Article

Evaluation of Collagen Alterations in Early Precursor Lesions of High Grade Serous Ovarian Cancer by Second Harmonic Generation Microscopy and Mass Spectrometry

Kristal L. Gant [1,2], Alexander N. Jambor [2], Zihui Li [3], Eric C. Rentchler [2], Paul Weisman [4,5], Lingjun Li [3,5,6], Manish S. Patankar [1,5,*] and Paul J. Campagnola [2,5,*]

[1] Department of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI 53706, USA; kgant@wisc.edu
[2] Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA; ajambor@wisc.edu (A.N.J.); rentchler@wisc.edu (E.C.R.)
[3] Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA; zli597@wisc.edu (Z.L.); lingjun.li@wisc.edu (L.L.)
[4] Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA; pweisman@wisc.edu
[5] Carbone Cancer Center, University of Wisconsin-Madison, Madison, WI 53706, USA
[6] School of Pharmacy, University of Wisconsin-Madison, Madison, WI 53705, USA
* Correspondence: patankar@wisc.edu (M.S.P.); pcampagnola@wisc.edu (P.J.C.)

Simple Summary: The collagen architecture in the extracellular matrix (ECM) is highly remodeled in high grade serous ovarian cancer (HGSOC). Many of these tumors begin in the fallopian tubes (FT) before metastasizing to the ovaries and it is important to study ECM alterations in carcinogenesis. Here, we used Second Harmonic Generation (SHG) microscopy to classify changes in the collagen fiber morphology in normal FT, and precursor pure p53 signatures and serous tubal intraepithelial carcinoma (STICs) in tissues with no HGSOC. Using a machine learning approach based on image features, we were able to discriminate the tissue groups with good classification accuracy. We additionally performed mass spectrometry analysis of normal and HGSOC tissues to associate the differential expression of collagen isoforms with fiber morphology alterations. This work provides new insights into ECM remodeling in early stage HGSOC and suggests the combined use of SHG microscopy and mass spectrometry as a new diagnostic/prognostic approach.

Abstract: Background: The collagen architecture in high grade serous ovarian cancer (HGSOC) is highly remodeled compared to the normal ovary and the fallopian tubes (FT). We previously used Second Harmonic Generation (SHG) microscopy and machine learning to classify the changes in collagen fiber morphology occurring in serous tubal intraepithelial carcinoma (STIC) lesions that are concurrent with HGSOC. We now extend these studies to examine collagen remodeling in pure p53 signatures, STICs and normal regions in tissues that have no concurrent HGSOC. This is an important distinction as high-grade disease can result in distant collagen changes through a field effect mechanism. **Methods:** We trained a linear discriminant model based on SHG texture and image features as a classifier to discriminate the tissue groups. We additionally performed mass spectrometry analysis of normal and HGSOC tissues to associate the differential expression of collagen isoforms with collagen fiber morphology alterations. **Results:** We quantified the differences in the collagen architecture between normal tissue and the precursors with good classification accuracy. Through proteomic analysis, we identified the downregulation of single α-chains including those for Col I and III, where these results are consistent with our previous SHG-based supramolecular analyses. **Conclusion:** This work provides new insights into ECM remodeling in early ovarian cancer and suggests the combined use of SHG microscopy and mass spectrometry as a new diagnostic/prognostic approach.

Keywords: HGSOC; STICs; precursor lesions; Second Harmonic Generation; collagen remodeling; mass spectrometry

1. Introduction

High Grade Serous Ovarian Cancer (HGSOC) is a highly metastatic disease, defined genetically by mutations in the tumor suppressor genes Tp53, BRCA I, and BRCA II, and DNA copy number alterations [1]. While these mutations are well-documented, the associated effects in the tumor microenvironment (TME), especially in terms of remodeling of the extracellular matrix (ECM), have not been well-studied. Such modifications occur in essentially all epithelial cancers, and are important in HGSOC because this disease can metastasize while the lesions are smaller than the resolution of clinical imaging modalities (e.g., ultrasound, CT, MRI, and PET) [2–4].

Serum tests (CA125 and HE4) do not have sufficient specificity and sensitivity for early reliable diagnosis [5–7]. As a result of these factors, in more than 70% of patients, HGSOC is detected at an advanced stage when the treatment options are limited. We postulate that the development of efficacious imaging/screening modalities requires a more thorough understanding of the HGSOC microenvironment.

For this purpose, we utilized the high-resolution collagen specific, optical modality of Second Harmonic Generation (SHG) microscopy to probe all levels of collagen structure (molecular through fiber). Importantly, we previously developed machine learning algorithms to differentiate between normal and high-risk ovarian stromal tissues as well as cancer sub-types based on the 3D collagen fiber morphology patterns [8]. We were able to discriminate HGSOC and normal tissues with excellent classification accuracy (~95%) [9] and other sub-types with good accuracy (~85%).

