 Precision, recall, and $F1$ score for the support vector classifier trained on abundance and image texture features, averaged across all five folds.

Class	Precision	Recall	$F1$ score	AUC (one versus rest)
Exocrine	0.78	0.74	0.76	0.91
T1	0.57	0.45	0.51	0.92
Stroma	0.36	0.39	0.37	0.76
T2 core	0.37	0.21	0.26	0.81
T2 rim	0.52	0.67	0.58	0.93
T3	0.22	0.41	0.28	0.86

Table 6 Precision, recall, and $F1$ score for the deep neural network model averaged across all 10-folds.

Class	Precision	Recall	$F1$ score	AUC (one versus rest)
Exocrine	0.92	0.94	0.93	0.99
T1	0.98	0.99	0.98	1.00
Stroma	0.89	0.82	0.85	0.98
T2 core	0.79	0.71	0.74	0.97
T2 rim	0.84	0.86	0.84	0.99
T3	0.85	0.91	0.88	1.00

five folds. Included in these tables is the AUC for the one-versus-rest classifiers shown later. For the two support vector classifiers, the overall performance is marginal. The classifier trained on abundance only (Table 4) is unable to classify the classes corresponding to Stroma and T2 Core in a true multiclass paradigm—this most likely is due to a lack of sufficient features in the model, as only five channels are present, whereas six classes exist. The total weighted average precision of the classifier was $43.8\% \pm 0.83\%$.

For the support vector classifier trained on image texture and abundance (Table 5), there is moderate improvement. The average weighted classification accuracy increases to $54.4\% \pm 0.54\%$ overall, and performance on a class-by-class basis increases nearly universally, with the exception of a slight decrease in performance for T1. Classes corresponding to T2 core and T3 remain the most challenging to classify accurately.

In stark contrast to the support vector classifiers with handcrafted features, using deep learning for feature extraction and classification, shown in Table 6, yields significant improvements, with an overall weighted classification accuracy of $89.4\% \pm 1.5\%$. The misclassification of T2 core is the primary factor affecting overall prediction accuracy, which also performed poorly for the two prior classifiers but improved significantly in this scenario.

Visualizations of ROC curves and confusion matrices for all classes are shown for each classifier in Fig. 10. Figures 10(a)–10(c) present the one-versus-rest ROC curves for all classes, highlighting the varying classification performance across different categories. As observed in the plot and tables, the AUC values are consistently higher than the $F1$ scores across all classes, as AUC evaluates the model's ability to distinguish between classes across a range of decision thresholds. Even if a suboptimal decision threshold results in classification errors and lower $F1$ scores, the model may still exhibit strong discriminative capability over the entire probability distribution. The ROC curve from the deep neural network model exhibits a sharp ascent toward

