Quantitative Characterization of Duodenal Gastrinoma Autofluorescence using Multi-photon Microscopy

Thomas G. Knapp B.S.,[1] Suzann Duan Ph.D.,[2] Juanita L. Merchant M.D.,Ph.D.,[2] Travis W. Sawyer Ph.D.[1,2,3]

[1] Department of Biomedical Engineering, [2] College of Medicine, [3] College of Optical Sciences, University of Arizona, Tucson, AZ, United States of America

Financial support for this study was provided by NIH grant GM132008, University of Arizona Core Facility Grant CA-CFPP April 2021 ICMarley-3502090, R01 DK45729-27, and the University of Arizona Cancer Center Support Grant (P30 CA023074).

The authors have no conflicts of interest to disclose.

A preprint of this work has been published on the Cold Spring Harbor Laboratory bioRxiv server.

DOI: https://doi.org/10.1101/2022.05.19.492747

Corresponding Author: Travis W. Sawyer **Email**: tsawyer9226@email.arizona.edu **Telephone**: (520) 621-8068, **Mailing address**: College of Optical Sciences, University of Arizona, 1630 E. University Blvd., Tucson, AZ 85721

ABSTRACT

Background

Duodenal gastrinomas are neuroendocrine tumors that develop in the submucosa of the duodenum and produce the hormone gastrin. Surgical resection of duodenal gastrinomas is complicated by the small size of these tumors and the tendency for them to develop diffusely in the duodenum. Endoscopic mucosal resection of duodenal gastrinomas is an increasingly popular method for treating this disease due to its low complication rate but suffers from poor rates of pathologically negative margins. Multiphoton microscopy can capture high-resolution images of biological tissue with contrast generated from endogenous fluorescence (autofluorescence) through two-photon excited fluorescence. Second harmonic generation is another popular method of generating image contrast with multi-photon microscopy and is a light-scattering phenomenon that occurs predominantly from structures such as collagen in biological samples. Some molecules that contribute to autofluorescence change in abundance from processes related to the cancer disease process (e.g., metabolic changes, oxidative stress, angiogenesis).

Study Design/Materials and Methods

Multi-photon microscopy was used to image 12 separate patient samples of formalin-fixed and paraffin-embedded duodenal gastrinoma slides with a second-harmonic generation channel four two-photon excited fluorescence channels. The excitation and emission profiles of each two-photon excited fluorescence channel were tuned to capture signal dominated by distinct fluorophores with well-characterized fluorescent spectra and known connections to the physiologic changes that arise in cancerous tissue.

Results

We found that there was a significant difference in the relative abundance of signal generated in the two-photon excited fluorescence channels for regions of duodenal gastrinomas in comparison to the neighboring tissues of the duodenum. Data generated from texture feature extraction of the multi-photon microscopy images were used in linear discriminant analysis models to create classifiers for tumor vs all other tissue types before and after principal component analysis. Principal component analysis improved the classifier accuracy and reduced the number of features required to achieve maximum accuracy. The linear discriminant classifier after principal component analysis distinguished between tumor and other tissue types with an accuracy of 90.6 - 93.8%.

Conclusions

These results suggest that multi-photon microscopy two-photon excited fluorescence and second-harmonic generation imaging is a promising label-free method for discriminating between duodenal gastrinomas and normal duodenal tissue which has implications for future applications of in vivo assessment of resection margins with endoscopic multi-photon microscopy.

1. INTRODUCTION

1.1 Gastroenteropancreatic Neuroendocrine Tumors

Neuroendocrine tumors (NETs) are neoplasms that are thought to originate from secretory cells of the endocrine system and are characterized by their ability to express peptide hormones and biogenic amines.[1] Highly heterogeneous in nature, NETs present with a wide spectrum of clinical features depending on their primary tissue site and molecular features such as their genome and epigenome.[1] The gastroenteropancreatic region is a common site for

gastroenteropancreatic NET (GEP-NET) formation and encompasses a wide variety of neoplasms that can be categorized based on their site of origin and characteristic hormone expression. Approximately 80% of GEP-NETs are found in the small intestine and colon, with most arising in the small intestine.[2] Small intestinal (SI) NETs account for ~2% of all gastrointestinal neoplasms and carry a higher potential for malignancy.[1] SI-NETs are classified as 'functional' or 'non-functional' based on their hormone-producing properties. As a result, functional hormone-secreting tumors occur in 30-40% of SI-NETs, with serotonin, somatostatin, or gastrin being the predominant secretory products.[3]

Gastrin-producing NETs, or gastrinomas, are a relatively rare group of GEP-NETs with an annual incidence of 0.05 – 0.2 cases per 100,000 [4] that develop in the pancreas and duodenum with 60 – 95% of cases occurring in the duodenum.[5] Gastrinoma secretion of the hormone gastrin can lead to the increased stimulation of gastric parietal cells, causing hypersecretion of gastric acid (hyperchlorhydria).[6, 7] The clinical manifestation of chronic gastric acid hypersecretion secondary to gastrinomas was first reported by Zollinger and Ellison and is thus named Zollinger-Ellison Syndrome (ZES).[8] ZES typically presents as pain secondary to heartburn and/or peptic ulcer disease, diarrhea, abnormal weight loss, gastrointestinal bleeding, strictures, or perforations/penetrations of the GI tract.[9] Hypergastrinemia from gastrinoma tumors has also been linked to the development of type I and type II gastric NETs[10] because the elevated blood levels of gastrin also stimulate the enterochromaffin-like cell also located in the body of the stomach along with parietal cells. These gastric NETs have been shown to regress following resection of gastrinomas.[11] Trends in NET cases show that this disease has been steadily growing in incidence over the last five decades which has been partially attributed to increased 'awareness' of the disease through research and improvements in screening and diagnosis.[3]

While conventional cross-sectional CT, MRI or ultrasound are commonly used for pre-operative detection and localization of gastrinomas, these modalities are at risk for missing duodenal gastrinomas (DGASTs) which tend to be small (<0.5cm) and develop diffusely, perpetuating the risk of future metastasis.[12] Improvements in functional imaging using ^{68}Ga-DOTATATE PET/CT has shown promise in diagnosing and localizing primary tumor sites and areas of metastases due to its high sensitivity and specificity for primary and metastatic GEP-NETs using radiolabeled somatostatin receptor analogues.[13] Endoscopic ultrasound (EUS) has been used to determine the depth and size of duodenal tumor growth which informs the surgeon whether an open surgical approach or endoscopic resection (ER) is appropriate.[14] These modalities are both hindered in their ability to assess small lesions due to their low resolution, while EUS is highly operator dependent and limitations with respect to contrast in smaller lesions.[15]

The 2017 guidelines on the surgical resection of small bowel NETs from the North American Neuroendocrine Tumor Society recommend palpation of the entire 20 feet of small bowel to ensure complete oncological resection. These guidelines include resecting the small bowel through a laparoscopic incision or, if needed, converting to a more invasive open abdominal surgery to facilitate adequate palpation.[16] Less invasive surgical approaches such as ER of duodenal NETs has gained popularity and is recommended for duodenal lesions < 10 mm in size by the European Neuroendocrine Tumor Society.[17] While retrospective studies of ER show this procedure as having relatively low complication rates, this method suffers from poor margin definition and a greater rate of recurrence compared to open surgery.[18] Due to the drastic improvement in prognosis with full resection of these tumors,[5] it is clear that methods for performing *in vivo* assessment of tumor clearance would greatly benefit patients that require resection of SI-NETs.

1.2 Autofluorescent Multi-photon Microscopy

The detection of endogenous fluorescence, or autofluorescence (AF), has shown promise in its use as an 'optical biopsy' tool, or as a means of distinguishing between tissue types and healthy versus diseased tissue based on their innate fluorescent properties. [19-22] AF imaging provides a minimally invasive method of probing tissue characteristics without the use of exogenous contrast agents, eliminating the risk of toxicity or potential reaction to the dyes. A major limitation of AF imaging is the relatively low signal-to-noise ratio produced when sensing light emitted from the naturally occurring fluorescent molecules within tissue which can lead to a high false-positive rate.[23] This can be mitigated with careful characterization of the AF properties of diseased and normal tissue. For example, quantitation of AF has been shown to be a promising method of distinguishing between adenomatous and hyperplastic colon polyps[20] and normal, hyperplastic, and dysplastic cells of isolated colon crypts.[21] Multi-photon microscopy (MPM) is a promising modality for quantifying tissue AF at a high resolution and contrast. Two-photon excited fluorescence (2PEF) is one method of generating image contrast in MPM and is the process of using an endogenous fluorophore that absorbs two photons within a narrow enough time interval to create fluorescence emission. The two-photon absorption process utilizes light of longer wavelengths (e.g., 700 - 1000 nm) to excite fluorescing electrons to a similar degree as single-photon fluorescence, which uses light of a shorter wavelength, (e.g., 360 – 550 nm), i.e., a greater photon energy. Second-harmonic generation (SHG) is another popular method of creating MPM contrast and is a nonlinear light scattering phenomenon from non-centrosymmetric structures, which is predominantly collagen in tissue samples.

Use of excitation light of longer wavelengths in MPM allows for imaging deeper into tissue samples due to the inverse relationship between wavelength and light scattering as it

passes through a medium. This would play an essential role for probing through tissue layers deep within the layers of the small intestine.[24] Signal generation from 2PEF and SHG events scales with the square of the excitation irradiance, confining photon emission from the tissue within the point of focus. The confined space of these events results in excitation of fluorophores within the volume of tissue from which the MPM optics can fully focus onto the detector. This allows one to image biologic samples by reducing the chance of inadvertent tissue damage or photobleaching during image acquisition and fluorescence emission outside of the focal plane which could degrade image quality.[26] Ultimately, MPM represents an imaging modality that can acquire volumetric images at a sufficiently high resolution to enable the accurate assessment of the molecular features comprising a tissue specimen.[27] Developing models of DGAST multiphoton fluorescence is an initial step towards *in vivo* label-free measurements of these lesions. This would bypass the need for fluorescent dyes that pose the risk of cytotoxicity while allowing for screening, diagnosis, and staging of DGASTs.

1.3 Multiphoton Image Analysis

Use of a tunable laser light source and multi-channel detector array in MPM of tissue AF allows for the selection of excitation and emission wavebands that are more likely to collect signal from different proportions of fluorescent molecules. For instance, imaging of fluorescent species such as NADH and FAD that are linked to established abnormalities in tumor cells, e.g., the Warburg effect[28] creates a contrast between regions of tumor and surrounding tissue. Other fluorophores with relevance to tissue changes during the cancer disease process include lipofuscin, porphyrins, riboflavins, collagen, elastin, and amino acids.[29]

Quantification of tissue AF and differences in the AF between tumor and normal tissue can be accomplished in several ways. A straightforward approach is the thresholding and direct comparisons of 2PEF and SHG signal abundances.[30] While easy to interpret, this approach is

strongly influenced by variations in light source power, instrument alignment, and tissue preparation. A more robust approach towards quantification is texture analysis, which can provide the spatially-varying statistical characteristics of image tone and texture through the co-occurrence of pixel gray-level values.[31] Texture analysis is valuable to characterize the distribution of image brightness, which can describe features such as contrast, homogeneity, and entropy within the image. In general, texture analysis has been used with success in the context of tissue characterization for cancer imaging, including ovarian cancer using optical coherence tomography[22], breast cancer lesions with MRI[32], and non-small cell lung cancer using non-contrast-enhanced CT.[33] Several different methods of texture analysis exist, which can be classified into four broad categories of statistical methods, structural, model-based, and transform-based methods. Each approach can provide different information about the image content and how that information is contained within the spatial distribution of pixel / voxel values.[34] This information can then be used to train classifiers to distinguish between different classes within images, which in this case would be tissue types.

In this work, we aim to assess the suitability of MPM for tissue characterization and margin definition of DNETs. To our knowledge, this is the first attempt to characterize optical properties of DGAST tumors with MPM. We collected a set of 2PEF and SHG images of fixed human DGAST samples from 12 patients. Image regions of interest (ROIs) corresponding to tumor tissue and surrounding regions of duodenal tissue were quantified by calculating the magnitude of fluorescence emission, as well as conducting statistics-based gray-level co-occurrence matrix (GLCM) texture analysis. Our results showed that multi-photon imaging provides excellent label-free contrast for identifying and delineating cancerous tissue. We found that statistically significant differences in AF and texture features between DGASTs and normal tissue of the duodenum exists. We then developed classifiers using the texture feature and show that the tumor tissue can be successfully differentiated from surrounding tissues with over

90% accuracy. These results are promising and demonstrate that MPM AF imaging can be used as a tool for demarcating areas of DGASTs from healthy tissue of the duodenum. Use of MPM in this context can eventually fill a vital role in the process of screening, localizing, and assessing surgical margins when combined with endoscopy or laparoscopy.