We have also documented sub-resolution macro/supramolecular changes (protein helix attributes) and fibril organization (size and packing) in the aberrant tissue classes [10,11]. Collectively, these studies showed that the collagen fibers are more aligned in HGSOC than in the corresponding normal tissues or other ovarian cancer sub-types and that the underlying supramolecular and fibril structures are more disordered. These results are consistent with improperly synthesized new collagen and/or faster turnover in normal Collagen I (Col I).

As the majority of HGSOC cases originate through precursors in the fallopian tube (FT) secretory epithelium [12–15], it is also important to investigate the corresponding collagen alterations as these can be biomarkers for early diagnosis of the disease. Both p53 signatures and serous tubal intraepithelial carcinomas (STICs) have been identified as two early precursors of HGSOC in the FT [16]. Aside from a loss of cilia and outgrowth of secretory cells, p53 signatures are defined by their aberrant and intense p53 staining [15].

In contrast, STICs are associated with high p53 intensity and the acquisition of cellular changes, where the morphology becomes more disorganized. Malignant disease in the FT (and primary ovary) that develops from the STICs has additional morphological alterations along with a high proliferative index [1]. Similar to the ovarian TME itself, the corresponding changes in the FT microenvironment, especially in terms of the collagen architecture, are not well-known beyond standard hematoxylin and eosin (H&E) histology, which is not sensitive to detailed fibrillar features.

We recently used SHG microscopy to investigate the collagen fiber structure in concurrent STIC and HGSOC fallopian tube tissues [17]. Interestingly, the collagen morphology in HGSOC resembled that occurring in the ovary itself, and using a multivariate analysis, excellent classification accuracy (~95%) was obtained relative to STICs and normal regions in the same tissue. However, a more modest accuracy (~75%) was obtained between normal and STIC regions, suggesting that more detailed analyses are required to define these collagen structural changes.

Here, we extended these studies to pure p53 and STIC precursors to determine if early changes in the collagen fiber morphology are detectable along with the p53 molecular changes. It is also important to complete these studies in the absence of concurrent HGSOC as these lesions can result in the transformation of distant collagen through a field effect mechanism [18–21] and obscure the morphologic changes associated only with the precursor states. We also used mass spectrometry to examine the molecular changes that

underlie remodeling, e.g., the up- or down-regulation of collagen isoforms. The correlation of differential isoform expression to changes in the collagen fiber morphology has not been previously investigated, and we suggest that this analysis can provide useful insights into remodeling in early disease progression.

2. Methods

2.1. Archived Human Tissues

In this retrospective study, archived fallopian tubes and ovarian tissues from the University of Wisconsin Carbone Cancer Center Tissue Bank and the University of Wisconsin Department of Surgical Pathology were analyzed under an IRB-approved protocol (protocol #2019-0211). Flash frozen normal (*N* = 3) and tumor fallopian tube tissues (*N* = 3) were analyzed via mass spectrometry, while archived fallopian tube tissues (*N* = 12) were analyzed via SHG microscopy. Table S1 provides additional information on the tissues evaluated.

2.2. Sample Processing, Histology, and Mapping of Precursor Lesions

The SEE-FIM protocol [12] was executed to identify fallopian tube samples with HGSOC and HGSOC precursors. Paraffin blocks of the cases with confirmed normal, p53 signatures, STICs, and HGSOC were serially sectioned to obtain 5–10 µm thick sections. The sections were stained with H&E to monitor the morphology and to confirm the presence of HGSOC and its precursor lesions. The slides were also immunostained for p53 using the DO-1 hybridoma (SantaCruz Biotechnology, SantaCruz, CA, USA). Adjacent sections were retained as unstained slides for SHG imaging. The stained slides were examined by a trained pathologist (P.W.) to confirm the diagnosis. This pathological review was used to map the normal tissues, precursor lesions, and HGSOC in the unstained slides.

2.3. Sample Processing for Mass Spectrometry-Based Proteomic Analysis

Protein extraction and digestion. Each sample was dissolved in 1 mL of extraction buffer (4% sodium dodecyl sulfate, 50 mM Tris-HCl, pH 8) and sonicated using a probe sonicator (Thermo Fisher Scientific, San Jose, CA, USA). Protein extracts were reduced with 10 mM dithiothreitol (DTT) for 30 min at room temperature and alkylated with 50 mM iodoacetamide for another 30 min in the Dark before quenching with DTT. The proteins were then precipitated with 80% (*v/v*) cold acetone (−20 °C) overnight. The samples were centrifuged at 14,000× *g* for 15 min after which supernatant was discarded.

