Computational Analysis and Classification of SHG Images of Cancerous Pancreatic Tissue Based on Collagen Fiber Alignment

Gerren Welch
COMPUTATIONAL ANALYSIS AND CLASSIFICATION OF SHG IMAGES OF CANCEROUS PANCREATIC TISSUE BASED ON COLLAGEN FIBER ALIGNMENT

by

Gerren Welch

A Thesis Submitted in Partial Fulfillment
Of the Requirements for a Degree with Honors
(Physics)

The Honors College
University of Maine
May 2021

Advisory Committee:
Karissa Tilbury, Assistant Professor of Biomedical Engineering, Advisor
Robert Bowie, Professor of Practice in Chemical and Biomedical Engineering
Kathleen Ellis, Lecturer of English and Preceptor in the Honors College
Andre Khalil, Associate Professor of Biomedical Engineering
Michael Wittmann, Professor of Physics
ABSTRACT

Pancreatic cancer is a deadly disease, with a low five-year survival rate partly due to the difficulty in diagnosing the cancer early in its development, as it shares symptoms with more common and less lethal conditions. Using Second Harmonic Generation (SHG) microscopy and computer analysis, our knowledge of the biophysics of the pancreatic tumor microenvironment increases which may lead to the development of more effective therapies. In collaboration with Maine Medical Center Research Institute (MMCRI), we have identified 20 pancreatic cancer patients. In these 20 pancreatic cancer patients, Dr. Jones, a pathologist at MMCRI has identified normal adjacent pancreas, fibrotic pancreas tissue, and tumorous pancreas tissue. Using SHG imaging microscopy with an 890 nm excitation laser and collection via a 445/20 bandpass filter in the forward direction, we imaged the collagen primarily around pancreatic ductal structures, which studies have shown to be common tumor origin sites. OrientationJ, a Java plugin in ImageJ, uses a structure tensor to quantify morphological changes in the collagen fibers within the extracellular matrix. In this analysis, the quantitative orientation measurement function built into OrientationJ provides orientation and average coherency values for regions of interest selected by the user. Collagen within the ECM generally has a random, basketweave pattern, but in other cancerous tissues there are studies which show a correlation between progression of cancer and the increased aligning of the collagen fibers. We found that quantitative biophysical alterations that are distinct between cancerous, fibrotic, and normal adjacent tissues, but the results were inconclusive. With a larger sample size, the results might demonstrate the pattern in the data more conclusively.
ACKNOWLEDGEMENTS

I would like to take this time to acknowledge the opportunity granted to me by the Center of Undergraduate Research in funding this research and allowing me to focus on my education. I would also like to thank my thesis advisor Karissa Tilbury and my academic advisor Michael Wittmann for their continued support and advice throughout my college career.
# TABLE OF CONTENTS

INTRODUCTION 1
  Troubles with Detection and Diagnosis 3
  Clinical Imaging Strategies 4
  Pathology 5

CHAPTER 1: THE TUMOR MICROENVIRONMENT 10
  The Role and Structure of Collagen 13
  Means of Collagen Detection and Analysis 13

CHAPTER 2: UNDERSTANDING AND USING SECOND HARMONIC GENERATION MICROSCOPY 16
  Second Harmonic Generation Apparatus 19

CHAPTER 3: QUANTIFICATION OF COLLAGEN MORPHOLOGY 24
  Reasons for Moving Away from CT-Fire 24
  The Use of OrientationJ 26
  The Analytical Process 30
  Results 34
  Why use ANOVA Tests and T-tests? 37

CHAPTER 4: CONCLUSIONS AND BASIS OF FUTURE WORK 40

REFERENCES 41

AUTHOR’S BIOGRAPHY

Error! Bookmark not defined.
LIST OF FIGURES

Figure 1: Exocrine and Endocrine Pancreas by OpenStax College 1
Figure 2: Example of H&E Stain 6
Figure 3: Pancreatic Cancer Tissue Types 7
Figure 4: Tumor Microenvironment at a Glance 11
Figure 5: Jablonski Diagrams of Two Photon Excitation and SHG Microscopy 15
Figure 6: Laser Scanning SHG Optical Microscope 18
Figure 7: Imaging Workflow 21
Figure 8: CT-Fire Applied to a Test Image 23
Figure 9: CT-Fire Problem 24
Figure 10: OrientationJ Visual Analysis Mode 27
Figure 11: OrientationJ Visual Analysis Mode Coherency Map 28
Figure 12: Attempted Sliding Window Approach 29
Figure 13: Contrast Enhanced Image in OrientationJ Measure 32
Figure 14: Average Coherencies with Weighting Applied 33
Figure 15: Unweighted Average Coherencies 35

LIST OF TABLES

Table 1: American Cancer Society SEER Survival Rates of Pancreatic Cancer 2
Table 2: Important Values From Coherency Analysis 35
<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td>Polarization Dependence of Second Harmonic Generation</td>
<td>17</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>Crystal Nonlinearity for the Pockels Effect</td>
<td>19</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>Definition of the Weighted Inner Product</td>
<td>25</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>OrientationJ Directional Derivative</td>
<td>26</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>Inner Product Manipulation</td>
<td>26</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td>OrientationJ Structure Tensor</td>
<td>26</td>
</tr>
<tr>
<td><strong>7</strong></td>
<td>Coherency Calculation</td>
<td>27</td>
</tr>
<tr>
<td><strong>8</strong></td>
<td>Mean Square of Treatments</td>
<td>37</td>
</tr>
<tr>
<td><strong>9</strong></td>
<td>Mean Square of Errors</td>
<td>37</td>
</tr>
<tr>
<td><strong>10</strong></td>
<td>Equation for the F-Value</td>
<td>38</td>
</tr>
</tbody>
</table>
INTRODUCTION

Pancreatic cancer is a deadly disease which is difficult to detect and diagnose. In discussing pancreatic cancer, a modicum of knowledge regarding the form and function of the pancreas is necessary. The pancreas is an organ situated behind the stomach, with about 95% of pancreatic tissue being made up of exocrine tissues which connect to the common bile duct (connected to the liver and depositing its contents into the small intestine). The remaining 5% is endocrine tissue responsible for the production of insulin and glucagon.