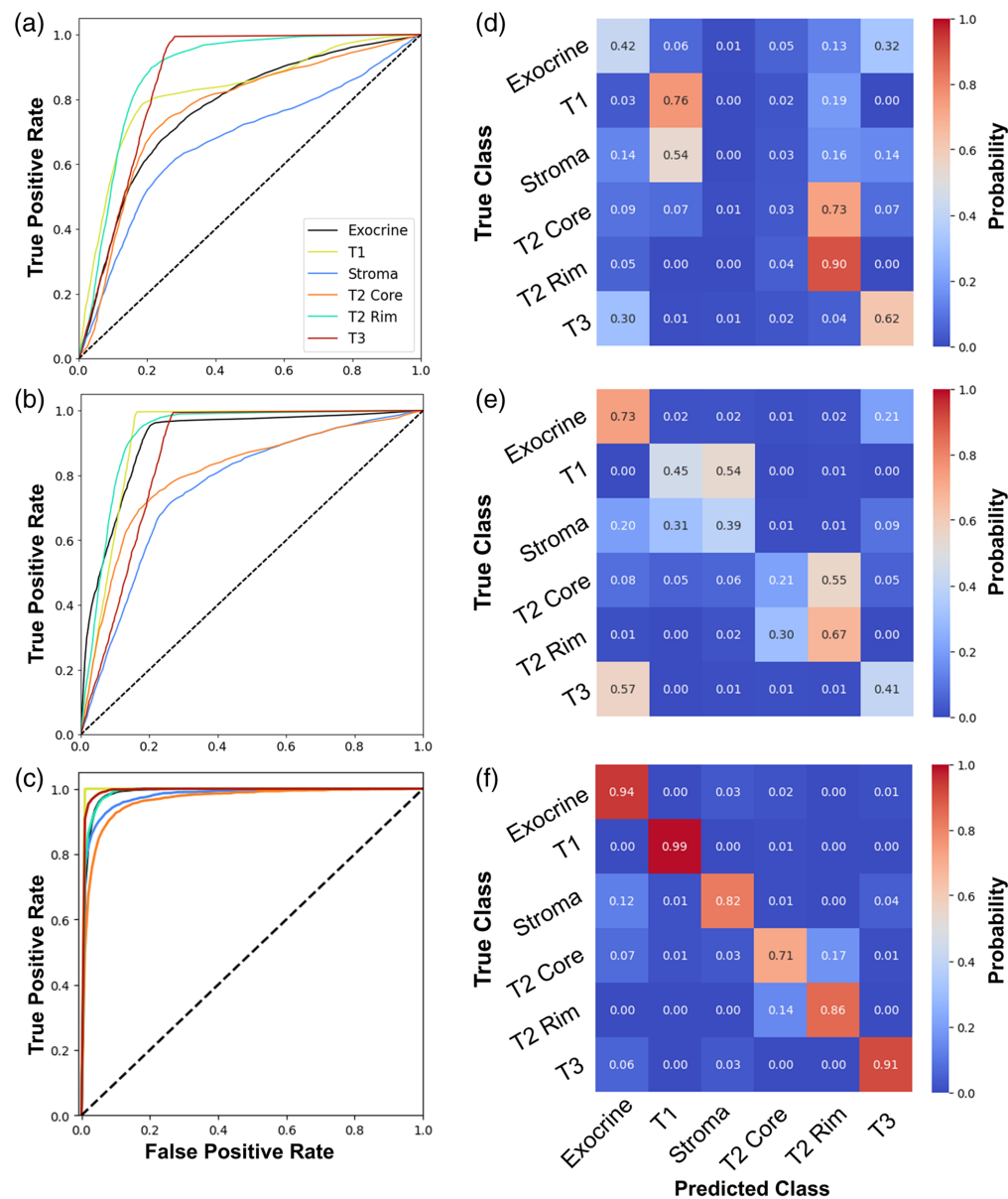


Fig. 10 Classification model results. ROC curves to assess one-versus-rest classification performance for (a) support vector machine with abundance features, (b) support vector machine with abundance and texture, and (c) deep neural network. Confusion matrices for multiclass classification for (d) support vector machine with abundance features, (e) support vector machine with abundance and texture, and (f) deep neural network.

the top-left corner, particularly for the T1 and T3 classes. This behavior reflects a strong balance between sensitivity and specificity, effectively minimizing false positives while maintaining a high true positive rate. For the simpler models using handcrafted features, the AUC is relatively high for a subset of classes (e.g., T2 Rim and T3) but relatively low for other classes (stroma, T2 core).

The issue of misclassification is most clearly illustrated in the confusion matrices [Figs. 10(d)–10(f)]. For the two support vector classifiers, the models consistently misclassify T1 and stroma, as well as T2 rim and core. The latter is not surprising given the earlier observations of the similarities between those two clusters both transcriptomically and optically. For the deep learning model, the ability to differentiate all classes is drastically improved; however, misclassification between exocrine and stroma persists, along with a notable number of misclassifications between the T2 core and rim. The latter two classes were also commonly misclassified

in the support vector approach. This aligns well with our discussion of gene expression and phenotype clustering as these two classes share similar transcriptomic profiles and are both highly represented in a single tissue specimen. The misclassification of stroma and exocrine is likely due to the fact that the ROIs contain a multitude of cell types and therefore are not “pure.” Some may contain both exocrine and stromal cells, resulting in some reduction in image classification performance.

Figure 11 presents a back-projection of all misclassified ROIs onto MPM WSI using the deep learning model from the test dataset across the 10-fold models, providing a detailed visualization of recurring misclassification patterns. By mapping these instances, we can identify spatial and structural trends contributing to classification errors, revealing potential model biases and overlapping feature distributions among classes. As illustrated in the legend patch on the third MPM WSI [Fig. 11(c)], the inner circle color of each spot represents the predicted class, whereas the outer circle indicates the true class. Across the four WSIs, the majority of mixed-color groups correspond to blue-black (exocrine and stroma) and cyan–orange (T2 rim and core), aligning with our previous analysis.