2. METHODS

2.1 Sample Acquisition and Preparation

Formalin-fixed paraffin-embedded (FFPE) slides of GEP-NETs were obtained from the University of Michigan Endocrine Oncology Repository (IRB #HUM00115310) through a Materials Transfer Agreement with the University of Michigan (Dr. Tobias Else). Original tissue specimens were collected from human subjects who received a diagnosis of Multiple Endocrine Neoplasia (MEN1) syndrome, gastrinoma, or both. Slides containing tissue that was not confirmed gastrinoma, either through positive immunohistochemistry for gastrin or documenting hypergastrinemia pre-surgery were excluded from this study. Informed consent was given by the patient prior to sample collection and all specimens were de-identified prior to transferring the slides. Samples were collected during upper endoscopy or during surgical procedure and placed in 10% neutral-buffered formalin prior to paraffin embedding and sectioning. Five-micron FFPE sections were used for downstream imaging studies. In total, 12 separate slides, some including multiple histological sections (19 sections total) were used for this study.

2.2 Multiphoton Imaging

Specimens were imaged using a Zeiss LSM 880 microscope (Carl Zeiss Microscopy, White Plains, NY) with a tunable MaiTai HP titanium:sapphire light source (Spectra-Physics, Mountain View, CA). **Figure** 1 shows a schematic of the imaging system, where a 690+ nm long-pass dichroic mirror and a 660nm short-pass emission filter were used as the main beam splitter and secondary beam splitter in the invisible light beam path. Coverslips were dry-mounted onto the

slides and images were taken at a 700 – 800 master gain and 50 mW average laser power for each of the fluorescence channels and the SHG channel using a plan-apochromat 20X/0.8 NA objective (420650-9902-000, Carl Zeiss Microscopy, White Plains, NY). Gain was adjusted on a per-sample basis and held fixed while acquiring the separate imaging channels to allow for the direct comparison of signal intensity from the separate tissue sites within the image. In total, five separate image channels were collected. The 2PEF excitation and emission bands (**Table 1**) were determined using the known two-photon fluorescent spectra of NADH[35,36], FAD[35], porphyrin[37], and lipofuscin.[38] We selected these fluorophores as they are established biomarkers of cancer development (e.g., due to irregularities in cell metabolism, senescence, and vascularization) and have well-characterized fluorescent properties.

Excitation and Emission Profiles of 2PEF Imaging Channels

Channel	Excitation (nm)	Emission (nm)	Primary Fluorophore
1	700 ± 1.05	425 – 465	NADH
2	800 ± 1.2	590 – 625	Porphyrin
3	830 ± 1.25	500 – 600	Lipofuscin
4	920 ± 1.38	475 – 600	FAD

Table 1: Excitation and emission (detection) bands for the four 2PEF imaging channels alongside the fluorophore expected to produce the dominant signal of the channel.

Large area scans were acquired by collecting individual image tiles of 256 x 256 pixels in size. Images were obtained with 16-bit gray scale resolution and 10% overlap between the separate tiles using an automatic motorized stage. Total area covered by the tiling scan was selected depending on the amount of each tissue type captured within the scan area for adequate ROI sampling within the image and analysis of bulk tissue morphology via image feature extraction. Our criteria for selecting the image size was based on the density of tissue types within the scan area, such that there were no artifacts and less than 90% brightness variation within the area to draw at least three 0.25 mm^2 ROIs for each class contained in the image. The tile-scanned images were collected in Z-stacks to address drop-off in brightness that occurs over the uneven

surface of the sample secondary to the sectioning process. The number of acquisitions in the Z-stack was selected depending on the site and severity of brightness variations. Z-stacks were acquired such that over 90% of the sample area was in-focus with the image, ranging between 3-9 acquisitions in the Z-axis.

2.2. Image processing

To process the acquired images, a flat-field correction was first applied to account for illumination variations. This was accomplished by averaging the pixel brightness for each 256 x 256 tile over all tiles and fitting the result to a linear brightness trend. All collected tiles were then divided by this trend to account for the illumination variation in the laser irradiance. The tiles were then combined using the ImageJ Grid/Collection stitching plugin[39] and a maximum projection image was created using the Z-stack images for the separate image channels. A custom MATLAB frequency filtering step was then used on the maximum projection images. For a more complete writeup on this image processing, please refer to Knapp et al.[40] A MATLAB implementation of the structural similarity index (SSIM)[41] was used to determine if image information was retained post-processing. Visual examination of the images was done to confirm the absence of artifacts and the SSIM score was used to quantitatively support the inclusion of the images in our analysis. An example of the processing steps and assessment with SSIM is shown in **Figure 2**. **Figure 2 A, B** are tiled 2PEF images of a DGAST sample that have been stitched together pre-/post-processing, respectively. **Figure 2 C** is an example of an SSIM map which is a visual representation of structural similarity between a ground-truth image and the post-processed image. In this representation, pixel values range from 0-1 and greater values signify a closer similarity between compared images. The example in **Figure 2 C** shows that image regions outside of the grid artifact are unaffected by the processing method. **Figure 2 D** represents the entire image processing pipeline, **Figure 2 D1** is the custom flat-field

correction, **Figure 2 D2** is the tile stitching process and **Figure 2 D3** is the frequency filtering of residual grid artifact.

ROIs from the 2PEF and SHG imaging channels were extracted using ImageJ software[41] for the separate tissue types present in the image, classified as either stroma, villi / lamina propria, Brunner's glands, abnormal Brunner's glands, and tumor. ROIs were individually inspected, and any containing artifacts were excluded resulting in a total of 209 ROIs from the 12 patient samples included in our analysis. Ground-truth classification of tissue using morphology and immunohistochemistry (IHC) of serial sections is shown in **Figure 3. Figure 3 columns 1-4** are comparisons of the tissue classifications used in this study with traditional H&E staining and MPM generated contrasts. Each row of **Figure 3** represents a different tissue class that has been isolated within an ROI.

Synaptophysin and chromogranin A are traditional markers for NETs. As gastrin-expressing NETs have been shown to stain weakly for chromogranin A[43], synaptophysin was used in combination with a gastrin antibody (**Figure 3 column 6**) to validate that the tumor regions were gastrin-expressing NETs, i.e., gastrinomas. PanCK is an epithelium marker used for determining tissue morphology during staining and has been used for identifying tumor-associated, or abnormal, Brunner's glands (**Figure 3 column 5**).[44]

2.3 Quantification and Comparison of 2PEF and SHG

All analyses were completed using Python 3 and statistical values were generated using the open source SciPy library[45]. A threshold was first applied to each image to remove saturated pixels and to limit the influence of background signal, as well as the inclusion of pixel values containing noise. Lower thresholds were determined for each channel by taking the average value of an image area with no tissue to determine background (**Table 2**). For the FAD-dominant channel, there were strong contributions in fluorescence from lipofuscin deposits that

manifest as localized bright spots, which is a common observation when measuring this biomarker.[46] This channel was set with an upper threshold of 50000 to remove the influence of these features (**Table 2, channel 4)**. Following thresholding, the mean 16-bit pixel value was recorded for each ROI and the measured values were compared between all tissue types within each image. Comparison of fluorescent and SHG intensity was constrained to within-image sample ROIs (e.g., paired comparisons between ROIs for each patient) to eliminate variations between the tissue slide sectioning, and instrument-level variations between image acquisition sessions. We compared different ROIs selected from a single image corresponding to a patient. Since the ROIs were selected from the patient and differed in tissue classification a paired t-test was used to determine the degree of significance between the comparisons of relative 2PEF and SHG signal intensity between the tissue classes. P-values generated from the paired t-tests are shown in **Figures 7** and **8**, where * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$. For all statistical comparisons, we averaged the values of all ROIs created for each tissue class within the slides from each patient to be treated as a different sample (n = 12).

Channel Threshold Values

Channel	Lower Threshold (16-bit pixel value)	Upper Threshold
1	3000	65535
2	3000	65535
3	3000	65535
4	1500	50000
SHG	1000	65535

Table 2: Values of the global thresholds that were applied to the 2PEF and SHG image channels. Channel 4 was the FAD-dominant excitation/emission band. Due to the broad detection band, a high proportion of bright lipofuscin deposits were present in the images of this channel, hence the lowered upper threshold. Lower thresholds were based on averaging of the image background noise.

2.4 Texture Analysis and Classification

We applied texture analysis to the collected ROIs from each of the MPM imaging channels. Haralick's method of extraction[31] based on the GLCM was used to generate thirteen values describing the texture of each ROI. Prior to texture extraction, the ROIs were normalized

and reduced to an 8-bit data range. The GLCM describes the spatial distribution of pixel gray-level values within an image. Values of the GLCM p(*i, j* | *d, θ*) are the measured occurrences of a gray-level (*i*) being a distance, *d*, away from gray-level (*j*) in the direction *θ*, where *θ* is 0 deg, 45 deg, 90 deg, and 135 degrees for a two-dimensional image. The thirteen texture values calculated from the GLCM were averaged for the four directions of *θ* (*see* ***Supp. 1*** *for formulae)*. With five MPM channels, this results in a total of 65 features for each ROI.

The ability to classify the collected ROIs was tested using different combinations of the subsets of the 65 texture features in Linear Discriminant Analysis (LDA) model using singular value decomposition.[47,48] We examined the classification performance between tumor and the four tissue types of villi / lamina propria (grouped as epithelium), stroma, Brunner's gland (BGs), and abnormal Brunner's gland (aBGs) to assess the quantitative ability to differentiate tissue images. The performance of the classifier to distinguish between normal and abnormal BGs was also assessed, as the occurrence of aBGs could represent a transition from normal tissue to disease state (e.g., pre-cancerous tissue). We first tested the LDA classifier using an input of the raw texture feature values and then evaluated the model after applying Principal Component Analysis (PCA) for dimensionality reduction of the feature space. PCA was applied to the dataset containing all ROIs and features, and the top 15 principal components were retained while the others discarded. The top 15 components accounted for 97% of the variance in the dataset.

We then compared the classification performance of different sizes of feature families from this dataset (i.e., families composed of different numbers of features). We tested feature families consisting of subsets ranging from two features to ten features, as texture classifiers have been shown to plateau in performance at around 6 or fewer features.[22,50,51] This process was repeated for the raw feature values, although it required considerably longer computation time to iterate through each feature combination.

We used a leave-one-out approach to measure the accuracy of classifiers using the feature families. We performed an exhaustive comparison of all possible feature families for each size. The top performing feature subsets for each feature subset family were selected as optimal for each size, and the classification accuracy and receive operator characteristic were recorded. Classification accuracy was determined by iterating through all possible combinations of the leave-one-out validation test of the LDA model. This consists of training the model on data from all samples except one (the test sample) and testing the classification on the remaining sample. This process is then repeated fresh with a different sample being the test sample until all samples have been tested against. A score of 0 or 1 was generated depending on if the LDA classifier incorrectly or correctly determined the sample class, respectively. For example, when assessing classifier performance in classifying images of tumor against normal gland tissue, the scores generated through all iterations of testing are cumulated and the final accuracy is the mean value of this score. The equation for determining classifier accuracy is as follows.

$$Accuracy = \frac{(N_c \times 1) + (N_i \times 0)}{N_c + N_i}$$

Where N_c is the number of correct classifications and N_i is the number of incorrect classifications after iterating through all possible leave-one-out combinations. Thus, the accuracy is essentially the proportion of correctly classified samples.

3. RESULTS & DISCUSSION

3.1 Qualitative Assessment of Tissue Contrast from 2PEF and SHG

Clear morphologic differences between the separate tissue classes can be distinguished with qualitative assessment of the 2PEF and SHG images. **Figure 4** shows the different morphologic information provided by SHG (**Figure 4 B**) and 2PEF (**Figure 4 C-F)** imaging. There is a clear increase in cell density within the tumor regions shown in the 2PEF images

while SHG shows the disorganized and thick stromal (S) changes near the tumor sites which is consistent with prior findings that exhibit desmoplastic changes occurring around tumor regions.[52] There is a loss of SHG signal and the appearance of thick collagenous encapsulation of the neoplasms. In comparison, the 2PEF and SHG images of the neighboring Brunner's glands (BGs) show clear contrast to the tumor in the organized arrangement of the glands with thin collagen walls composing their structure. Immunostaining of the DGAST samples (**Figure 4, H)** shows a high level of gastrin and synaptophysin expression within the tumor regions which is characteristic of DGASTs.[53] While clear distinctions between normal and abnormal BGs (aBGs) could not be made with qualitative assessment of 2PEF and SHG imaging, aBGs were shown to stain for gastrin and the epithelial marker PanCK. aBGs were typically located near tumor regions and were demarcated due to BGs being a suspected precursor to DGASTs.[44,54] H & E staining of the DGAST samples (**Figure 4 A)** shows the trabecular patterning of DGASTs[55] in the tumor region corresponding to the marked (T) region of the NADH channel image.