Figure 1: Exocrine and Endocrine Pancreas
The exocrine functions are performed as the epithelial tissues lining the pancreatic ducts produce enzymes necessary for digestion such trypsin, amylase, and lipase. As their exocrine designation suggests, the products of these tissues are passed to the common bile duct, which connects to the duodenum where they can enter the small intestine to digest food. As for endocrine function, the clumps of cells called islets of Langerhans deposit hormones such as insulin and glucagon directly into the bloodstream to regulate blood sugar throughout the body.4

Pancreatic cancer is a deadly disease, with a low five-year SEER survival rate of 10% across all types partly due to the difficulty in diagnosing the cancer early in its development.6 This means that an individual diagnosed with pancreatic cancer has a 10% chance to survive five years after diagnosis relative to a healthy person (who by this metric would have a 100% chance of survival).

<table>
<thead>
<tr>
<th>SEER Stage</th>
<th>5-year Relative Survival Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Localized</td>
<td>39%</td>
</tr>
<tr>
<td>Regional</td>
<td>13%</td>
</tr>
<tr>
<td>Distant</td>
<td>3%</td>
</tr>
<tr>
<td>All SEER stages combined</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 1: American Cancer Society SEER Survival Rates of Pancreatic Cancer6

According to the American Cancer Society, pancreatic cancer is the deadliest of common cancer types. There are several factors contributing to the low survival rate, such as vague
symptoms, late stage detection, and poor response to treatments. Currently treatments for pancreatic cancer predominantly rely on surgical resection and adjuvant chemotherapy, though other treatment methods including radiation therapy and immunotherapy with varying efficacy are also used in some cases. While surgery and chemotherapy can prove effective, they are best used early in the development of cancer, making the lives of pathologists more difficult. The product of this research might make the lives of these pathologists easier, allowing for the easier and more reliable diagnosis of cancer.

Troubles with Detection and Diagnosis

Due to the anatomical location of the pancreas and its far-reaching effects within the body, pancreatic cancer symptoms are difficult to pinpoint and broadly applicable to many less severe illnesses, and include the likes of 1) jaundice, 2) abdominal/back pain, and 3) new or worsening diabetes. These and the other potential symptoms are too broadly applicable to be useful in determining that an illness is in fact pancreatic cancer and not some more common and less serious ailment. Furthermore, there are no biochemical markers specific to pancreatic cancer. Diagnosis is often delayed significantly as a result, since many other conditions need to be eliminated from consideration before pancreatic cancer can be diagnosed.

The lack of specific biochemical markers means that tests such as blood tests are not capable of differentiating pancreatic cancer from other cancers. If signs or symptoms arise, a doctor may benefit from inquiring about the patient’s medical history. Certain behaviors and conditions such as cigarette smoking and excessive alcohol consumption, being overweight, and having diabetes have been linked to pancreatic cancer.
Approximately 5–10% of pancreatic cancer patients reporting family history of pancreatic cancer, making genetics a significant factor.8

Earlier detection would make treatments considerably more reliable, and significantly improve the outlook for patients. While improving the ability of the pathologist to make accurate diagnoses quickly would not speed up the process of eliminating symptoms, it could still allow for somewhat faster diagnoses and be less dependent on pathologist skill. The search for biomarkers of pancreatic cancer research has extended to the extracellular matrix surrounding the cancer cells in recent years3,9–11. Based on research done in other cancer types including ovarian and breast cancers, collagen remodeling in the extracellular matrix has been demonstrated to be associated with cancer progression, with an increase in abundance and alignment between fibers as cancer advances.2,3,9 In this research we made use of Second Harmonic Generation (SHG) microscopy to target the collagen fibers in the extracellular matrix surrounding the pancreatic cancer cells. By utilizing the structural dependence of SHG microscopy to isolate the collagen fibers, we endeavored to develop a method of reliably quantifying the changes made by collagen remodeling in pancreatic cancer, which may allow for earlier diagnosis in future work.

Clinical Imaging Strategies

An individual with any common risk factors described above might warrant the use of various imaging techniques to either spot cancerous growths directly or indirectly. Computed Tomography (CT) scans and Magnetic Resonance Images (MRIs) allow pathologists to see the region of interest with some detail and will show cancerous tissues
directly. Positron Emission Tomography (PET) scans require the use of a radioactive sugar molecules likely to be absorbed for energy by fast-growing cancer cells to show the shape of probable cancer tissues. Ultrasounds can be used as they are cost-effective and minimally invasive, but they typically do not allow for the same image quality CT scans can offer, and are thus less favored for the purpose of pancreatic cancer detection. Cholangiopancreatography and angiography can also be used to detect the presence of tumors by showing blockages in the pancreatic ducts and blood vessels within the pancreas respectively.\textsuperscript{12}

**Pathology**

The imaging methods above in combination with the patient’s medical history may provide a strong indication regarding the presence of pancreatic cancer, but conclusive results generally require the performance of a biopsy, the gold standard of pathological diagnosis. A tissue sample from the suspected tumor can be collected through the skin by a long, thin needle (percutaneous biopsy) or by use of an endoscope (endoscopic biopsy).\textsuperscript{8,12} After this a pathologist examines thinly sliced biopsy material stained with the dyes hematoxylin and eosin. This method is known and H&E staining, and is used as the dark blue hematoxylin dye stands out in contrast to the reddish eosin dye. These dyes bind to different structures, resulting in nuclei stained blue and protein structures stained in red, providing a high contrast and facilitating the identification of features within the sample. An example of this is depicted in Figure 2, where fibers of the extracellular matrix are shown in a reddish pink color and cell nuclei and contents of ductal cells appear in a dark blue or purple.
Figure 2: Example of H&E Stain