Notably, the second MPM WSI represents an all-tumor slide, containing both T2 core and T2 rim classes. In Fig. 11(b), most misclassified images cluster within the T2 core region, with

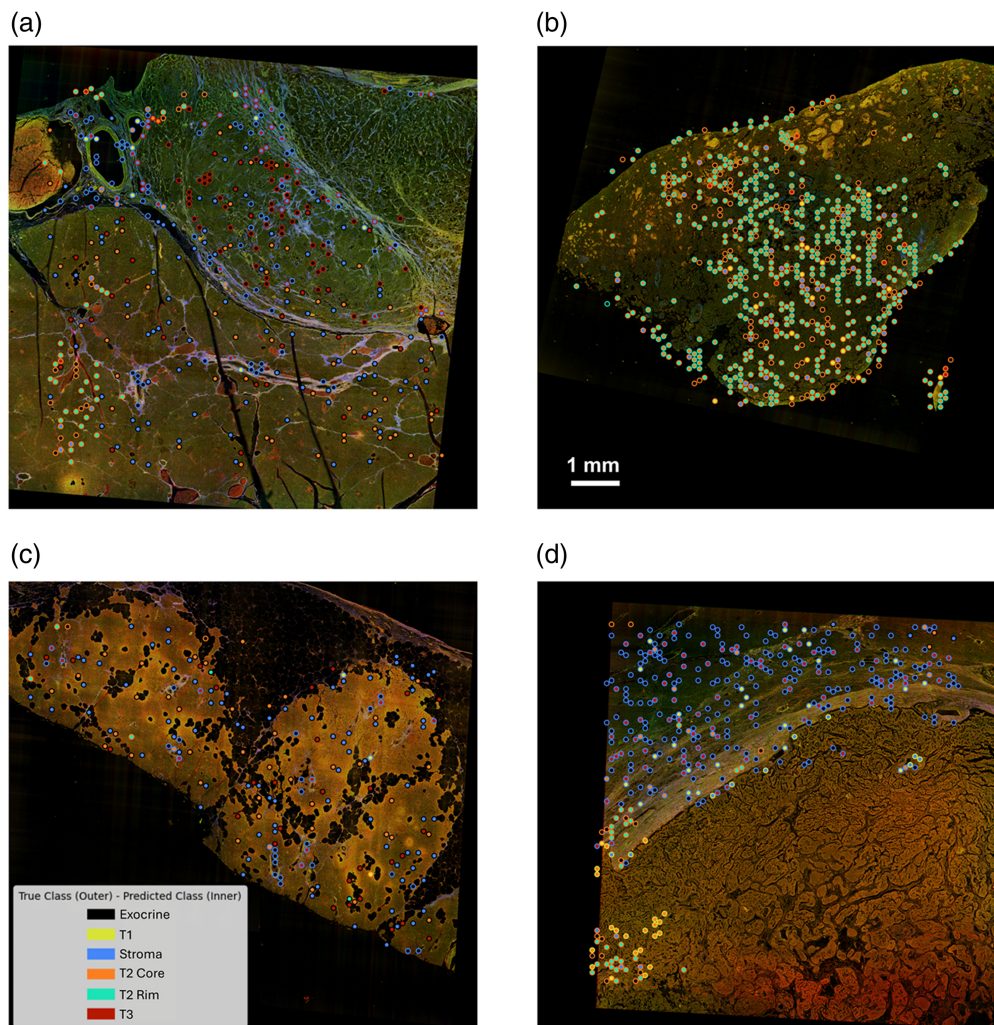


Fig. 11 Back projection of all misclassified ROIs from the test dataset across the 10-fold models onto the four whole-slide images corresponding to our patient specimens - (a) S1, (b) S2, (c) S3, (d) S4 - provides a comprehensive visualization of recurrent misclassification patterns. The color of the center of the circle represents the predicted class, whereas the color of the rim represents the true class.

errors predominantly occurring at the interface between two clusters. This misclassification may partially stem from the expanded input image size, where adjacent tiles share overlapping regions, potentially introducing bias, particularly at class boundaries. However, given the limited dataset, the benefits of this expansion outweigh the loss introduced by overlapping biases. Furthermore, T1 [Fig. 11(d), lower right region] and T3 [Fig. 11(a), upper right region] exhibit an exceptionally high true positive rate, with nearly no false predictions. There are also some misclassified instances that lie outside the WSI region, indicating imperfections in data preprocessing that may introduce additional bias.

These observations suggest potential improvements to enhance model performance. Refining the data preprocessing pipeline by implementing a more precise tile screening method would help accurately exclude dark images and irrelevant tiles, reducing unnecessary biases. In addition, once more data become available, limiting the overlap between adjacent tiles could minimize misclassification at class boundaries. Another possible refinement involves developing specialized sub-models for exocrine versus stroma and T2 core versus rim, which may improve classification accuracy for these challenging class pairs. By addressing these factors, the model's robustness and predictive performance can be further improved.