Figures 5-7 show H&E, SHG, and 2PEF images for three other patient samples. In **Figure 5**, we see a comparison of bordering BGs and aBGs near a large region of tumor. While not yet clear in its significance, there appears to be a reduction in the SHG signal (**Figure 5 B**) and thickening of the gland tissue seen in the FAD-dominant 2PEF channel (**Figure 5 C**) seen in the aBGs. **Figure 6** shows the bordering region of a tumor and normal BGs. **Figure 6 B** shows the loss of SHG signal within the tumor that is normally occurring within the gland structure. The tumor exhibits many bright deposits in the lipofuscin-dominant 2PEF channel (**Figure 6 C**). **Figure 7** is an example of the thickened stroma surrounding sites of tumor which is accentuated in SHG imaging (**Figure 7 B**). Bright deposits, like those seen in **Figure 6 C,** were present in the DGAST region when imaged with the porphyrin (**Figure 7 C**), lipofuscin, and FAD-dominant channels. These clusters are thought to be dye used for tumor localization prior to resection and/or accumulation of lipofuscin deposits.

3.2 Quantitative Assessment of 2PEF and SHG Image Channel Intensities

T-test comparisons of the raw relative mean signal magnitude collected from the tissue classes in the separate 2PEF and SHG imaging channels (**Supp. 2)** shows a significant difference ($p < 0.001$ for the NADH, lipofuscin, and porphyrin-dominant channels, (**S2 A, C, D**), $p < 0.01$ for the FAD-dominant channel (**S2 B**) in 2PEF between the tumor regions and glandular tissue. Epithelial tissue (villi and lamina propria) exhibited a significantly greater signal in the lipofuscin ($p < 0.05$, **S2 C**) and NADH-dominant ($p < 0.001$, **S2 A**) channels in comparison to the tumor tissue. This reflects earlier findings of broad-spectrum AF signal emitted from lipofuscin and lipofuscin like pigment within the lamina propria of mice and humans.[45] As expected, stromal tissue produced a significantly greater ($p < 0.001$, **S2 E**) SHG signal in comparison to the epithelial, glandular, and tumor tissue classes.

A correlation matrix of the image channel data was produced to assess crosstalk between imaging channels (**S2 F)**, which suggested that a large degree of overlap existed between the lipofuscin, FAD, and porphyrin channels. This was unsurprising considering the broad emission band of lipofuscin[56] and comparisons were made following normalization of each ROI through division of the corresponding lipofuscin-dominant image channel as shown in **Figure 8**. The correlation matrix of the normalized data (**Figure 8 F**) shows a lower degree of overlap between the image channels. The process of normalizing using the channel tuned to collect lipofuscin fluorescence was done to mitigate the crosstalk between 2PEF channels that could be dominated by tissue lipofuscin fluorescence, which could occur in each imaging channel simultaneously due to its broad emission spectrum.[56] There was a mild decrease in the level of difference of the 2PEF signal magnitudes between the tissue types following normalization. Notably, the tumor tissue no longer produced a significantly greater magnitude of signal in the NADH (**Figure 8 A**), FAD (**Figure 8 B**) and porphyrin-dominant (**Figure 8 D**) channels in comparison to the aBGs tissue class. While signal magnitude decreased in all 2PEF

channels following normalization, the greater loss in signal for the aBGs may suggest that lipofuscin accumulation plays a larger role in the AF signal produced by this tissue in comparison to normal BGs (**Figure 8 C**). The unchanged relationship in SHG signal from the respective tissue classes (**Figure 8 E**) following normalization suggests that little to no crosstalk occurred between this and the 2PEF channels.

3.3 Classification Using Texture Features

Table 3 shows the results of the LDA accuracy for classifying the tumor vs all other tissue classes and BGs vs aBGs using the extracted texture features from the ROIs prior to PCA. Accuracy is defined as the total number of correct classifications divided by the total number of attempts made in the leave-one-out process. We chose the tissue comparisons based on the clinical relevance of tumor localization and the identification of potentially pre-cancerous tissues as described by the aBG state. The results show that, using raw feature values, the accuracy for classifying tissue types tends to plateau with five to six features. The accuracy for delineating tumor, BG, aBG, and VLP all exceeds 90%, yet is limited to 71.9% for stroma. Additionally, we see that the accuracy for classifying BG and aBG is maximized at 79.2%. For most comparisons, we see a decrease in accuracy with the addition of features after the plateau point, which is a common observation when performing feature selection and classification with high-dimensional datasets. This plateau occurs when additional features no longer increase the information content for classification.[57,58] This observation suggests that six features can capture most of the relevant information.

Pre-PCA LDA Classifier Accuracy

Number of features used in classifier	% accuracy for classifying BGs vs aBGs	% accuracy for classifying tumor vs BGs	% accuracy for classifying tumor vs aBGs	% accuracy for classifying tumor vs VLP	% accuracy for classifying tumor vs stroma
2	75.0	87.5	83.3	93.3	62.5
3	75.0	87.5	87.5	96.7	65.6
4	75.0	87.5	91.7	96.7	68.8
5	79.2	90.6	91.7	96.7	68.8
6	75.0	90.6	91.7	96.7	71.9
7	79.2	90.6	87.5	96.7	68.8
8	75.0	90.6	83.3	96.7	68.8

Table 3: *Number of features prior to principal component analysis of the texture feature dataset included in the classifier and the resulting mean level of accuracy of the classifier in separating the tissue groups.*

Principal component analysis (PCA) was used to reduce the dimensionality of the large set of texture features and the top 15 principal components were used in the LDA classifier. **Table 4** shows the results of measuring the LDA classifier accuracy for discriminating between ROIs of tumor vs all other tissue classes using the texture features following PCA.

Post-PCA LDA Classifier Accuracy

Number of features used in classifier	% accuracy for classifying BGs vs aBGs	% accuracy for classifying tumor vs BGs	% accuracy for classifying tumor vs aBGs	% accuracy for classifying tumor vs VLP	% accuracy for classifying tumor vs stroma
3	83.3	87.5	87.5	90.0	90.6
4	83.3	87.5	91.7	90.0	93.8
5	83.3	90.6	87.5	90.0	93.8
6	83.3	90.6	91.7	90.0	93.8
7	83.3	90.6	87.5	93.3	93.8
8	83.3	93.6	87.5	93.3	90.6

Table 4: *Number of features following principal component analysis of the texture feature dataset included in the classifier and the resulting mean level of accuracy of the classifier in separating the tissue groups.*

Dimension reduction with PCA had an insignificant effect in improving classification accuracy between tumor and aBGs; however, we do observe that with PCA applied, the classifier reaches maximum accuracy with fewer features. There was a marked improvement in accuracy for the classification of tumor vs stroma following PCA (71.9 to 93.8%) and mild improvement in performance for classifying tumor vs BGs (90.6 to 93.8%) and BGs vs aBGs (79.2 to 83.3%). The classifier's accuracy in separating the tumor and VLP classes decreased following PCA

(96.7 to 93.3%). Overall, the LDA classifier reached maximum accuracy with fewer features overall. The decrease in accuracy for the classification of tumor and VLP is not surprising, as while PCA will identify the directions of maximum variance in data, it does not necessarily lead to the maximum separability between all groups. However, considering the marked increase in classification accuracy between the other four groups and the achievement of maximum accuracy with fewer features, the results suggest that PCA is an effective technique for dimension reduction and feature selection with this dataset.

The PCA transformed texture feature data was used to generate receiver operating characteristic (ROC) curves and LDA projections for tumor vs VLP, stroma, and BGs using the first two linear discriminants as shown in **Figure 9**. The ROC curves show a similar trend in the plateauing accuracy of classifiers using texture features[22,50,51], although this plateau did range between 4-8 features depending on the tissue class. The ability for LDA to cluster image classes based on texture feature data appeared to vary when discriminating between different tissue types. There was greater clustering of the normal tissue classes vs tumor (**Figure 9 A-C**) in comparison to the aBGs vs tumor (**Figure 9 D**). While potential similarities in image texture for aBGs and DGAST tumor would need further investigation, the reduced clustering performance of the LDA classifier is a likely contributor to the reduced accuracy of its discrimination between tumor and aBGs (**Table 4**).

4. CONCLUSION

ER is an increasingly popular treatment for DGASTs due to its relatively low complication rate in comparison to open surgical resection. Currently, there exists a discrepancy between endoscopic and pathologically negative margins of tumors resected with this procedure. Optical imaging using MPM can capture AF and SHG information of tissue at sub-cellular resolution. Modulating endoscopy with MPM imaging could allow for real-time *in vivo* analysis of resection sites, reducing the likelihood of missed margins. In this work, we have shown that it is possible

to distinguish between DGAST tumors and the normal duodenal tissue surrounding these lesions using multi-channel AF images collected with MPM. The separate tissue classes showed clear morphological differences in both the 2PEF and SHG image channels which corresponded well with H & E and immunostained images of the tissue. Collecting multiple 2PEF channels provided a robust means of differentiating between tissue classes through varying levels of significant differences existing in signal magnitude depending on the tissue and imaging channel. Imaging channels that collected signals dominated by fluorophores tied to cell metabolism and senescence, such as NADH, FAD and lipofuscin, respectively, provided a high degree of contrast between tumor and the neighboring BGs ($p < 0.001$, $p < 0.05$, $p < 0.001$ for each channel post-normalization). Normalization of the imaging channels showed that significant differences in 2PEF signal intensity exists between the DGAST tumors and normal tissue even while controlling for the high cellular density in the cancer. Texture features (e.g., contrast, entropy, autocorrelation) were extracted for regions of interest from five separate image channels for each tissue type. Subsets of the 65 texture features were used to create classifiers through linear discriminant analysis that classified the DGAST tumor against surrounding tissues with up to 96% accuracy. Principal component analysis (PCA) was used to reduce the dimensionality of the 65 texture features to 15 principal components containing 97% variance of the data. A new classifier using subsets of the transformed data showed improved classification accuracy on average of 4.5% between tumor and surrounding tissues. Use of PCA also resulted in achieving higher accuracy with fewer features included in the subset of data used. In both cases, determination of the optimal number of features for highest classification accuracy was made based on the plateauing of accuracy in relation the feature number shown in **Tables 3** & **4**. Another method of determining the optimal number of features included in the LDA classifiers would be to extract the true positive and true negative predictive values of the classifier at each level of feature inclusion.

These early results show promising use of AF as a label-free approach for determining regions of DGAST tumors and potential for differentiating between areas of normal and abnormal glandular tissue which may be areas of early neoplastic change. With the high-resolution achievable with optical imaging, refinement of this classification process may offer the ability to perform margin analysis, screening, and/or diagnosis of NETs *in vivo*. Optimization of the classifier will be approached through further assessment of the amount of spatial and spectral information required to produce discrete levels of accuracy (i.e., the optimum spatial resolution and selection of spectral bands to collect during imaging). The combination of streamlining data collection and feature extraction while limiting the data size will make real-time analysis of NETs and the surrounding tissue during multi-photon imaging feasible with the ultimate goal of applying this methodology to live tissue.