Using these dyes, the pathologist searches for several characteristics of cancerous tissue. Among the most prominent of these is the nucleus to cytoplasm ratio of the cells in the image. A high nucleus to cytoplasm ratio is an indicator that there are many immature cells in an area, which is frequently a sign of cell growth, a hallmark of cancer. In the case of pancreatic cancer, a majority of these abnormal cells are epithelial cells originating in the ducts.\textsuperscript{4,13} Since cancer cells do not undergo apoptosis, or programmed cell death, their reproduction remains unchecked. They divide infinitely so long as nutrient supplies allow, and will spread throughout the body if given the opportunity. The result of this unchecked proliferation is a tissue densely populated by cells surrounding the epithelial tissue of the duct, while the surrounding area is dominated by the collagen fibers grown from cancer associated fibroblasts.
Figure 3: Pancreatic Cancer Tissue Types

- Normal Adjacent
- Fibrotic Tissue
- Cancerous Tissue

Key Features:
- Cells interspersed throughout
- Ductal Structure
- Beginnings of cancerous fiber alignment
- High Nucleus to Cytoplasm Ratio
- Growth of cells on edge of duct
- Visibly aligned collagen fibers

Nucleus to Cytoplasm Ratio less than cancerous tissue
In healthy tissues, the extracellular matrix (ECM) is disordered in a sort of “basket-weave” and many cells within have a low nucleus to cytoplasm ratio indicating cell maturity as depicted in Figure 3. The cells are also interspersed throughout the tissue, as opposed to cancer tissues which have clumps of cells surrounded by ECM.

A precursor to full tumorigenesis in pancreatic cancer is fibrotic tissue, characterized by the increased density and ordering of the collagen caused by fibroblast overactivity induced by chemical signals from the malignant cells. Here the beginnings of multicellular spheroids start to form, replacing the cell monolayers more common in healthy tissue. These spherical malignant structures are capable of increasing tumor drug resistance. These spheroid structures typically contain hypoxic regions, which are effective in stopping the penetration of immune cells and treatments. Fibrotic tissues are sometimes quite difficult to distinguish from cancerous tissues, and provide a challenge for pathologists attempting to diagnose a patient. In Figure 3, fibrotic tissue and cancer are separated by the red line with the goal of distinguishing them from each other visibly.

In cancerous tissue, there is a great abundance of relatively orderly collagen, and the average nucleus to cytoplasm ratio of most of the cancerous cells is larger than that of healthy tissues. We also see a fully formed iteration of the aforementioned spheroid structures in Figure 3.

Due to the difficulty of diagnosis, pancreatic cancer is often found in later stages, making patient outlook rather bleak. This brings us to the primary motivation for this research. By developing a reliable quantitative process for identifying cancer-driven collagen remodeling, earlier and more certain diagnosis of pancreatic cancer might be achieved. To improve patient outlook, this research is a stepping stone towards new
treatment methods requiring reliable quantitative methods of detecting cancerous changes in pancreatic tissues. The remainder of the thesis is divided into four chapters. First, the tumor microenvironment is explored, giving an overview of the changes this project intends to track. Second, the theoretical basis for the physical processes dictating second harmonic generation is provided. Third, the methods of collagen quantification and the results of the experiment are discussed. And finally, the implications of these results are considered, as well as the future work that can build upon this.
In the pancreas and elsewhere, tumors rely on the extracellular matrix (ECM) to spread around the body. This ECM acts as a scaffold for the cells and also plays an active role in the progression of cancer.\textsuperscript{16,17} It contains chemical growth factors, which interact with cell surface receptors and affect mechanical properties of the tissue such as elasticity and compressive strength.\textsuperscript{17} Collagen makes up a large portion of the ECM, and conveniently with its tri-helical structure it is a suitable target for second-harmonic generation (SHG) microscopy. Collagen is abundant in both healthy and cancerous tissues, but tumors act as “rogue organs” which reshape the extracellular matrix to their own purposes of proliferating. This rogue organ description as used by Balkwill et al. is valuable in understanding the tumor microenvironment as a complex system depicted in Figure 4, containing not just malignant cells but also a support network of cells ranging from fibroblasts to immune cells to things such as adipocytes (fat cells).\textsuperscript{17}

This support network of cells is chemically coerced into supporting the malignant tissue as they would healthy growing tissue with signals akin to those used in wound healing and inflammation.\textsuperscript{17} Since SHG microscopy is not sensitive to cells themselves, the most relevant parts of the rogue tumor organ are the cancer-associated fibroblasts (CAFs) displayed near the bottom of Figure 4. These cells have play two key roles: They produce growth factors vital for cell growth and division, which cancer cells co-opt for
their own growth. They are also producing collagen fibers, which serve the dual purpose of facilitating the spread of malignant cells and inhibiting the access of immune cells (and agents used in chemotherapy) to the malignant tumor cells.17
Figure 4: Tumor Microenvironment at a Glance
The Role and Structure of Collagen

Collagen is the single most abundant protein in the animal kingdom, and its primary purpose is to allow tissues to withstand stretching. As mentioned before, collagen fibers also play a role in the facilitation of tumor progression, by containing the growth factors necessary for tumor expansion which were generated by the CAF cells and . Their triple-helical structure is dictated by the abundance of three amino acids, being glycine, proline, and hydroxyproline. These make up three polypeptide chains which twist together to form a structure with great tensile strength, which is used primarily to bind tissues together. This triple helical structure makes collagen fibers noncentrosymmetric, allowing for the use of SHG microscopy to provide a highly detailed image and isolate the fibers for study.

Means of Collagen Detection and Analysis

There are numerous ways to detect and quantitively analyze collagen. In terms of detection, there are the following. Spatial light interference microscopy (SLIM) is one method being researched. It is cheaper and faster to use than SHG, but is sensitive to structures other than collagen, which introduces additional image components which a computer would need to sift out of the image before it could analyze the collagen fiber morphology as desired here. The LC-PolScope is an enhanced form of polarized light microscopy which has also been suggested as a means to image collagen fibers, but the technology was relatively new in 2020 when the Fanous et al. paper was published. Other forms of two-photon excitation microscopy (TPEF) which rely on fluorescence are also capable of imaging collagen fibers, but run the risk of damaging and photobleaching
the sample, as well as being mechanically more complex than SHG setups due to their requiring multiple photon sources (i.e. lasers).