Figure 12 presents complementary plots to Fig. 11, showing all the correct predictions back-projected onto the four WSIs. These two figures map all the prediction results back onto the MPM WSIs, showing consistency with the spatial transcriptomics profiles. Ultimately, the strong

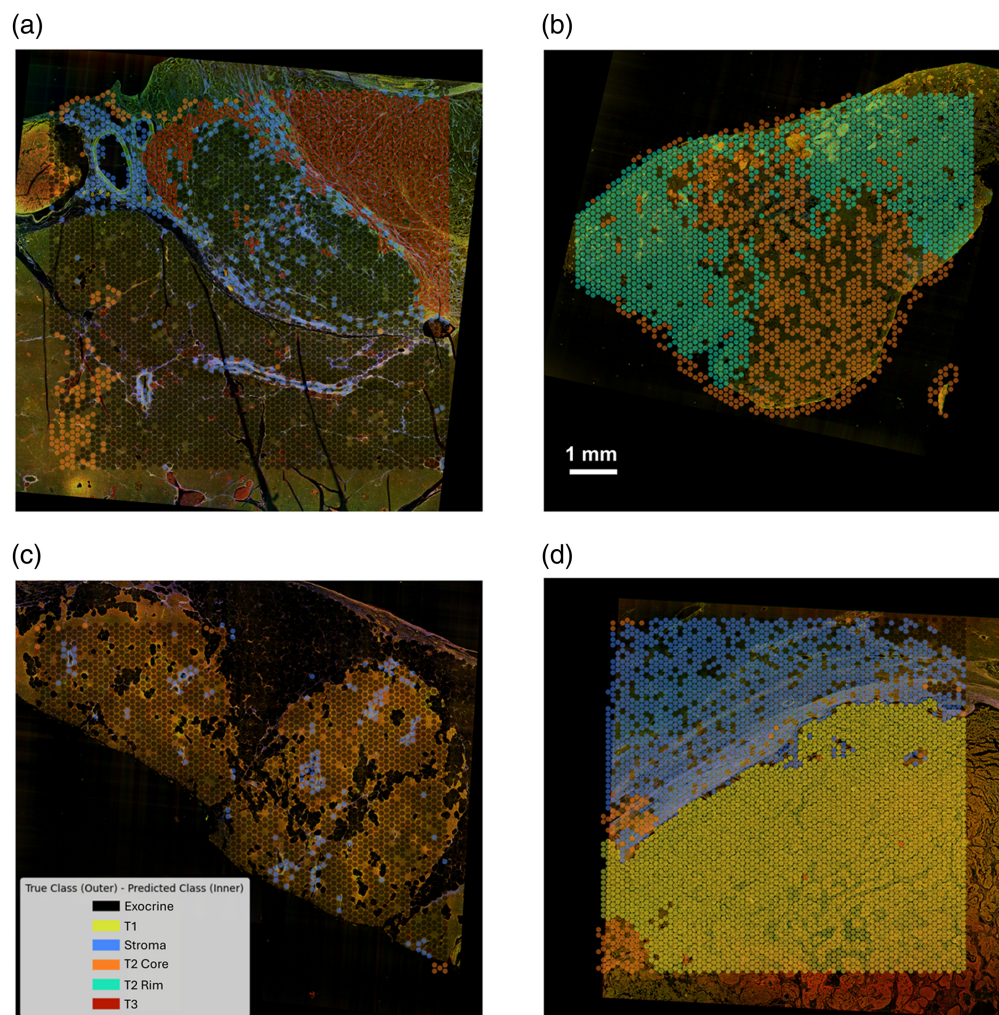


Fig. 12 Back projection of all correctly predicted ROIs from the test dataset across the 10-fold models onto the four whole-slide images corresponding to our patient specimens - (a) S1, (b) S2, (c) S3, (d) S4.

performance on the ROC curves, coupled with the high overall accuracy (89.4%) on the test dataset, further validates our deep learning model's predictive capability and reinforces the potential of integrating label-free microscopy and spatial transcriptomics as a powerful tool for optical phenotyping.