Although the images for this study were collected from cross-sectioned specimens, *in vivo* measurement would necessarily be conducted in the enface direction. Contrast and texture features generated from autofluorescence is expected to be consistent regardless of the imaging direction for tumor and glandular tissues, which tend to have isotropic features. However, the penetration depth in thick tissues using the wavelengths in this study is limited to 400 – 1000 microns.[59] DGASTs tend to form directly beneath and invade into the epithelial layer of the duodenum which, making up approximately half the duodenal wall thickness, can be 400 – 800 microns thick.[60] Thus, it is still feasible to form high quality images of DGASTs *in vivo.* A further application of MPM for managing DGASTs could be in the form of margin definition during resection, where the exposed tissue surface post-resection is the object of interest. In this case, the need for greater penetration depth would be reduced. Further challenges for the potential clinical translation of this methodology lie in controlling the peristaltic / random motion of the small intestine during the scanned imaging and the mitigation of spherical aberrations with greater imaging depths. Potential solutions would be stabilizing the scanning area with

endoscopic suction and introducing an index-matching medium between the imaging probe and the lumen of the duodenum. While challenges exist for translating this technology to *in vivo* application, our results are promising for establishing feasibility in applying MPM for DGAST localization and margin definition. Further challenges are balancing patient exposure to laser light while achieving adequate signal-to-noise and imaging depth. Guidelines on the allowable exposure to laser light during medical procedures do not specify the maximum permissible exposure for pulsed lasers. Per the OSHA technical manual section 3: chapter 6 guidelines, thermal heating is the major mechanism behind tissue damage for repetitively pulsed or scanning lasers. The FDA/CDRH Federal Laser Product Performance Standard accepts a cumulative radiant energy exposure of 3.85 mJ will not cause a biological effect. The image parameters used in this study would result in a radiant exposure of 0.19 mJ in each 1.28 μm^2 pixel scan region and an irradiance of 14.17 mW / mm^2 at the greatest average laser power (see Supp. 3 for laser power measurements). Podgorski et al.[61] have shown the effects of MPM in murine brain tissue, where irradiances greater than 100 mW / mm^2 resulted in tissue heating but only continuous illumination above 250 mW / mm^2 resulting in lasting tissue damage. However, these results have limited transferability given differences in organ and species. Therefore, further in vivo assessment of the thermal load introduced with illumination conditions for MPM is necessary to fully determine if adequate imaging can be performed without the risk of tissue damage.

ACKNOWLEDGEMENTS

Funding support for this study was provided by NIH grant GM132008, University of Arizona Core Facility Grant CA-CFPP April 2021 ICMarley-3502090, R01 DK45729-27, and the University of Arizona Cancer Center Support Grant (P30 CA023074). Special thanks to Patty Jansma and the University of Arizona Marely Imaging Core Facility.

REFERENCES

1. Cives, M., & Strosberg, J. R. (2018). Gastroenteropancreatic Neuroendocrine Tumors. *CA: A Cancer Journal for Clinicians*, *68*(6), 471–487. https://doi.org/https://doi.org/10.3322/caac.21493
2. Frilling A, Akerstrom G, Falconi M, et al. Neuroendocrine tumor disease: an evolving landscape. *Endocr Relat Cancer.* 2012; **19**: R163- R185.
3. Clift, A. K., Kidd, M., Bodei, L., Toumpanakis, C., Baum, R. P., Oberg, K., Modlin, I. M., & Frilling, A. (2020). Neuroendocrine Neoplasms of the Small Bowel and Pancreas. *Neuroendocrinology*, *110*(6), 444–476. https://doi.org/10.1159/000503721
4. Lloyd RV, Osamura RY, Klöppel G, Rosai J. WHO classification of tumours of endocrine organs 4th edition, Lyon, France: IARC Press: pp 229-232, 2017.
5. Norton, J. A., Foster, D. S., Ito, T., & Jensen, R. T. (2018). Gastrinomas: Medical or Surgical Treatment. In *Endocrinology and Metabolism Clinics of North America* (Vol. 47, Issue 3, pp. 577–601). W.B. Saunders. https://doi.org/10.1016/j.ecl.2018.04.009
6. Maton, P. N., Lack, E. E., Collen, M. J., Cornelius, M. J., David, E., Gardner, J. D., & Jensen, R. T. (1990). The effect of Zollinger-Ellison syndrome and omeprazole therapy on gastric oxyntic endocrine cells. *Gastroenterology*, *99*(4), 943–950.https://doi.org/10.1016/0016-5085(90)90611-4
7. Roy PK, Venzon DJ, Feigenbaum KM, et al. Gastric secretion in Zollinger-Ellison syndrome: correlation with clinical expression, tumor extent and role in diagnosis - A prospective NIH study of 235 patients and review of the literature in 984 cases. Medicine (Baltimore) 2001; 80:189–222
8. ZOLLINGER, R. M., & ELLISON, E. H. (1955). Primary peptic ulcerations of the jejunum associated with islet cell tumors of the pancreas. *Annals of Surgery*, *142*(4), 709–728. https://pubmed.ncbi.nlm.nih.gov/13259432
9. Roy, P. K., Venzon, D. J., Shojamanesh, H., Abou-Saif, A., Peghini, P., Doppman, J. L., Gibril, F., & Jensen, R. T. (2000). Zollinger-Ellison Syndrome: Clinical Presentation in 261 Patients. *Medicine*, *79*(6). https://journals.lww.com/mdjournal/Fulltext/2000/11000/Zollinger_Ellison_Syndrome__Clinical_Presentation.4.aspx
10. Rindi G, Azzoni C, La Rosa S, et al. ECL cell tumor and poorly differentiated endocrine carcinoma of the stomach: prognostic evaluation by pathological analysis. Gastroenterology 1999; 116:532-542
11. Delle Fave, G., Capurso, G., Milione, M., & Panzuto, F. (2005). Endocrine tumours of the stomach. *Best Practice & Research Clinical Gastroenterology*, *19*(5), 659–673. https://doi.org/https://doi.org/10.1016/j.bpg.2005.05.002
12. Ito T, Igarashi H, Jensen RT. Zollinger-Ellison syndrome: recent advances and controversies. *Curr Opin Gastroenterol*. 2013;29(6):650-661. doi: 10.1097/MOG.0b013e328365efb1
13. Sadowski, S. M., Neychev, V., Millo, C., Shih, J., Nilubol, N., Herscovitch, P., Pacak, K., Marx, S. J., & Kebebew, E. (2016). Prospective Study of 68Ga-DOTATATE Positron Emission Tomography/Computed Tomography for Detecting Gastro-Entero-Pancreatic Neuroendocrine Tumors and Unknown Primary Sites. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *34*(6), 588–596. https://doi.org/10.1200/JCO.2015.64.0987
14. Yoon, J. Y., Kumta, N. A., & Kim, M. K. (2021). The Role of Endoscopy in Small Bowel Neuroendocrine Tumors. *Clinical Endoscopy*, *54*(6), 818–824. https://doi.org/10.5946/ce.2020.296
15. Howe JR, Cardona K, Fraker DL, et al. The Surgical Management of Small Bowel Neuroendocrine Tumors: Consensus Guidelines of the North American Neuroendocrine Tumor Society. *Pancreas*. 2017;46(6):715-731. doi:10.1097/MPA.0000000000000846
16. Kim, M. K. (2012). Endoscopic ultrasound in gastroenteropancreatic neuroendocrine tumors. *Gut and Liver*, *6*(4), 405–410. https://doi.org/10.5009/gnl.2012.6.4.405
17. Brito H P, Torres I T, Turke K C et al. Comparison of endoscopic resection techniques for duodenal neuroendocrine tumors: systematic review. Endoscopy International Open 2021; 09(08): 1214 - 1221. doi:10.1055/a-1487-5594
18. Kumar, S., Mahmud, N., Roses, R. E., Katona, B. W., Ginsberg, G. G., & Metz, D. C. (2020). Resection Trends for Duodenal Carcinoid Tumors: A Single-Center Experience. *Pancreas*, *49*(1), e11–e13. https://doi.org/10.1097/MPA.0000000000001448

19. Nie, Z., Yeh, S.-C. A., LePalud, M., Badr, F., Tse, F., Armstrong, D., Liu, L. W. C., Deen, M. J., & Fang, Q. (2020). Optical Biopsy of the Upper GI Tract Using Fluorescence Lifetime and Spectra. *Frontiers in Physiology*, *11*. https://www.frontiersin.org/article/10.3389/fphys.2020.00339
20. McCallum, A. L., Jenkins, J. T., Gillen, D., & Molloy, R. G. (2008). Evaluation of autofluorescence colonoscopy for the detection and diagnosis of colonic polyps. *Gastrointestinal Endoscopy*, *68*(2), 283–290. https://doi.org/10.1016/j.gie.2007.10.039
21. DaCosta, R. S., Andersson, H., Cirocco, M., Marcon, N. E., & Wilson, B. C. (2005). Autofluorescence characterisation of isolated whole crypts and primary cultured human epithelial cells from normal, hyperplastic, and adenomatous colonic mucosa. *Journal of Clinical Pathology*, *58*(7), 766 LP – 774. https://doi.org/10.1136/jcp.2004.023804
22. Sawyer, T. W., Koevary, J. W., Rice, F., Howard, C. C., Austin, O. J., Connolly, D. C., Cai, K. Q., & Barton, J. K. (2019). Quantification of multiphoton and fluorescence images of reproductive tissues from a mouse ovarian cancer model shows promise for early disease detection. *Journal of biomedical optics*, *24*(9), 1–16. https://doi.org/10.1117/1.JBO.24.9.096010
23. Wong Kee Song, L.-M., Banerjee, S., Desilets, D., Diehl, D. L., Farraye, F. A., Kaul, V., Kethu, S. R., Kwon, R. S., Mamula, P., Pedrosa, M. C., Rodriguez, S. A., & Tierney, W. M. (2011). Autofluorescence imaging. *Gastrointestinal Endoscopy*, *73*(4), 647–650. https://doi.org/10.1016/j.gie.2010.11.006
24. Zipfel, W. R., Williams, R. M., & Webb, W. W. (2003). Nonlinear magic: Multiphoton microscopy in the biosciences. In *Nature Biotechnology* (Vol. 21, Issue 11, pp. 1369–1377). https://doi.org/10.1038/nbt899
25. Perry, S. W., Burke, R. M., & Brown, E. B. (2012). Two-photon and second harmonic microscopy in clinical and translational cancer research. *Annals of Biomedical Engineering*, *40*(2), 277–291. https://doi.org/10.1007/s10439-012-0512-9
26. Niesner, R., Andresen, V., Neumann, J., Spiecker, H., & Gunzer, M. (2007). The power of single and multibeam two-photon microscopy for high-resolution and high-speed deep tissue and intravital imaging. *Biophysical Journal*, *93*(7), 2519–2529. https://doi.org/10.1529/biophysj.106.102459
27. Chen D, Nauen DW, Park HC, et al. Label-free imaging of human brain tissue at subcellular resolution for potential rapid intra-operative assessment of glioma surgery. *Theranostics*. 2021;11(15):7222-7234. Published 2021 May 24. doi:10.7150/thno.59244
28. Vaupel, P., & Multhoff, G. (2021). Revisiting the Warburg effect: historical dogma versus current understanding. *The Journal of Physiology*, *599*(6), 1745–1757. https://doi.org/https://doi.org/10.1113/JP278810
29. Khosroshahi, M. E., & Rahmani, M. (2012). Detection and Evaluation of Normal and Malignant Cells Using Laser-Induced Fluorescence Spectroscopy. *Journal of Fluorescence*, *22*(1), 281–288. https://doi.org/10.1007/s10895-011-0958-4
30. Ara A. Ghazaryan, Jo-Ya Tseng, Wen Lo, Yang-Fang Chen, Vladimir Hovhannisyan, Shean-Jen Chen, Hsin-Yuan Tan, and Chen-Yuan Dong "Multiphoton imaging and quantification of tissue glycation", Proc. SPIE 7895, Optical Biopsy IX, 789509 (16 February 2011); https://doi.org/10.1117/12.874536
31. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314
32. Waugh, S. A., Purdie, C. A., Jordan, L. B., Vinnicombe, S., Lerski, R. A., Martin, P., & Thompson, A. M. (2016). Magnetic resonance imaging texture analysis classification of primary breast cancer. *European Radiology*, *26*(2), 322–330. https://doi.org/10.1007/s00330-015-3845-6
33. Ganeshan, B., Abaleke, S., Young, R. C. D., Chatwin, C. R., & Miles, K. A. (2010). Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging : The Official Publication of the International Cancer Imaging Society*, *10*(1), 137–143. https://doi.org/10.1102/1470-7330.2010.0021
34. Armi, L., & Fekri Ershad, S. (2019). *Texture image analysis and texture classification methods - A Review*. *2*, 1–29.
35. Schaefer, P. M., Kalinina, S., Rueck, A., Arnim, C. A. F. von, & Einem, B. von. (2019). NADH Autofluorescence—A Marker on its Way to Boost Bioenergetic Research. *Cytometry Part A*, *95*(1), 34–46. https://doi.org/10.1002/CYTO.A.23597
36. Kolenc, O. I., & Quinn, K. P. (2019). Evaluating Cell Metabolism Through Autofluorescence Imaging of NAD(P)H and FAD. *Antioxidants & Redox Signaling*, *30*(6), 875–889. https://doi.org/10.1089/ars.2017.7451