For these reasons above and the ability of SHG to isolate the collagen fibers from the rest of the tumor microenvironment, SHG remains well-suited for the task of studying these collagen-based phenomena. The vast majority of pancreatic cancers start in ductal regions,\textsuperscript{13} which is why pathologists focus on these regions when they search for pancreatic cancer. Considering that CAFs can derive from endothelial cells, smooth muscle cells and myoepithelial cells\textsuperscript{17} common in exocrine glands (e.g. the pancreas), they are a prime target for further scrutiny in regards to pancreatic cancer, as are their products.

As for analysis, there are numerous methods which have been proposed and used for the quantification of collagen fibers. Currently free-to-use programs which perform this task include FibrilTool, OrientationJ, Cytospectre, CurveAlign, and CT-Fire.\textsuperscript{11} FibrilTool borrows concepts from liquid crystal physics to calculate coherency and alignment but requires significant user input which makes it time-consuming to use for larger sample sizes.\textsuperscript{11,20} OrientationJ makes use of a Gaussian function in a manner commonly used in image processing, and will be discussed at length later in the paper. Cytospectre uses a Fourier transform technique to find orientation and anisotropy information, but is only capable of providing global measurements for an image.\textsuperscript{11,21}

Both CurveAlign and CT-Fire were created by the Eliceiri lab at University of Wisconsin for the task of fiber quantification, though they approach this in different ways. CurveAlign makes use of curvelet transforms to enhance fiber edges and reduce noise in the image, and CT-Fire uses an individual fiber tracking algorithm in order to
provide data on a fiber-by-fiber basis, rather than across an entire image or region of interest.\textsuperscript{11} CT-Fire was the original method slated for use in this research, though due to issues discussed in the quantification chapter OrientationJ was eventually used. Both of these programs are described in greater depth at that point.

Out of these methods, we opted to use SHG microscopy and both CT-Fire and OrientationJ. Because fibrosis and collagen remodeling are hallmarks of pancreatic cancer,\textsuperscript{3} isolating this effect using the above techniques could prove tremendously useful. By quantitatively tracking the dynamics of collagen remodeling we could gain valuable insights into the early growth of pancreatic cancer and provide pathologists with the means to detect it reliably and reduce analysis time.\textsuperscript{11}
Spectroscopy is the study of interactions between matter and light, and thus is integral to the techniques mentioned in Chapter 2, including SHG. Key interactions that fall under this topic include absorption and scattering. Absorption is the process by which matter takes in the energy of a photon, and scattering is the process of photons colliding with matter. SHG is considered an inelastic scattering phenomenon, deriving the inelastic moniker from term used in classical mechanics describe a collision in which energy is lost to the system, as opposed to the mostly idealized elastic collision where energy is not lost to the system in question. In second harmonic generation, two lower energy photons convert to twice the incident frequency (half the wavelength) of an excitation laser upon striking a noncentrosymmetric object. It is a special case of two photon excitation microscopy (TPEF) and was first developed in 1961, but since 2002 with the seminal publication by Campagnola and colleagues, SHG has seen increased biological application.
Compared to other imaging methods, SHG has several advantages for research on biological samples. Chief among these are the lack of energy absorption into the sample evidenced by the virtual electronic state (dotted line in Figure 5), which allows researchers to avoid photobleaching the sample. In order to achieve this state the photons hitting the sample must be both coherent (matching phase) and simultaneous. SHG microscopy satisfies both of these requirements. By the nature of two-photon imaging, we also receive the benefit of a high signal to noise ratio, which generally aids in the production of clear images.

SHG can be treated theoretically by solving Maxwell’s equations directly for wave propagation in a nonlinear medium, but for the purposes of this project a summation of the second harmonic fields generated from the oscillations of the medium in question is a preferred method. The direct solution of Maxwell’s equations is best reserved for regularly shaped objects, and become quite complicated to solve for complex shapes such as those found in biological structures. In order to avoid significant
information loss by oversimplification, the shapes calculated need to be particularly small.

The summation method used by Freund, Deutsch, and Sprecher handles irregularly shaped biological structures more gracefully. Its caveat that the objects being imaged be embedded in an index-matching medium. This makes the phase differences of the various waves arriving at the detector plane dependent solely on the signal emitted from the sample itself. The requirement is met by the pancreatic tissues studied in this research. Dr. Tilbury uses a textbook version of Equation 1, which is similar to the situation-specific way Freund uses to express the general equation for the polarization of an electric field shown in Equation 1. This equation dictates all interactions of electromagnetic waves (e.g. light).

\[
P = \varepsilon_0 (\chi_1 E + \chi_2 E^2 + \cdots + \chi_n E^n)
\]

\(P\) is the second order polarization, \(\chi 2 = \beta * n\) where \(n\) = number of molecules, \(\beta_{ijk}(r)\) is the hyperpolarizability of the material (second order electric susceptibility), and the fields \(E^{(1)}_{n}(r)\) are the electric fields in the j and k directions of an orthogonal coordinate system.

The collagen fibers are noncentrosymmetric, which enables an SHG response. A centrosymmetric fiber would result in the destructive interference of the signal. Due to the squaring of the electric field term (or any even powered polarization term) and the phase matching involved, structures such as cells which have inversion centers do not produce SHG signals.
Second Harmonic Generation Apparatus

The apparatus required for SHG microscopy is more substantial than the basic theory might lead one to believe. In order to get photons of the correct wavelength and sufficient energy to strike the sample consistently and in an even fashion, there must be multiple intermediate steps. These must modulate the power for the laser and thus the energy of said photons, ensure a circular polarization of the beam at the focal point to cause the SHG effect evenly across the sample, and scan the beam across the sample.