4 Discussion

The primary goal of this work was to develop a method for optical phenotyping of tissues using label-free microscopy. The results of this work demonstrate several important findings. First, label-free optical imaging contrast, in this case related to autofluorescence and second harmonic generation, carries significant biologically relevant information that can be used for optical phenotyping. The evaluation of different feature extraction methods and classification approaches supports our hypothesis that multiscale hierarchical features are necessary to more comprehensively capture the tissue microenvironment and enable optical phenotyping.

Despite our promising findings, there are limitations that must be acknowledged and addressed in future work. First, this work is carried out using *ex vivo* FFPE specimens, and the fixation process is known to alter tissue autofluorescence. This limits our ability to draw specific conclusions about what fluorescence species are varying in abundance as there is significant disagreement in the literature regarding what fluorescence species are preserved.^{40,41} In addition, although the approach of using label-free imaging for *in vivo* optical phenotyping is theoretically possible, all findings obtained using fixed tissues will need to be validated and tested in fresh/*in vivo* tissues given the differences in optical properties between the two. Likely, the features learned by the DNN using fresh tissues will differ significantly from FFPE, indicating that a different model would need to be developed. Another outstanding question is to understand which label-free contrast mechanisms are most effective for phenotyping and/or estimating other transcriptomic features. We utilize MPM in this study primarily to elicit the SHG signal, but it may be possible to achieve similar performance using simpler systems (e.g., single-photon autofluorescence) or measuring different optical properties.

Similarly, our dataset was limited to only four specimens due to both the cost and availability of specimens that yielded sufficient RNA quality. Although the number of barcode regions numbered close to 15,000 total, the small sample size in terms of independent patients required us to segment the dataset, treating each of the barcode regions as independent. Although this is common in multi-omic studies, a more rigorous approach with a larger dataset and more patient specimens is required to fully evaluate the statistical properties and generalizability of our classification models.

Another important discussion is the clustering approach to define our phenotypes. The approach presented in this study involved a completely unbiased and unsupervised method to reduce any subjective interpretation and selection of phenotypes. Although powerful for maintaining reproducibility and rigor, the approach has some inherent limitations. The mathematical assignment of clusters using *k*-means clustering does not necessarily integrate specific biological information such as spatial correlation of the signatures, and weighting of different genes that may have greater significance. This may lead to some tissues being clustered incorrectly due to numerical noise. Indeed, the clustering of our specimens seemed to suggest there was a small degree of this; for example, cluster 4 appearing in the normal tissue region of specimen S1 (see Fig. 6). It is also important to recognize that there are many more than 6 cell types present overall. The spatial sampling for transcriptomics used in this study was limited to spacing the sampling points by 100 microns; thus, within each cluster, there could be a variety of cell populations contributing to the overall transcriptomic signature, which will result in a wider distribution within each cluster. Single-cell sequencing would limit this effect but would also lose the surrounding microenvironment information given by the larger field of view of our approach. One of the most significant outcomes of this work is that our results suggest that the local environment is critical to effectively performing optical phenotyping. Furthermore, our phenotyping is focused on tissue-level properties, not cell-level properties; therefore, single-cell sequencing is likely not appropriate for measuring these effects. Recently, the technological platforms for spatial transcriptomics have been advanced to enable higher resolution spatial sequencing over what was used in this study. Integrating these next-generation

technologies would be a valuable way to assess phenotypes with higher precision without losing local environmental characteristics.

Finally, integrating a deep learning approach requires future work investigating the explainability of the model to determine specifically what features are used for performing high-accuracy classification and to evaluate whether these features are generalizable to other tissues and patients. In addition, an interesting avenue of future inquiry is to investigate how the transcriptomic signatures could be analyzed to assess higher-level biological processes such as pathway regulation. This information could be used to inform and validate what image features offer the most predictive value. Along these lines, other label-free imaging modalities such as polarization and fluorescence lifetime imaging could be integrated into this architecture and evaluated for their ability to perform optical phenotyping. Ultimately, the framework we present in this paper has wide potential for addressing a number of intriguing scientific questions and lays the foundation for accelerating discoveries in the field of biophotonics.