37. Mathai, S., Bird, D. K., Stylli, S. S., Smith, T. A., & Ghiggino, K. P. (2007). Two-photon absorption cross-sections and time-resolved fluorescence imaging using porphyrin photosensitisers. *Photochemical & Photobiological Sciences*, *6*(9), 1019–1026. https://doi.org/10.1039/B705101H
38. di Guardo, G. (2015). Lipofuscin, lipofuscin-like pigments and autofluorescence. *European Journal of Histochemistry: EJH*, *59*(1), 2485. https://doi.org/10.4081/ejh.2015.2485
39. Preibisch S, Saalfeld S, Tomancak P. Globally optimal stitching of tiled 3D microscopic image acquisitions. Bioinformatics. 2009;25(11):1463-1465. doi:10.1093/bioinformatics/btp184
40. Thomas Knapp, Natzem Lima, Suzann Duan, Juanita L. Merchant, Travis W. Sawyer, "Evaluation of tile artifact correction methods for multiphoton microscopy mosaics of whole-slide tissue sections," Proc. SPIE 11966, Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXIX, 119660D (2 March 2022); https://doi.org/10.1117/12.2609634
41. Zhou, W., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity." IEEE Transactions on Image Processing. Vol. 13, Issue 4, April 2004, pp. 600–612.
42. Rasband W., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland: (2012).
43. Bellizzi, A. M. (2020). Immunohistochemistry in the diagnosis and classification of neuroendocrine neoplasms: what can brown do for you? *Human Pathology*, *96*, 8–33. https://doi.org/10.1016/j.humpath.2019.12.002
44. Rico K, Duan S, Pandey RL, et al. Genome analysis identifies differences in the transcriptional targets of duodenal versus pancreatic neuroendocrine tumours
BMJ Open Gastroenterology 2021;8:e000765. doi: 10.1136/bmjgast-2021-000765
45. Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. (2020) **SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python**. *Nature Methods*, 17(3), 261-272.
46. Wizenty, J., Ashraf, M. I., Rohwer, N., Stockmann, M., Weiss, S., Biebl, M., Pratschke, J., Aigner, F., & Wuensch, T. (2018). Autofluorescence: A potential pitfall in immunofluorescence-based inflammation grading. *Journal of Immunological Methods*, *456*, 28–37. https://doi.org/https://doi.org/10.1016/j.jim.2018.02.007
47. Welling, M. Fisher linear discriminant analysis. University of Toronto, Department of Computer Science, 2007.
48. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011
49. Lingley-Papadopoulos C. A., et al., “Computer recognition of cancer in the urinary bladder using optical coherence tomography and texture analysis,” J. Biomed. Opt. 13(2), 024003 (2008).10.1117/1.2904987
50. Sawyer T., et al., “Three-dimensional texture analysis of optical coherence tomography images of ovarian tissue,” Phys. Med. Biol. 63(23), 235020 (2018).10.1088/1361-6560/aaefd2
51. Duda R. O., Hart P. E., Stork D. G., Pattern Classification, 2 ed., Wiley, New York: (2001).
52. L. Li *et al*., "Label-Free Identification of Early Gastrointestinal Neuroendocrine Tumors via Biomedical Multiphoton Microscopy and Automatic Image Analysis," in *IEEE Access*, vol. 8, pp. 105681-105689, 2020, doi: 10.1109/ACCESS.2020.3000289.
53. Ramírez-Rentería, C., Ferreira-Hermosillo, A., Marrero-Rodríguez, D., Taniguchi-Ponciano, K., Melgar-Manzanilla, V., & Mercado, M. (2020). An Update on Gastroenteropancreatic Neuroendocrine Neoplasms: From Mysteries to Paradigm Shifts. *Archives of Medical Research*, *51*(8), 765–776. https://doi.org/https://doi.org/10.1016/j.arcmed.2020.06.018
54. Klöppel, G., Anlauf, M., & Perren, A. (2007). Endocrine Precursor Lesions of Gastroenteropancreatic Neuroendocrine Tumors. *Endocrine Pathology*, *18*(3), 150–155. https://doi.org/10.1007/s12022-007-0025-5
55. Milione, M., Parente, P., Grillo, F., Zamboni, G., Mastracci, L., Capella, C., Fassan, M., & Vanoli, A. (2021). Neuroendocrine neoplasms of the duodenum, ampullary region, jejunum and ileum. *Pathologica*, *113*(1), 12–18. https://doi.org/10.32074/1591-951X-228
56. Croce, A. C., & Bottiroli, G. (2014). Autofluorescence Spectroscopy and Imaging: A Tool for Biomedical Research and Diagnosis. *European Journal of Histochemistry: EJH*, *58*.

57. Dale J. Waterhouse, A. Siri Luthman, and Sarah E. Bohndiek "Spectral band optimization for multispectral fluorescence imaging", Proc. SPIE 10057, Multimodal Biomedical Imaging XII, 1005709 (15 February 2017); https://doi.org/10.1117/12.2253069
58. Travis W. Sawyer, Sarah E. Bohndiek, "Towards a simulation framework to maximize the resolution of biomedical hyperspectral imaging," Proc. SPIE 10412, Diffuse Optical Spectroscopy and Imaging VI, 104120C (28 July 2017); https://doi.org/10.1117/12.2284109
59. Cho, H. J., Chun, H. J., Kim, E. S., & Cho, B. R. (2011). Multiphoton microscopy: an introduction to gastroenterologists. In *World journal of gastroenterology* (Vol. 17, Issue 40, pp. 4456–4460). https://doi.org/10.3748/wjg.v17.i40.4456
60. Svein Ødegaard, Lars B Nesje, Ole Didrik Lærum & Michael B Kimmey (2012) High-frequency ultrasonographic imaging of the gastrointestinal wall, Expert Review of Medical Devices, 9:3, 263-273, DOI: 10.1586/erd.12.6
61. Podgorski, K., & Ranganathan, G. (2016). Brain heating induced by near-infrared lasers during multiphoton microscopy. *Journal of Neurophysiology*, *116*(3), 1012–1023. https://doi.org/10.1152/jn.00275.2016

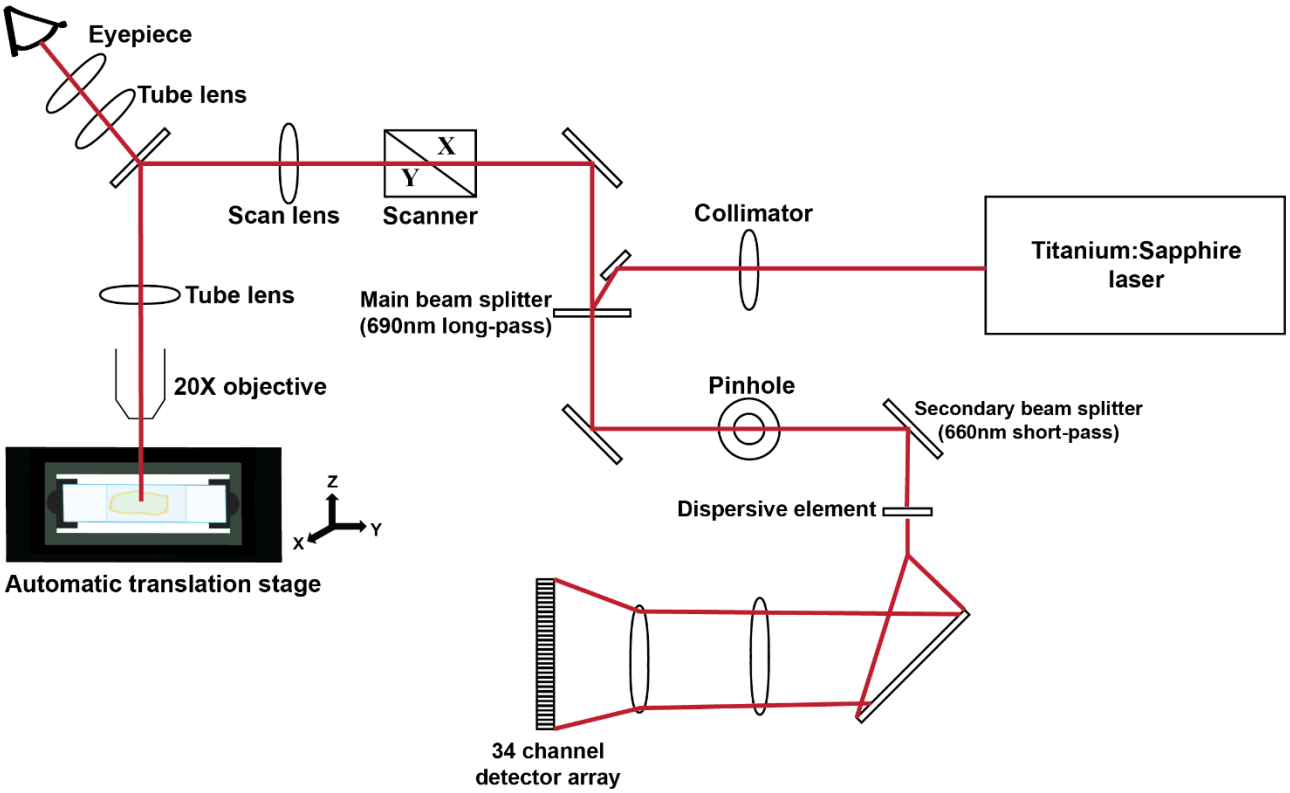


Figure 1: Block diagram of the multiphoton imaging system using a tunable titanium sapphire laser light source, tunable detector array, X,Y,Z translation stage and 20X air objective.

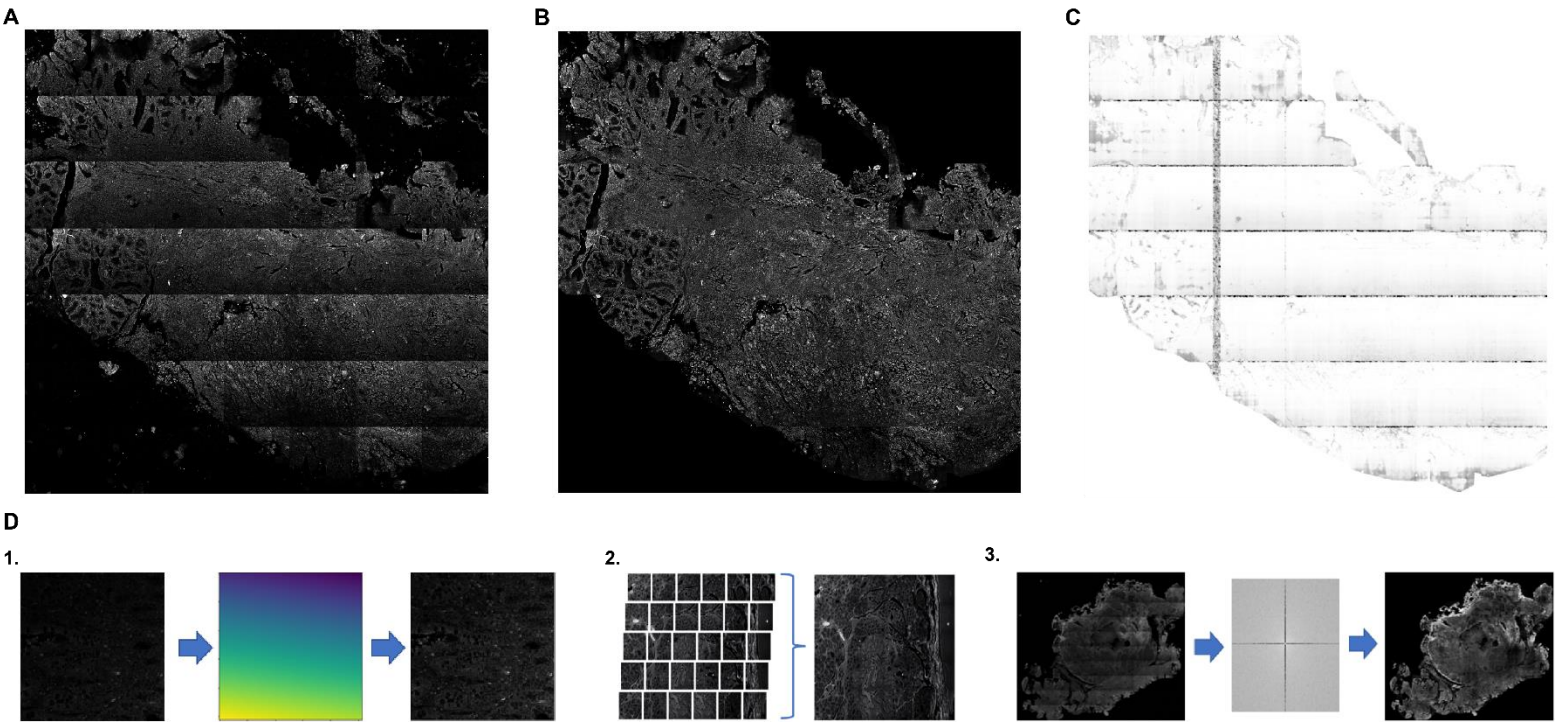


Figure 2: 2PEF channel pre-processing (A) and post-processing (B) showing the presence and removal of the tiling artifact. The SSIM image of the same region post-processing (C) shows good retention of image structural information, indicated by the peak value of 1 represented as white pixel values in areas where image structure was retained through the processing steps. The dark lines on the SSIM show where the grid line artifact was removed. On the bottom row of the figure (D) is an abbreviated illustration of the image processing steps. 1 = custom flat-field approach for correcting image tile brightness variations, 2 = ImageJ stitching methods to combine image tiles, 3 = frequency filtering to remove residual grid lines following the combination of the image tiles.