![Diagram of SHG Apparatus](image)

To accomplish the necessary polarization in the pancreatic tissue samples, an apparatus modelled on the one depicted in Figure 6 was used. For a femtosecond excitation source, we used a mode-locked Titanium Sapphire laser (Coherent Chameleon Ultra II, Santa Clara, California) which was tuned to produce pulses at wavelengths of 890 nanometers.
To modulate the power of these pulses, the laser beam was sent through a Pockels cell (ConOptics, Danbury, Connecticut).

The Pockels Effect used by the Pockels cell is an electro-optic effect which results from the application of a direct current or low frequency electric field to a medium. Since crystals are the most common media used for producing this effect, it is customary to express the nonlinearity of the refractive index $n$ in a manner analogous to the susceptibility\textsuperscript{23} used in Equation 2.

\[ 2) \frac{1}{n^2} = \frac{1}{n_0^2} + rE + RE^2 \]

In Equation 2, $n$ is the induced refractive index of the crystal the laser travels through, $n_0$ is the refractive index of the crystal in the absence of an applied field, and the coefficients $r$ and $R$ are the linear and quadratic electro-optic coefficients, which are tensors reflecting the symmetry of the crystal. The Pockels Effect results from the linear term of Eq. 2, where $E$ is an applied DC field.\textsuperscript{23} This field shifts the electrons within the crystal, causing birefringence (alteration of the refractive index). When passed through the Pockels cell, the different components of the electric field experience different refractive indices, and stretch/contract accordingly, resulting in an elliptically polarized beam and power modulation.

After the Pockels cell, a quarter and half waveplate ensures that the laser beam will be of circular polarization at the focus, exciting the dipoles within the sample equally. From here the laser beam is scanned across the field of view via electronically-controlled mirrors within the scanning system (FluoView 300; Olympus, Center Valley, Pennsylvania). From there the beam travels through an upright microscope stand.
(BX50WI; Olympus, Center Valley, Pennsylvania), which directs the laser beam on the sample. At the focal plane, the beam is focused into a circular polarization at a power range of 5-50mW using a LUMPlanFLN 40x 0.8NA (Olympus, Center Valley, Pennsylvania) water immersion objective. From here the light is collected with a 0.9 NA condenser. SHG signal emitted from the sample is collected in the forward direction thanks to the geometry of the sample (a thin slice of tissue), as the signals travelling along the axis of propagation are most pronounced. The forward signal sent through a dichroic, which serves to direct the SHG signal to the photon detectors (Hamamatsu Type #7421-40) set to photon counting mode to record the forward and backward channels data. The forward channel was favored for the data analysis, recording the SHG signal filtered with a FF01-448/20 bandpass filter (Semrock, Rochester, New York). The backward channel was set up to record fluorescence using a FF01-582/64 nm bandpass filter (Semrock, Rochester, New York). Finally, the recorded data from the photon detectors is processed and stored as tiff files using the FluoView software associated with the scanning system. On the user end, the process of imaging pancreatic tissue slides is as described in Figure 7 and the following pages.
Figure 7: Imaging Workflow

Annotated slide scans were provided by a pathologist

Target regions were identified under brightfield conditions (e.g. ductal structure in cancerous tissue)

Target regions imaged using fluorescence and SHG microscopy simultaneously
Forty slides of pancreatic samples were intended to be imaged, with regions identified using bright-field images as healthy, fibrotic, and cancerous by Dr. Michael Jones, a pathologist from the Maine Medical Center. Of the original set of forty slides, twenty of these contain some combination of cancerous tissue and fibrotic or normal adjacent tissues, and the other twenty slides were healthy samples. Due to time constraints and alignment troubles which developed during the course of the experiment, only twenty-one of these slides were able to be fully imaged and thus analyzed. Of these, fourteen were cancer-containing slides, which were focused on for analysis as they contained tissues of all three types (normal adjacent, fibrotic, and cancerous) and comprised 128 regions of interest within them.

We proceeded to acquire the images, which consisted of the following steps summarized in Figure 7: 1) The slide was placed on the microscope, and drops of water were applied to the slide surface to allow for the use of the 40x water objective. 2) The region of interest was found under brightfield conditions before switching the microscope back to its SHG configuration. 3) The lights were shut off to minimize light interference, and the imaging commenced by scanning in the FluoView software. For each region of tissue of a given type, five to seven images were acquired of various features surrounding the pancreatic duct within the sample region in order to provide a large dataset for the statistical analysis. Some especially small regions had fewer images taken if there were relatively few significant features within. In the healthy slides (distinguished from normal adjacent regions), ten images were taken for each.
CHAPTER 3

QUANTIFICATION OF COLLAGEN MORPHOLOGY

Reasons for Moving Away from CT-Fire

This project was originally supposed to make use of CurveAlign and CT-Fire, a pair of plugins for ImageJ which makes use of curvelet transforms\textsuperscript{24} and individual fiber tracking algorithms\textsuperscript{11,25,26} in order to obtain fiber alignments in significant detail. Using test images, the results of CT-Fire appear promising. It is capable of recreating and measuring fibers in the image as shown in Figure 8.

![Figure 8: CT-Fire Applied to a Test Image](image)

The output can be manipulated by numerous configurable settings, allowing the user great flexibility in the sensitivity of the analysis. In our attempts to use CT-Fire on the pancreatic images we collected we encountered problems, which are shown in Figure 8.
Unlike the test images, CT-Fire struggled to distinguish collagen fibers from their surroundings, and attempted to place fibers in empty spaces to fill the whole image.

This problem persisted through dozens of settings adjustments, and it is not entirely clear what caused the insertion of fibers into empty space as depicted in Figure 9. Multiple pancreatic images were tested, and aside from the test image, all exhibited this issue. CT-Fire also proved time-consuming to use due to its in-depth analysis, and would require a powerful processor to speed the process. Due to time constraints, the decision was made to abandon CT-Fire and use OrientationJ instead, as OrientationJ was capable of producing usable results in seconds as opposed to the overly sensitive CT-Fire which could take over a minute to analyze a single image. Despite these troubles, it should be noted that CT-Fire still holds potential, but some issues would need to be resolved in order to make it useful for future work related to this project.