5 Conclusion

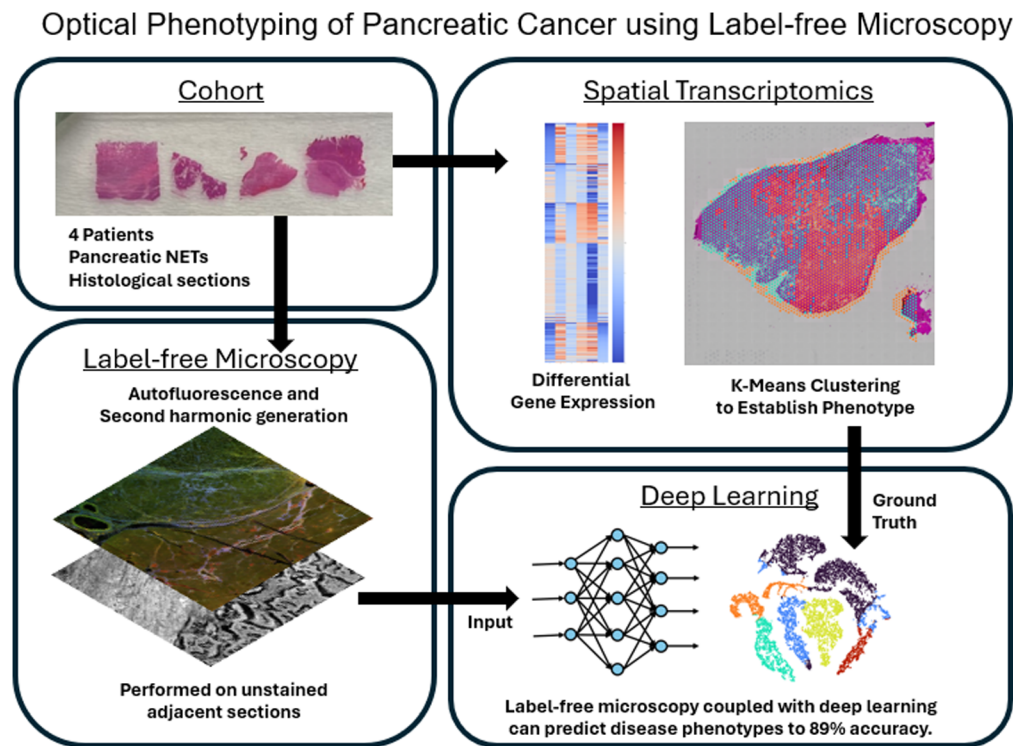
Phenotyping is a powerful tool in diagnostics and for developing precision therapeutics, but most phenotyping methods are costly, time-intensive, and not compatible with *in vivo* use. In this study, we introduce a new approach for “optical phenotyping” using label-free microscopy images, deep learning, and tissue phenotypes as defined by spatial transcriptomics. Our results show significant promise for optical phenotyping. The presented method is able to achieve over 89% accuracy in a complex multiclass classification problem. Although promising, a number of outstanding challenges are to be addressed in future work; namely, expanding the patient cohort, addressing nonbiological heterogeneity that may influence phenotype assignment, as well as eventual testing in fresh tissues.

6 Appendix: Supplemental Materials

This appendix includes additional clinical information for each patient specimen extracted from pathology reports, shown in Table 7. In addition, a graphical abstract of the work is shown in Fig. 13.

Table 7 Available metadata for each patient specimen used in this study. Specimens used for analysis were formalin-fixed paraffin-embedded histological sections.

	S1	S2	S3	S4
Diagnosis	Grade 2 well-differentiated pancreatic NET	Grade 2 well-differentiated pancreatic NET	Grade 1 well-differentiated pancreatic NET	Grade 2 well-differentiated pancreatic NET
Tumor site	Pancreatic tail	Pancreatic body	Pancreatic head	Pancreatic tail
Lymph node involvement	Metastases located on three lymph nodes	All benign	All benign	All benign
Tissue on histological section	Tumor and adjacent normal	Tumor only	Normal only	Tumor and adjacent normal
Chromogranin	+	+	+	N.A.
Synaptophysin	+	+	+	N.A.
Gastrin	N.A.	Focal +	—	—
Glucagon	—	Rare +	+	+
Ki-67	Up to 4.4%	Up to 8.2%	<1%	Up to 4%



Disclosures

The authors declare that there are no financial interests, commercial affiliations, or other potential conflicts of interest that could have influenced the objectivity of this research or the writing of this paper.

Code and Data Availability Statement

All Python code and data associated with this work can be found in the repository at the following: <https://doi.org/10.25422/azu.data.28664747>.

Acknowledgments

We would like to thank Carole Kepler and Alexandra Jimenez and the Tissue Acquisition and Cellular/Molecular Analysis service or their assistance with histological processes, as well as Martin Deymier at the University of Arizona Genetics Core. We would also like to thank the UC Davis Comprehensive Cancer Center for supporting this work. This study was supported by the National Institutes of Health (Grant No. P30 CA023074), the Arizona Department of Health Services (Grant No. RFGA2024-022-005), and the Department of Defense (Grant No. W81XWH2210211).