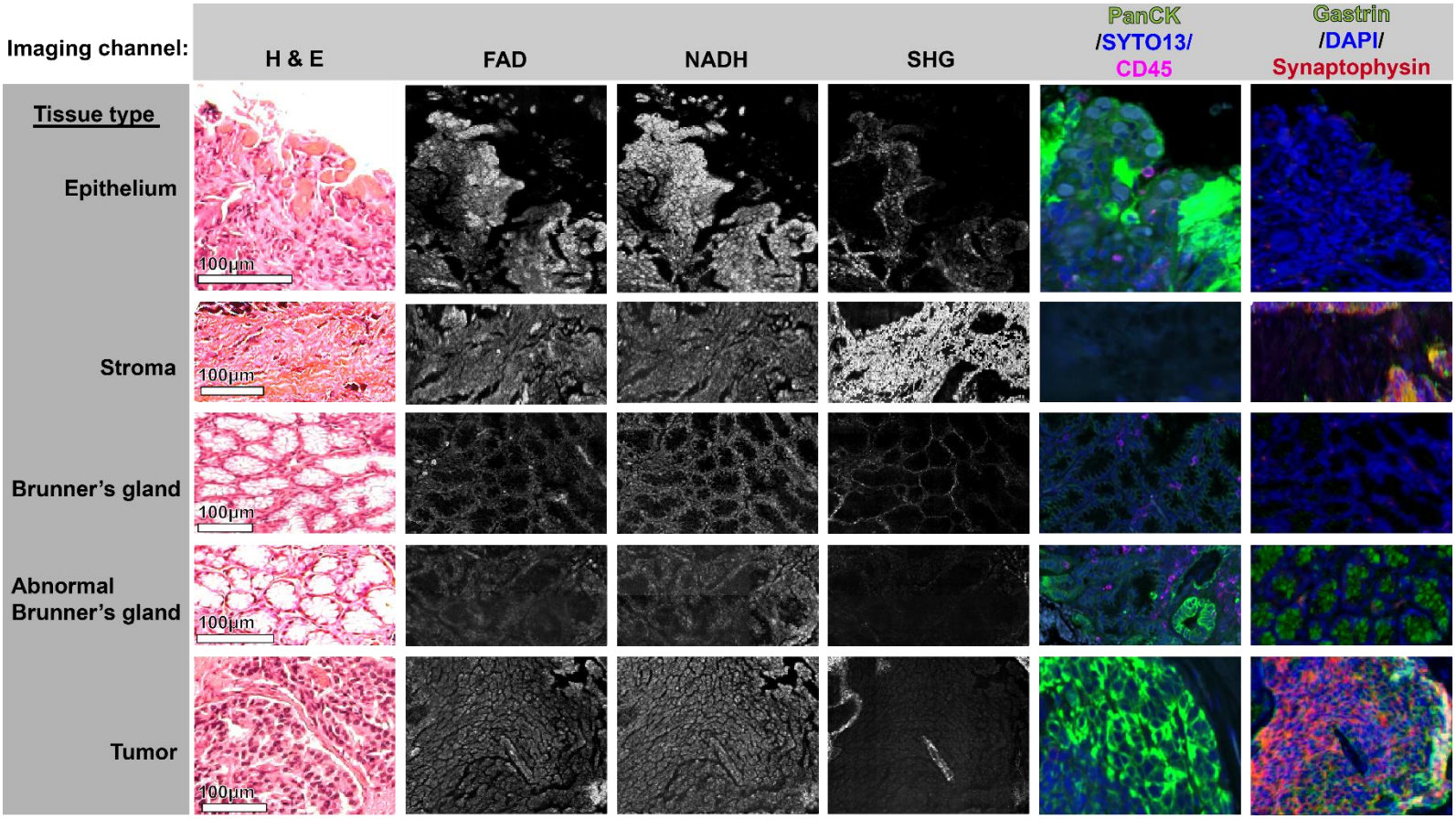


Figure 3: ROIs of the different tissue types showing examples of differences in morphology and IHC staining used for classification. The regions of gastrinoma tumor were verified through gastrin and synaptophysin staining. Regions of collagenous stroma produced a strong SHG signal with a distinct fibrous morphology. Villi and lamina propria were classified together and determined to be regions of the sample obviously oriented towards the lumen in-situ and/or strongly stained by the epithelial marker PanCK and having the characteristic appearance of villi. Brunner's glands had a characteristic glandular appearance in 2PEF, SHG, H&E and IHC stained images, while abnormal Brunner's glands had an atypical expression of epithelial markers that shown with PanCK staining.

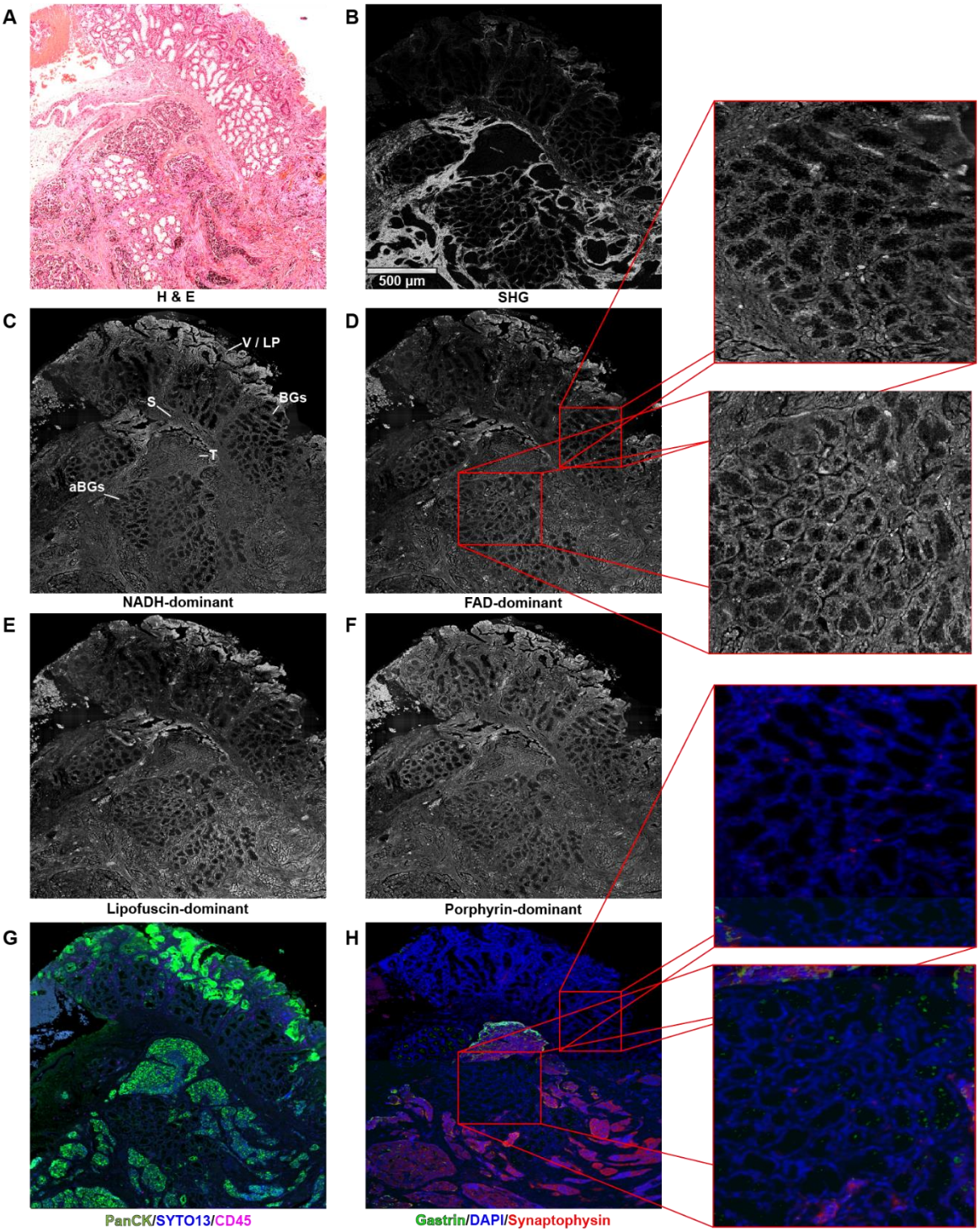


Figure 4: H & E (A), SHG (B), 2PEF channels (C-F), and IHC-stained images (G,H) of a duodenal gastrinoma specimen. Tissue classes are labeled in the NADH-dominant channel image (C), V / LP = villi / lamina propria, BGs = Brunner's glands, S = stroma, aBGs = abnormal Brunner's glands, T = tumor. Gastrin staining (H) is specific for gastrin-expressing cells, while synaptophysin is specific for neuroendocrine tumors. PanCK stain (G) is specific for epithelial markers, SYTO stains DNA, and CD45 for lymphocytes (immune cells). aBGs were classified based on their abnormal expression of epithelial marker and gastrin which was determined with PanCK and gastrin staining. Zoomed views

show BGs and aBGs imaged with the FAD-dominant 2PEF channel and gastrin/DAPI/synaptophysin staining. The H & E images of the DGAST samples were used to assess tumor morphology and as landmarks for regions of tumor neighboring other tissue types of interest for comparing the measurable autofluorescence.

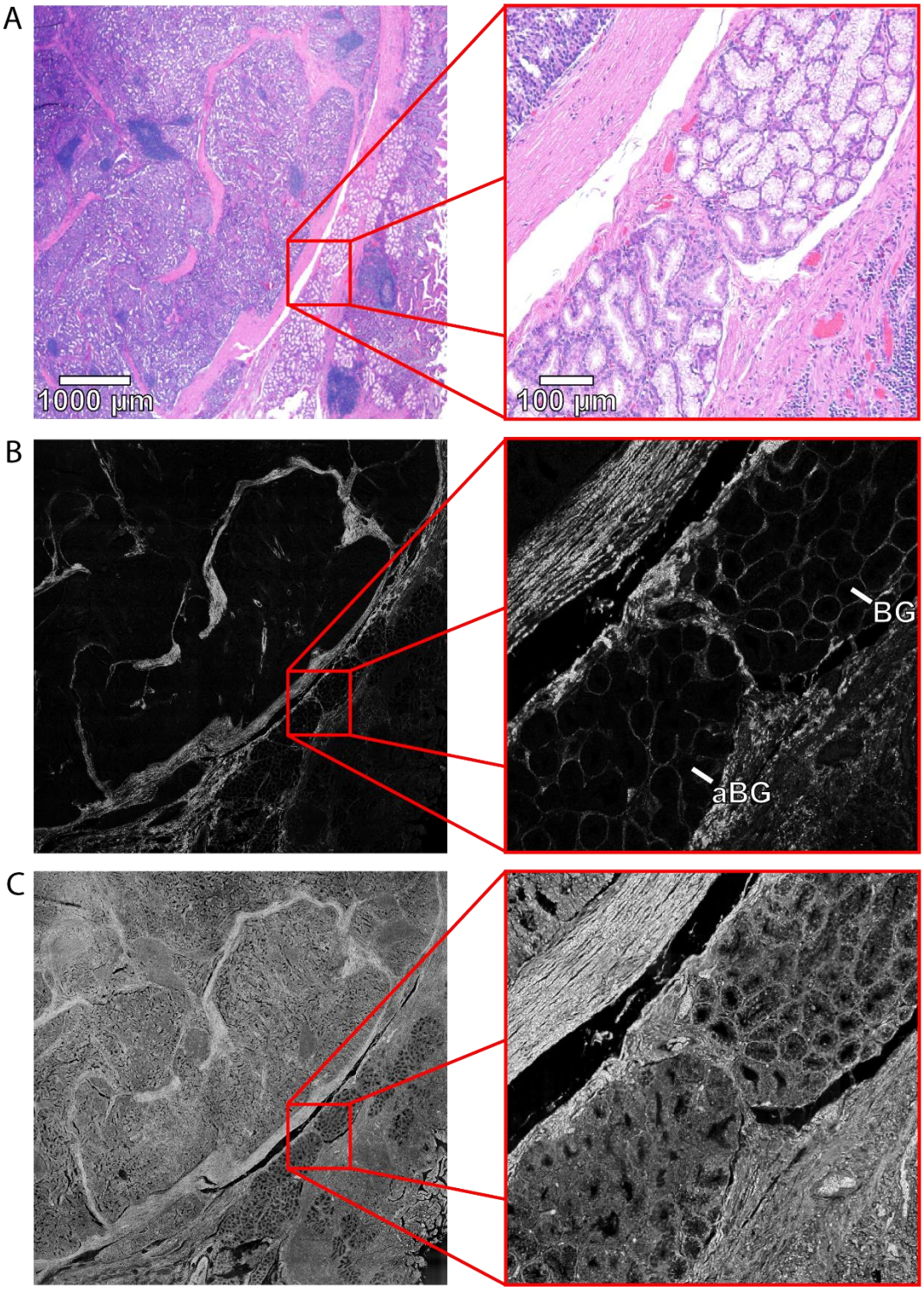


Figure 5: H&E (A), SHG (B) and FAD-dominant (C) channel images of DGAST tumor with zoomed views showing the border between abnormal (aBG) and normal Brunner's glands (BG).