Figure 9: CT-Fire Problem
After some trial and error, the data measurement was finally done in OrientationJ, a plugin built for the National Institute of Health’s ImageJ program. Unlike the curvelet transforms used by CurveAlign and CTFire, OrientationJ is a tensor-based approach. It calculates the primary orientation and coherency of the fibers in the image, with the coherency being a measure of the overall alignment of the fibers. A coherency closer to its maximum value of one indicates that the fibers are close in orientation, while a coherency closer to zero indicates that the fibers are less ordered, being oriented in various directions.

The purpose of the program is to characterize the dominant fiber orientation and isotropy properties of a region of interest (ROI) in an image. In order to do this, we need to first define the weighted inner product in Equation 3.

\[
3) \langle f, g \rangle_w = \iint_{\mathbb{R}^2} w(x, y) f(x, y) g(x, y) dx dy
\]

where \( w(x, y) \geq 0 \) is a weighting function which specifies the area of interest\(^{27}\). The functions \( f \) and \( g \) are smaller regions of interest over which a two-dimensional Gaussian function is calculated. The Gaussian function is used for the purpose of reducing the effect of noise while calculating the coherency values. The area of interest was left as a general function in the theoretical treatment of OrientationJ to allow for the possibility of irregularly shaped regions of interest, but like Sage suggests, this project made use of square regions of interest (ROIs) due to the image size of 512 pixels by 512 pixels.\(^{27}\) Were the images to be analyzed with multiple ROIs, the square region of interest also allows for the use of a sliding window approach, which removes the user bias from the selection of regions to analyze as the entire image is subjected to
analysis. Clemons et al. used ellipse-shaped regions of interest selected by hand instead.\(^\text{14}\)

The norm of this inner-product, a function associating the vector space with the nonnegative real numbers, is \(\|f\|_w = \sqrt{\langle f, f \rangle_w}\). After this, consider the derivatives in the direction of unit vector \(u_\theta = (\cos \theta, \sin \theta)\), which is given by \(D_{u_\theta} f(x, y) = u_\theta^T \nabla f(x, y)\)

where \(\nabla f(f_x, f_y)\) is the gradient of the image. The next step is to finding the directions \(u\) along which the directional derivatives is maximized over the ROI. It is given by

\[
4) u_\theta = \arg \max_{\|u\| = 1} \| D_u f \|_w^2
\]

These directional derivatives are used to calculate the orientation angle. Using a standard inner-product manipulation, we find that

\[
5) \| D_u f \|_w^2 = \langle u^T \nabla f, \nabla f^T u \rangle_w = u^T J u
\]

In this case,

\[
6) J = \langle \nabla f, \nabla f^T \rangle_w = \begin{bmatrix}
(\langle f_x, f_x \rangle_w, \langle f_x, f_y \rangle_w, \langle f_x, f_y \rangle_w, \langle f_y, f_y \rangle_w)
\end{bmatrix}
\]

Is the structure tensor, which is a 2 x 2 symmetric positive-definite matrix. Solving for the maximum value of (5) can be done by setting the derivative of \(u^T J u + 1 - \frac{\lambda}{2} u^T J u\) with respect to \(u\) to zero, yielding the eigenvector equation: \(J u = \lambda u\).\(^\text{27}\)

This tensor given above allows for OrientationJ to calculate the eigenvalues for the dominant orientation and the minimized direction, both of which are needed to calculate the coherency for each pixel. The orientation was not used directly in this project as it depends on the angle from which the image was taken and the fact that this general direction of the collagen fibers is irrelevant. The direction is arbitrary, but whether or not
the fibers point in similar directions (e.g. coherency) to each other is not, and is consistent regardless of the sample orientation as the image is taken.

The eigenvalues are as follows:

\[ a) \quad \lambda_{max} = \max\|D_u f\|^2 \]
\[ b) \quad \lambda_{min} = \min\|D_u f\|^2 \]

They correspond to the dominant orientation and the minimized orthogonal direction respectively. Sage gives the formula used to calculate the coherency:

\[ 7) \quad C = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} = \sqrt{\left(\langle f_y, f_y\rangle_w - \langle f_x, f_x\rangle_w\right)^2 + 4\langle f_x, f_x\rangle_w \langle f_y, f_y\rangle_w} / \langle f_x, f_x\rangle_w + \langle f_y, f_y\rangle_w, \quad C \in [0 \ldots 1] \]

As mentioned earlier, C has a maximum value of 1, which occurs if the collagen fibers are all parallel. The minimum value of zero occurs when the collagen fibers are equally distributed in all directions.

Figure 10: OrientationJ Visual Analysis Mode
There are several functions built into the OrientationJ plugin which make use of this tensor calculation. The visual directional analysis provides a qualitative view of the fiber alignments, encoding the hue as the fiber orientation, the saturation as the coherency at a given point, and the brightness is determined by the source image, all shown in Figure 10. The color scale in Figure 10 demonstrates what colors correspond to what fiber angle relative the overall alignment of the fibers calculated by OrientationJ. The associated coherency map created (Figure 11) also serves to provide the user with insight into the computer’s calculations by allowing the user to see what the computer is considering fibers and what it is considering blank space. This coherency map is a visual representation of the coherencies calculated. For this project I took the mean coherency value for the image, which was based on these calculated coherencies.

The most valuable functions for this project were the OrientationJ Measure function and the OrientationJ Distribution function, which provide numerical outputs and graphs. Due to time constraints the Measure function was the one used for analysis. It
allowed the user to select their own regions of interest and gave a numerical output of the coherency values, giving the ability to run further statistical analyses upon the data, as done in this paper. The Distribution function provides the user with a histogram of collagen alignments, which can provide a wealth of information.