References

1. R. Wang and Z. Wang, "Precision medicine: disease subtyping and tailored treatment," *Cancers* **15**(15), 3837 (2023).
2. A. Sahu et al., "In vivo tumor immune microenvironment phenotypes correlate with inflammation and vasculature to predict immunotherapy response," *Nat. Commun.* **13**, 5312 (2022).
3. C. Torres and P. Grippo, "Pancreatic cancer subtypes: a roadmap for precision medicine," *Ann. Med.* **50**(4), 277–287 (2018).
4. C. Weng, N. H. Shah, and G. Hripcsak, "Deep phenotyping: embracing complexity and temporality—towards scalability, portability, and interoperability," *J. Biomed. Inf.* **105**, 103433 (2020).
5. P. Harrison, A. Wright, and J. Mank, "The evolution of gene expression and the transcriptome-phenotype relationship," *Semin. Cell Dev. Biol.* **23**(2), 222–229 (2012).

6. N. Shaked et al., “Label-free biomedical optical imaging,” *Nat. Photonics* **17**, 1031–1041 (2023).
7. A. Croce and G. Bottioli, “Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis,” *Eur. J. Histochem.* **58**(4), 2461 (2014).
8. T. Knapp et al., “Quantitative characterization of duodenal gastrinoma autofluorescence using multi-photon microscopy,” *Lasers Surg. Med.* **55**(2), 208–225 (2023).
9. A. T. Shah, T. M. Heaster, and M. C. Skala, “Metabolic imaging of head and neck cancer organoids,” *PLoS One* **12**, e0170415 (2017).
10. J. Bec et al., “Label-free visualization and quantification of biochemical markers of atherosclerotic plaque progression using intravascular fluorescence lifetime,” *JACC Cardiovasc. Imaging* **14**(9), 1832–1842 (2021).
11. N. Ghosh and I. Vitkin, “Tissue polarimetry: concepts, challenges, applications, and outlook,” *J. Biomed. Opt.* **16**(11), 110801 (2011).
12. P. Tang et al., “Polarization sensitive optical coherence tomography with single input for imaging depth-resolved collagen organizations,” *Light Sci. Appl.* **10**(1), 237 (2021).
13. T. Cannon et al., “Measuring collagen injury depth for burn severity determination using polarization sensitive optical coherence tomography,” *Sci. Rep.* **12**(1), 10479 (2022).
14. L. Wang et al., “Hot and cold tumors: immunological features and the therapeutic strategies,” *Med. Comm.* **4**(5), e343 (2020).
15. F. Kai, A. Drain, and V. Weaver, “The extracellular matrix modulates the metastatic journey,” *Dev. Cell* **49**(3), 332–346 (2019).
16. M. Chaplain, *Multiscale Modelling of Cancer: Micro-, Meso- and Macro-Scales of Growth and Spread*, pp. 149–168, Springer International Publishing, Cham (2020).
17. Y. Wu and L. Xie, “Ai-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships,” *Comput. Struct. Biotechnol. J.* **27**, 265–277 (2025).
18. F. Wang and A. Preininger, “AI in health: state of the art, challenges, and future directions,” *Yearb. Med. Inf.* **28**, 16–26 (2025).
19. H. Chan et al., “Deep learning in medical image analysis,” *Adv. Exp. Med. Biol.* **1213**, 3–21 (2020).
20. P. Prasetyanti and J. Medema, “Intra-tumor heterogeneity from a cancer stem cell perspective,” *Mol. Cancer* **16**(1), 41 (2017).
21. M. Haffner et al., “Genomic and phenotypic heterogeneity in prostate cancer,” *Nat. Rev. Urol.* **18**(2), 79–92 (2021).
22. M. Chen et al., “Radiomics and artificial intelligence for precision medicine in lung cancer treatment,” *Semin. Cancer Biol.* **93**, 97113 (2025).
23. C. Haase et al., “Image-seq: spatially resolved single-cell sequencing guided by in situ and in vivo imaging,” *Nat. Methods* **19**, 1622–1633 (2022).
24. Y. Zhang et al., “From genotype to phenotype: Raman spectroscopy and machine learning for label-free single-cell analysis,” *ACS Nano* **18**, 18101–18117 (2024).
25. K. Kobayashi-Kirschvink et al., “Prediction of single-cell RNA expression profiles in live cells by Raman microscopy with Raman2RNA,” *Nat. Biotechnol.* **42**, 1726–1734 (2024).
26. A. Schroeder et al., “The RIN: an RNA integrity number for assigning integrity values to RNA measurements,” *BMC Mol. Biol.* **7**, 3 (2006).
27. T. Matsubara et al., “Dv200 index for assessing RNA integrity in next-generation sequencing,” *Biomed. Res. Int.* **2020**, 9349132 (2020).
28. T. Knapp et al., “Combined flat-field and frequency filter approach to correcting artifacts of multichannel two-photon microscopy,” *J. Biomed. Opt.* **29**(1), 016007 (2025).
29. A. Cardona et al., “TrakEM2 software for neural circuit reconstruction,” *PLoS One* **7**, e38011 (2012).
30. S. Saalfeld et al., “As-rigid-as-possible mosaicking and serial section registration of large system datasets,” *Bioinformatics* **26**, i57–i63 (2010).
31. E. Zorpas et al., “Mapping the transcriptome: realizing the full potential of spatial data analysis,” *Cell* **186**, 5677–5689 (2023).
32. R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst. Man Cybern. SMC-3(6), 610–621 (1973).*
33. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv: 1409.1556v6 (2015).
34. A. Weisbrod et al., “Altered PTEN, ATRX, CHGA, CHGB, and TP53 expression are associated with aggressive VHL-associated pancreatic neuroendocrine tumors,” *Horm. Cancer* **4**, 165–175 (2013).
35. S. Duan et al., “GFAP-directed inactivation of men1 exploits glial cell plasticity in favor of neuroendocrine reprogramming,” *Cell Mol. Gastroenterol. Hepatol.* **14**(5), 1025–1051 (2022).
36. C. Shelton et al., “PRSS1-related hereditary pancreatitis,” *GeneReviews*, M. P. Adam et al., Eds., University of Washington, Seattle, Washington (2019).
37. J. Hofland et al., “European neuroendocrine tumor society 2023 guidance paper for functioning pancreatic neuroendocrine tumour syndromes,” *J. Neuroendocrinol.* **35**, e13318 (2023).