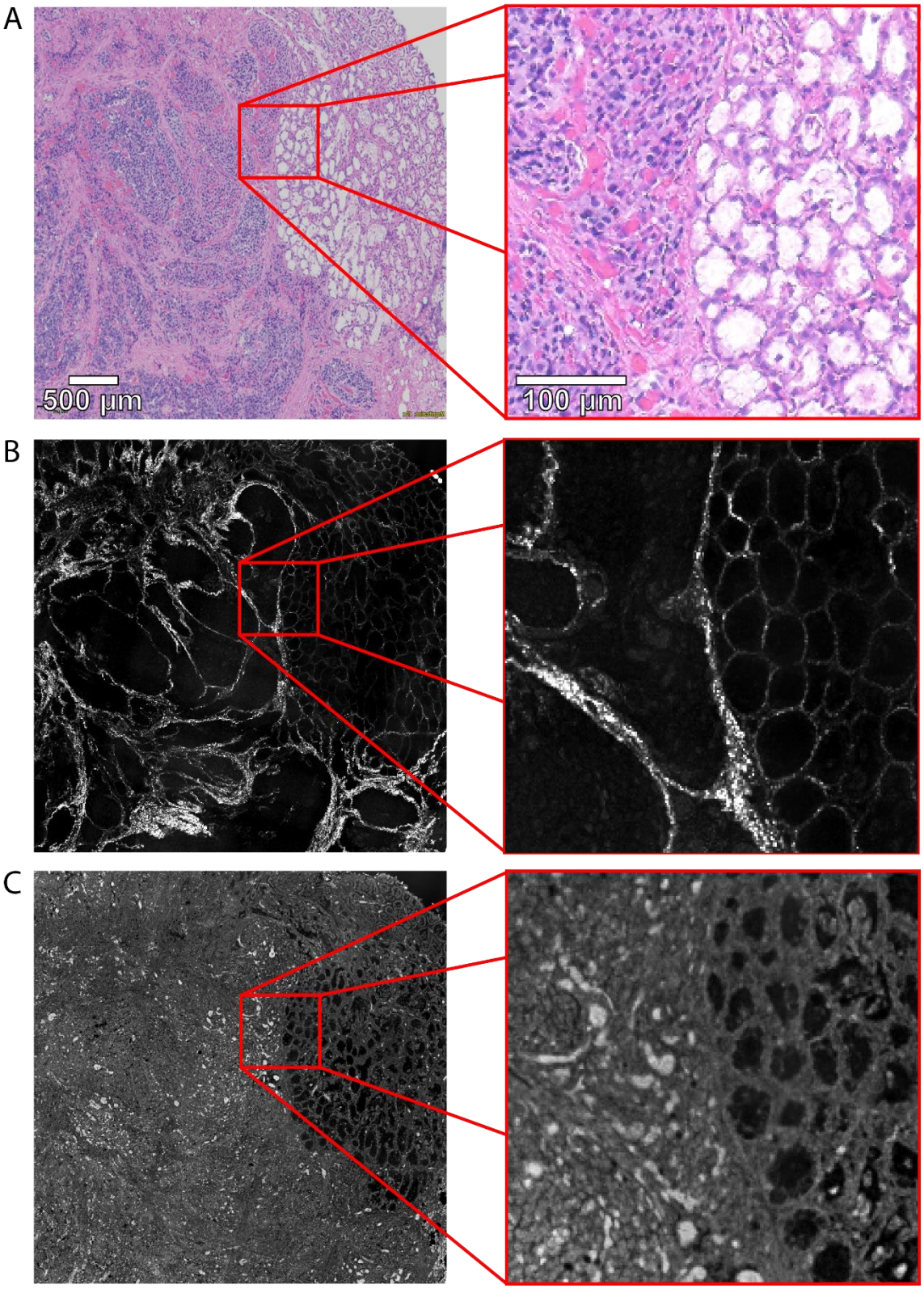


Figure 6: H&E (A), SHG (B) and lipofuscin-dominant (C) channel images of DGAST tumor with zoomed views showing the border between tumor (T) and Brunner's glands (BG).

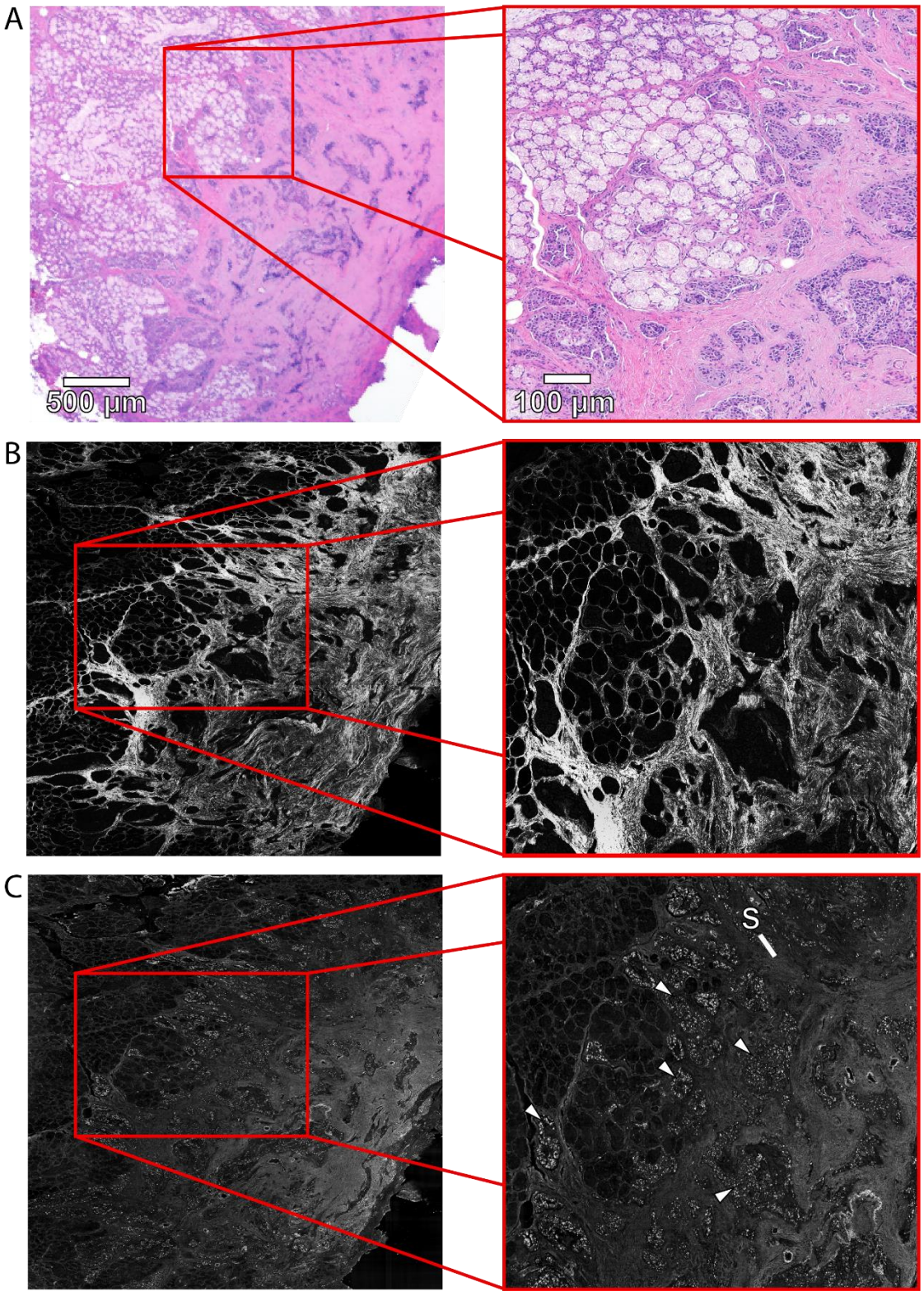


Figure 7: H&E (A), SHG (B) and porphyrin-dominant (C) channel images of DGAST tumor with zoomed view showing the highly disorganized stroma (S) with the characteristic trabecular appearance surrounding the diffuse clusters of tumor (marked with arrows in C).

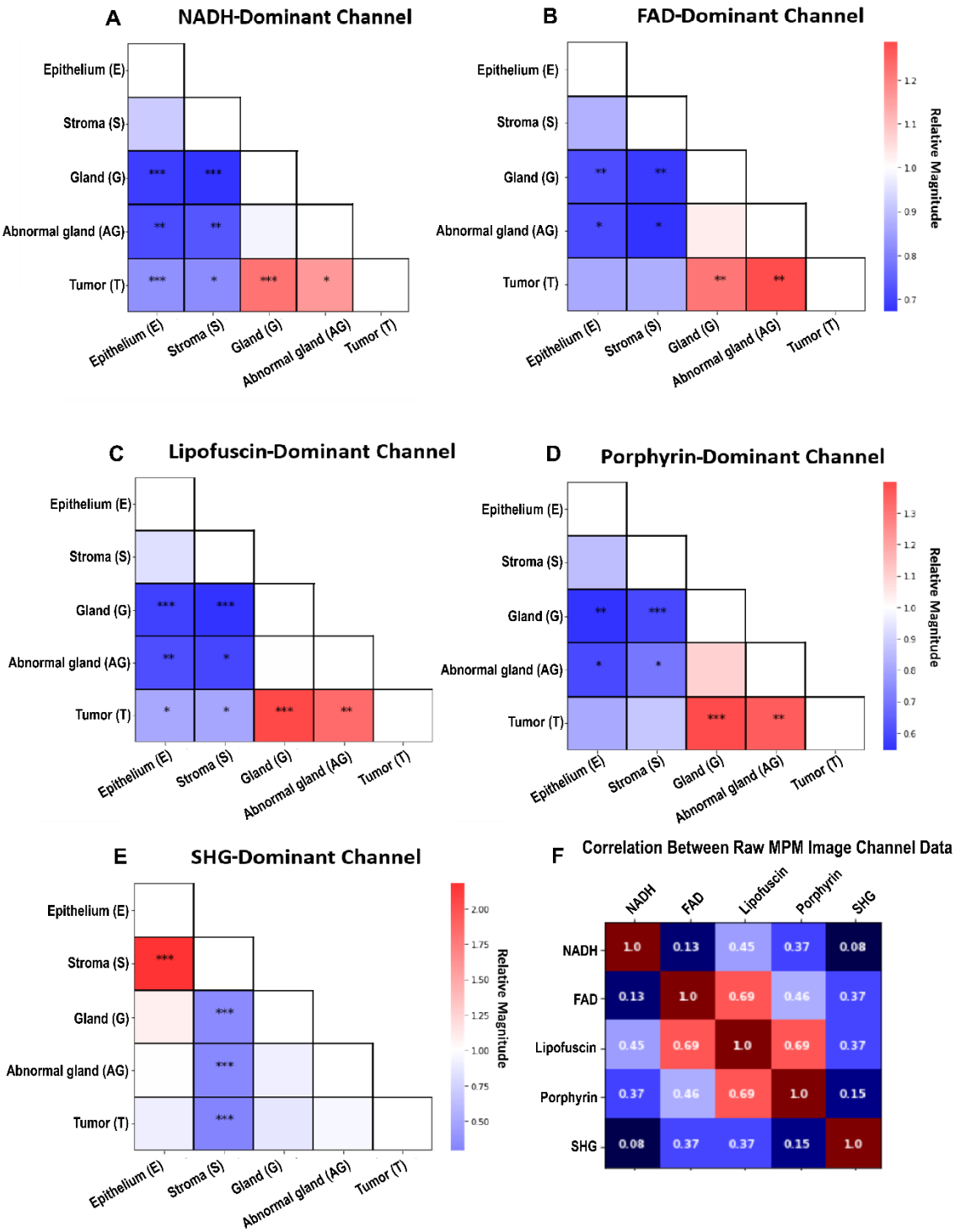


Figure 8: Heat matrices of the relative signal intensity comparison between tissue classes post-normalization by the lipofuscin imaging channel. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. The correlation matrix (F) shows a lower degree of correlation between the lipofuscin, FAD, and porphyrin channels as expected. This does result in a decrease in the degree of difference between the tissue classes for the FAD and porphyrin-dominant channels. A significant difference between the tumor and BGs still exists for the porphyrin-dominant channel and the tumor, stroma, and normal/abnormal glands for the FAD-dominant channel.