The Analytical Process

After completing the SHG imaging and storing the resulting .tif files, we used OrientationJ Measure in order to obtain the coherency values for each sample, treating the entire image as the region of interest (ROI) in order to avoid user bias in selecting regions to image. Attempts were made to use a sliding window approach, where the ROIs were laid out in a grid of sixteen 128 by 128 pixel squares covering the entire image.

Figure 12: Attempted Sliding Window Approach
The benefits of this approach lie in the smaller ROIs. The incoherent empty spaces could be discarded due to a lack of collagen making them irrelevant to the analysis, and the remaining ROIs provide a more detailed picture of the structure being imaged. This approach was unfortunately scrapped for the purposes of this paper due to time constraints, because OrientationJ Measure requires the user to input each ROI individually and performs its calculations for each ROI after each new input. The whole image approach is significantly faster, but is not without its own complications. OrientationJ treats empty space in an image (areas without collagen fibers) as incoherent. When images contain significant empty regions, the calculated coherency value is skewed closer to zero, indicating disordered “basket-weave” collagen. The problem would be less significant with the sliding window approach, as whole (smaller) ROIs could be discarded on the basis of containing too little collagen to be useful.

To address this problem, the images were analyzed both unaltered and by weighting them based on their collagen coverage as a percentage of the image area. This weighting was accomplished using the built-in thresholding function of ImageJ, and was generally left on the automatic setting in order to reduce user bias. In some especially noisy images, ImageJ was unable to exclude the noise, which would be manually removed by increasing the threshold, excluding pixels below a certain value. While this approach does introduce the potential for user bias, it also partially counters this in the data analysis as the user selections have a reduced weight in the calculation of weighted average coherencies.
The process of obtaining the coherency data used is outlined below:

1. We opened the .tif file of the desired image in ImageJ prior to opening the OrientationJ measure plugin.

2. While optional, we adjusted the ImageJ Brightness/Contrast setting under the “Image” tab on several images. This process does not impact the data collected from the image, but simply adjusts the display for the user. This was beneficial for the weighting process later.

3. At this point the OrientationJ Measure plugin was opened and the region of interest was selected. In this experiment the entire image was selected.

4. The coherency value calculated by OrientationJ was recorded at this point.

5. This step was done only once per data recording session. Under the “set measurements” menu, we needed to ensure that the “Area fraction” box was selected.

6. Using the ImageJ Analyze tab, the threshold image was measured (separately from OrientationJ). The “area fraction” was recorded, which was the percentage of pixels kept. Generally speaking, the thresholding was left on the automatic setting to reduce user bias, but in files with significant noise the threshold would be increased to reduce the amount of noise particles being considered. This

7. All data recording was done manually in Excel. After obtaining measurements, the threshold image was closed, and the process was repeated for each SHG image gathered.
This process produced the average coherency value across the entire image and the area coverage of the collagen within, displayed in the fashion shown in Figure 13.

![Coherency visualization](image)

**Figure 13: Contrast Enhanced Image in OrientationJ Measure**

OrientationJ version 2.0.5 does not have the ability to analyze multiple images at once built in, nor is there a means to analyze subsequent ROIs independently. This means that the process shown above would need to be done for every image, and every region of interest within an image after the first results in the measurement of every region of interest present, or a new macro making use of the software would need to be developed. This presented a problem for the user, as the analysis of several regions of interest within a single image took up a significant amount of extra time, making it unfeasible to continue the sliding window approach.
Results

The coherency values recorded for each image were recorded in Excel alongside their weight values. The average of these was taken both with and without weighting taken into account. We compared the mean values and the variance of the three tissue types established earlier (normal adjacent, fibrotic, and cancerous) in order to see if there were statistically significant differences between these groups, indicating that OrientationJ did indeed distinguish them by their average collagen alignments.

There was a statistically significant difference between the tissue types when weighted. This was determined by one-way ANOVA (F(30,2) = 3.3158, p = 0.03). The input values
for the F-function are the degrees of freedom within and between samples, respectively, and are calculated as follows: \((N - k) = 30\) and \((k - 1) = 2\). In both Figures 14 and 15, we ran a one-way ANOVA test in order to determine if there was a statistically significant difference in the coherency values across the different tissue types. The null hypothesis stated that there was no statistically significant difference between the tissue types regarding collagen coherency, with the p-value (also described as an alpha value) being set to the common value of 0.05. This value of 0.05 corresponds to a five percent chance that the null hypothesis is true. If the ANOVA test returns a p-value less than 0.05, the data is statistically significant as the probability that the results were up to chance is small. On the other hand, a p-value greater than 0.05 indicates that there is a large enough probability that the data is up to chance that the null hypothesis cannot be discounted.

Because this result was deemed statistically significant, the next step was to perform subsequent unequal variance t-tests comparing the tissue types two at a time. T-tests operate on a similar principal to ANOVA tests, so the p-value is again the relevant output, and the same rules regarding where the p-value is relative to 0.05 apply. Comparing the normal tissues to cancerous tissues, the p-value of 0.03 again confirms that the data is statistically significant. For the comparison between fibrotic and cancerous tissue, the p-value of 0.02 nets the same result. Important values factored into these ANOVA tests are listed in Table 2.
Figure 15: Unweighted Average Coherencies

Figure 15 yields a similar result, though the statistical significance is questionable. Applying the same ANOVA test used for the weighted coherencies, the p-value of 0.05 is on the cusp of being statistically insignificant. Being potentially significant, a T-test was performed comparing the fibrotic tissue to the cancerous tissue, finding a p-value of 0.01.

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Weighted Avg.</th>
<th>Average</th>
<th>Std. Deviation (W)</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancerous</td>
<td>0.353</td>
<td>0.335</td>
<td>0.101</td>
<td>0.078</td>
</tr>
<tr>
<td>Fibrotic</td>
<td>0.286</td>
<td>0.272</td>
<td>0.083</td>
<td>0.073</td>
</tr>
<tr>
<td>N. Adjacent</td>
<td>0.251</td>
<td>0.242</td>
<td>0.1</td>
<td>0.086</td>
</tr>
</tbody>
</table>

Table 2: Important Values From Coherency Analysis
In both the weighted and unweighted coherency data, the same trend is preserved. The mean coherency increases as the tissue progresses from normal to fibrotic to cancerous, but the process of weighting magnifies these differences.