38. R. Maréchal et al., “Sonic hedgehog and Gli1 expression predict outcome in resected pancreatic adenocarcinoma,” *Clin. Cancer Res.* **21**, 1215–1224 (2015).
39. K. Taki et al., “Significance of lymph node metastasis in pancreatic neuroendocrine tumor,” *Surg. Today* **47**, 1104–1110 (2017).
40. T. Cannon et al., “Characterization of nadh fluorescence properties under one-photon excitation with respect to temperature, pH, and binding to lactate dehydrogenase,” *OSA Contin.* **4**, 1610–1625 (2021).
41. A. Sánchez-Hernández, C. Polleys, and I. Georgakoudi, “Formalin fixation and paraffin embedding interfere with the preservation of optical metabolic assessments based on endogenous NAD(P)H and FAD two-photon excited fluorescence,” *Biomed. Opt. Express* **14**(10), 5238–5253 (2023).

Shuyuan Guan is currently a PhD candidate at the University of Arizona’s Wyant College of Optical Sciences. She works in the “Biomedical Optics and Optical Measurement” Laboratory with Dr. Travis Sawyer. Her research interests involve the design and development of optical imaging systems and leveraging deep learning techniques to enhance the outcomes of biomedical imaging.

Thomas Knapp: Biography is not available.

Alba Alfonso-Garcia is an assistant professor of biomedical engineering at UC Davis, studied physics at the University of Barcelona and earned her PhD in biomedical engineering from UC Irvine. She leads the Advanced Light-based Applications Laboratory, developing label-free optical spectroscopy and imaging tools for biomedical research and clinical use, including disease detection, therapy monitoring, and surgical guidance. Her research interests include label-free fluorescence lifetime imaging, spontaneous and stimulated Raman spectroscopy, and translational biophotonics.

Suzann Duan is an assistant professor in the Department of Pharmaceutical Sciences at UC Irvine. Her lab focuses on defining genetic and epigenetic mechanisms that contribute to cancer cell plasticity. She completed her postdoctoral training at the University of Arizona Comprehensive Cancer Center, and she holds her BS degree from Boston University and her PhD in cellular and molecular pharmacology and physiology from the University of Nevada, Reno.

Travis W. Sawyer is an assistant professor of optical sciences at University of Arizona’s Wyant College of Optical Sciences. His laboratory focuses on developing novel optical technology for healthcare applications, including cancer screening, minimally invasive procedures, and biomarker identification, with a particular emphasis on gastrointestinal cancer.