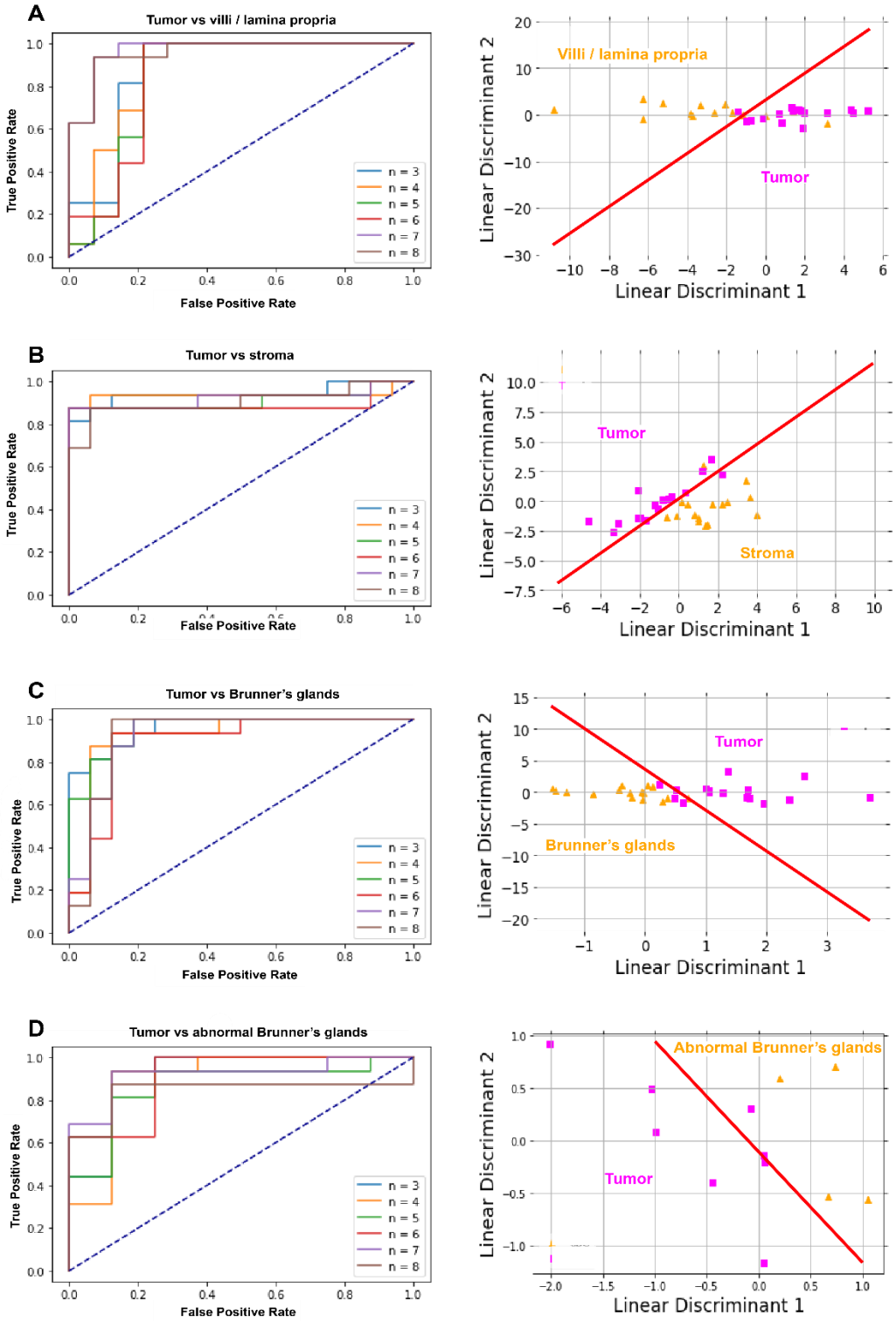


Figure 9: Receiver operating characteristic curves of the LDA classifier and LDA projections with hyperplane plotted in red distinguishing between tumor and BGs, aBGs, VLP, and stroma. n = number of features used to generate classifier. LDA projections for the tumor vs BG, stroma, and VLP classifier are shown with classification using 6 features, the tumor vs aBG plot is with 4 features used for classification.

SUPPLEMENTAL MATERIAL

Supplemental 1: Equations defining the texture features extractable from image gray-level co-occurrence matrix (GLCM)

1. *Angular Second Moment* :

$$f_1 = \sum_i \sum_j \{p(i,j)\}^2$$

2. *Contrast* :

$$f_2 = \sum_{n=0}^{N_g} n^2 \{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j))\}$$

3. *Correlation* :

$$f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

4. *Sum of Squares : Variance*

$$f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j)$$

5. *Inverse Difference Moment* :

$$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$$

6. *Sum Average* :

$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i)$$

7. *Sum Variance* :

$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$$

8. *Sum Entropy* :

$$f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) log\{p_{x+y}(i)\}$$

9. *Sum Entropy* :

$$f_9 = -\sum_i \sum_j p(i,j) log(p(i,j))$$

10. *Difference Variance* :

$$f_{10} = variance\ of\ p_{x-y}$$

11. *Difference Entropy* :

$$f_{11} = -\sum_{i=0}^{N_{g-1}} p_{x+y}(i) log\{p_{x-y}(i)\}$$

12. , 13. *Information Measures of Correlation* :

$$f_{12} = \frac{HXY - HXY1}{max\{HX, HY\}}$$

$$f_{13} = (1 - e^{[-2.0(HXY2 - HXY)]})^{1/2}$$

$p_x(i)$ = *i*th entry in the marginal-probability matrix created by summing rows of

$p(i,j) = \sum_{j=1}^{Ng} P(i,j)$

N_g = number of distinct gray levels in quantized image,

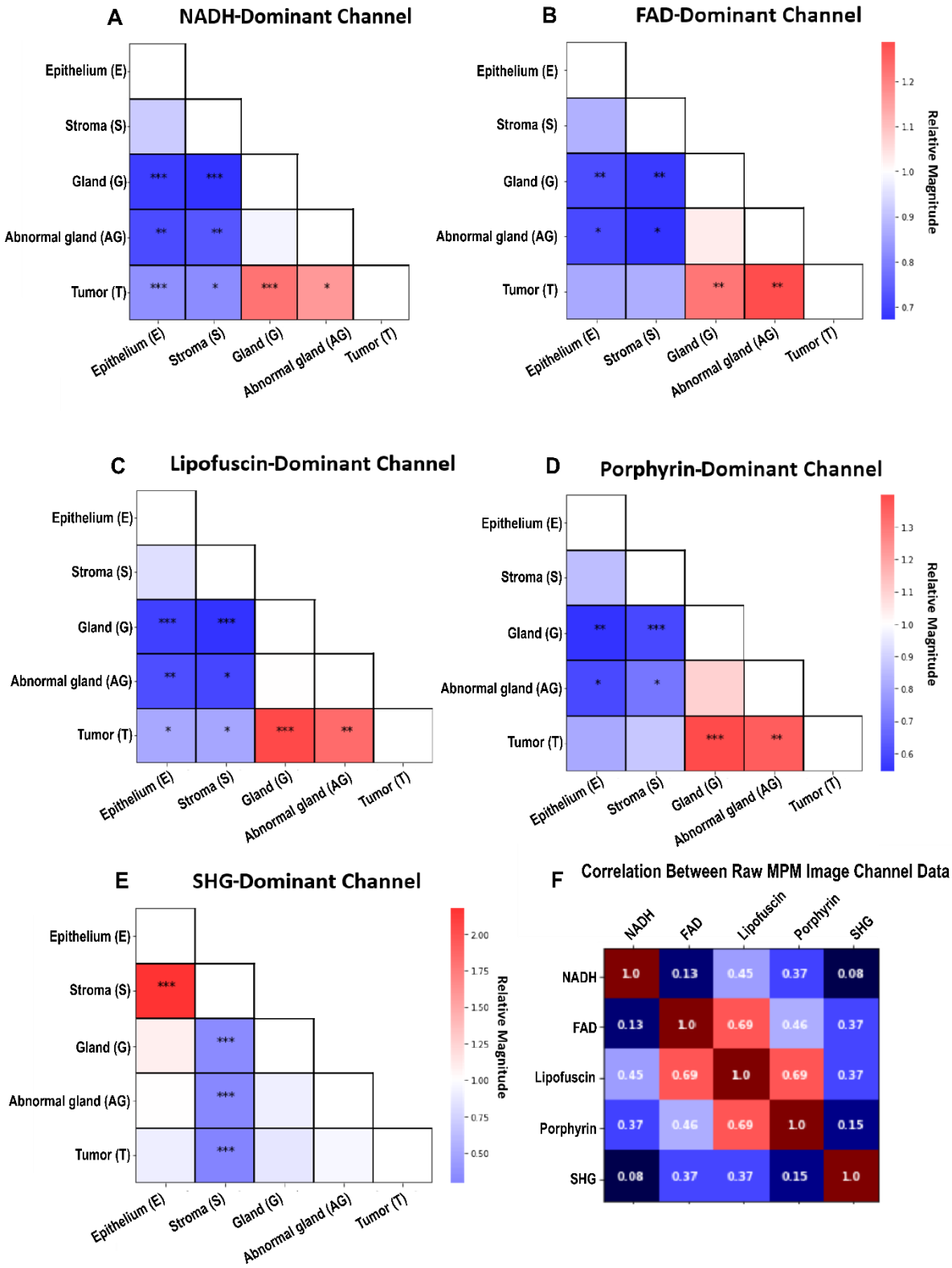
$p_y(j) = \sum_{i=1}^{Ng} p(i,j)$,

$p_{x+y}(k) = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j),\ i+j=k,$ $k=2,3,..2N_g$,

$p_{x-y}(k) = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j),\ |i+j|=k,$ $k=0,1,..N_g-1$,

μ_x, μ_y, σ_x, σ_y are the means and standard deviations of p_x and p_y.

SUPPLEMENTAL MATERIAL



Supplemental 2: *Heat matrices of the relative raw signal intensity comparisons between each tissue class. * = p < 0.05, ** = p < 0.01, *** = p <0.001. The matrices are interpreted as the row divided by the column. The correlation matrix (F) shows a high positive correlation between the lipofuscin, FAD, and porphyrin-dominant image channels prior to normalization.*

SUPPLEMENTAL MATERIAL

Wavelength (nm)	Peak Power (mW)	Average Power (mW)
700	26.77	19.96
800	51.15	46.48
830	48.37	45.06
880	34.88	32.78
920	22.36	21.24

Supplemental 3: *Power measurements of the MaiTai Ti:Sa laser at the specimen site. Power measurements were made with a ThorLabs S120 standard photodiode power sensor and PM100USB power and energy meter with laser parameters the same as those used during data collection as described in the METHODS section. Scan frame time was 0.63 seconds for a FOV of 328 μm^2 Mean power measurements calculated over continuous laser exposure for a duration of 1 second.*