**Why use ANOVA Tests and T-tests?**

For the analysis of the collagen coherency data, the one-way Analysis Of Variance (ANOVA) test was used in order to determine the statistical significance of said data. As touched upon in the prior section, the standard p-value of 0.05 was chosen because it refers to a 95% confidence in the data not being up to chance. In order to make use of this test, the following assumptions were made.

1. The dependent variable needed to be continuous. That was the case, as coherency is a continuous measurement ranging from 0 to 1.

2. The independent variable needed to be categorical data. Given the independent variables were tissue types, this assumption was upheld.

3. One-way ANOVA tests assume independence of observations, meaning that no observation is impacting another. While each image was separate from the others, to say that the tissues had no impact on the other types is technically false. The normal adjacent tissue within a cancer patient can be impacted by the cancerous tissue even centimeters away, though it is still distinct from fibrotic or cancerous tissue. The assumption made in this case was while the tissues coming from the same person (e.g. cancerous tissue and fibrotic tissue from the same patient) may have been capable of interacting, the contribution to the data from these
interactions would be small, and therefore the ANOVA test would not be skewed significantly by such contributions.\textsuperscript{28}

4. The samples are assumed to be normally distributed. The ANOVA test is considered fairly robust in this regard, being able to tolerate moderate deviations from this rule, and the results of a prior Shapiro-Wilks test indicates some semblance to a normal distribution (p > 0.05).

5. Finally, the one-way ANOVA test assumes homogeneity of variances.\textsuperscript{28} While the variances for the samples were not all equal, they were all within an order of magnitude of each other. Like the normal distribution assumption, this assumption is considered robust (but not as much as the normality assumption) and can accommodate some deviations.\textsuperscript{28}

The ANOVA test results take the form of an F-distribution.\textsuperscript{28} This distribution is a single-tailed distribution, where the p-value mentioned above associated with it refers to the percent of the area past a certain F-value. The F-value is dependent on two values, the MST which refers to the Mean Square for Treatments and MSE which refers to the Mean Square for Errors, defined in Equations 8 and 9.

\begin{align*}
8) \text{MST} &= (k - 1)^{-1} \sum_{i=1}^{k} n_i (X_i - X)^2 \\
9) \text{MSE} &= (N - k)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_j} (X_{ij} - X_i)^2
\end{align*}

The coefficients at the front of these sums refer to the degrees of freedom. N is the number of data points in all categories and k is the number of categories. As mentioned
earlier, \((N - k) = 30\) and \((k - 1) = 2\). \(X\) is the mean of the means of each sample, \(X_i\) is the mean of a sample \(i\), and \(X_{ij}\) is a particular value within one of the samples \(i\). The F-value is the test statistic for the ANOVA test and is defined by Equation 10.

\[
F = \frac{MST}{MSE}
\]

What this ANOVA test does is compare the values of the samples with their variances. If the variance is large relative to the values of the samples as evidenced by a small F-value, the data is considered statistically insignificant as random (or unaccounted for) variances are large enough to cast doubt on the validity of the measurements. On the other hand, samples with values significantly larger than their variances will demonstrate a higher F-value, which corresponds to a smaller probability of the data being up to chance. T-tests perform a similar function, but do so between two samples. These were done to account for a weakness of the ANOVA test, which is that it does not indicate which samples are contributing to differences between them. In the unweighted case, this is demonstrated by the ANOVA p-value indicating a lack of statistical significance across the three samples, while a subsequent T-test between the fibrotic tissue and the cancerous tissue demonstrated a p-value of 0.01. This indicated that while the fibrotic tissue and cancerous tissue were different in a statistically significant manner, the healthy and fibrotic tissue were too close to be considered different, which obscured the nuance of the situation when analyzed by the ANOVA test alone.
Due to the statistically insignificant result of the unweighted sample, the conclusions which can be drawn from this research are limited. The trend between samples of collagen coherency increasing with cancer progression, which has been demonstrated in other research,\(^3\) was upheld in this project. Below is a possible option to improve the process for future work. One way would be to automate and implement the sliding window approach based on OrientationJ or the theory it is built on. This would ideally remove the need to weight ROIs based on collagen content and allow for empty regions to be discarded while keeping the collagen within an image, as opposed to deciding whether an image is suitable for analysis as a whole.

This project was a piece of a larger project in collaboration with the Maine Medical Center Research Institute (MMCRI). Using fiber morphology analysis as a base we will collaborate with Dr. Peter Brooks (MMCRI) who has developed an antibody that specifically targets cryptic collagen epitopes only exposed during collagen remodeling. This antibody (HU177) will be co-stained in future imaging studies and will be correlated with specific morphological collagen alterations in the future. Beyond collagen remodeling, there is evidence that this antibody decreases the motility of cancer cells while increasing the motility of T-cells; offering a double-edged sword to our understanding and treatment opportunities of pancreatic cancer.
REFERENCES


24. Candes, E., Demanet, L., Donoho, D. & Ying, L. Fast Discrete Curvelet


AUTHOR’S BIOGRAPHY

Gerren Welch was born in Brunswick, Maine on October 24th, 1997. Other than a brief stint in Louisiana, Gerren spent his childhood in Maine, graduating from Messalonskee High School in 2016. Gerren graduated with a physics major and minors in mathematics and biology. An avid musician, Gerren spent his college career playing and singing in a variety of roles, playing euphonium in the Symphonic Band, trombone in the Jazz ensemble, and singing in the Black Bear Men’s Choir. He enjoys science, but has found over the course of his college career that he has no idea what he wants to do with that information. His future plans include expanding his horizons by living somewhere new and discovering what work suits him.