The pellets were rinsed with cold acetone and air-dried at room temperature. Eight molar urea was added to dissolve the pellets, and 50 mM Tris buffer was used to dilute the samples to a urea concentration <1 M. On-pellet digestion was performed with LysC/trypsin (Promega, Madison, WI, USA) at a 50:1 ratio (protein: enzyme, *w/w*) at 37 °C overnight. The digestion was quenched with 1% trifluoracetic acid, and the samples were desalted with Sep-Pak C18 cartridges (Waters, Milford, MA, USA). The concentrations of peptide mixture were measured by peptide assay (Thermo Fisher Scientific, San Jose, CA, USA). Ten micrograms of peptide were aliquoted for each sample and dried in vacuo.

Liquid chromatography (LC)-tandem mass spectrometry analysis. The samples were analyzed on a Q-Exactive quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to a Waters nanoAcquity Ultra Performance LC. Each sample was dissolved in 15 µL 4% acetonitrile (ACN) and 0.1% formic acid (FA) in water before loading onto a 75 µm inner diameter homemade microcapillary column, which was packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 µm, 130 Å, Waters, Milford, MA, USA) and fabricated with an integrated emitter tip. Mobile phase A was composed of water and 0.1% FA, while mobile phase B was composed of ACN and 0.1% FA.

LC separation was achieved across a 120-min gradient elution of 4% to 30% mobile phase B at a flow rate of 300 nL/min. Survey scans of peptide precursors from 300 to 1500 *m/z* were performed at a resolving power of 70,000 with an automatic gain control (AGC) target of 1×10^6 and maximum injection time of 250 ms. The top 15 precursors

were then selected for higher-energy collisional dissociation (HCD) fragmentation with a normalized collision energy of 30, an isolation width of 2.0 Da, a resolving power of 17,500, an AGC target of 1×10^5, a maximum injection time of 150 ms, and a lower mass limit of 120 m/z. The precursors were subject to dynamic exclusion for 45 s with a 10 ppm tolerance. Each sample was acquired in technical triplicates.

Data analysis. The raw files were searched against the UniProt *Homo Sapiens* reviewed Database (February 2020) using MaxQuant (version 1.5.2.8) [22] with trypsin/P selected as the enzyme and two missed cleavages allowed. Carbamidomethylation of cysteine residues (+57.02146 Da) was chosen as fixed modification and variable modifications included oxidation of methionine residues (+15.99492 Da), acetylation at protein N-terminus (+42.01056 Da), and hydroxylation on proline residues (+15.99492 Da). The "LFQ quantification" and "Match between runs" features were enabled in MaxQuant. Search results were filtered to a 1% false discovery rate (FDR) at both the peptide and protein levels.

Peptides that were found as reverse or potential contaminant hits were filtered out, and all other parameters were set as the default. ECM proteins were identified and classified by matching the results to Human Matrisome Dataset [23]. Proteins were considered as identifiable when detected in at least one sample and quantifiable when detected in at least two samples in each group. Missing intensities were replaced using the "replace missing values from normal distribution" feature in Perseus [24] (version 1.6.0.7) prior to further processing.

Two-sample Student's *t* tests with a two-tailed distribution for binary comparison and hierarchical clustering analysis were conducted using Perseus. The volcano plot was generated using R packages. The mass spectrometry proteomics Data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [25] with the Dataset identifier PXD025864.

2.4. Second Harmonic Generation (SHG) Microscopy

The SHG laser scanning microscope used here has been described in detail previously and only the salient features are given here [26]. The excitation source was a mode-locked Titanium Sapphire laser (Mira; Coherent, Santa Clara, CA, USA), providing 890 nm excitation and coupled to an upright microscope (BX61; Olympus, Tokyo, Japan). Laser scanning and Data acquisition were achieved through home written LabVIEW code and National Instruments FPGA (National Instruments, Austin, TX, USA). The SHG Excitation used a 40×0.8 NA water immersion lens (LUMPlanFL/IF; Olympus, Tokyo, Japan) and a 40×0.9 NA condenser for collection of the forward propagating signal. The lateral and axial resolutions of the system were approximately 0.7 and 2.5 μm, respectively, as this is sufficient for resolving collagen fibers.

The SHG emission has an associated directionality resulting from the sub-resolution fibril structure [27], and we acquire the forward and backward propagating signals [28]. These respective components were collected using identical photon-counting detectors (7421 GaAsP; Hamamatsu, Hamamatsu City, Japan) with the backward detector in a non-descanned geometry. For each channel, the SHG wavelength (445 nm) was isolated with a dichroic mirror and 10 nm FWHM bandpass filter (Semrock, Rochester, NY, USA). Circular polarization was used for imaging as this state excites all fiber orientations equally. This polarization of the excitation laser was determined at the focus by imaging dye-labelled vesicles [26]. The collected images were 512×512 pixels with a field of view of 180×180 μm. The image acquisition time was 3 s per frame with three-frame Kalman averaging.

A total of 81 image stacks were analyzed and utilized for classification across three tissue groups: distal normal ($N = 37$), p53 only signature ($N = 7$), and STIC lesion ($N = 37$). We also included a comparison with HGSOC tissues ($N = 33$) from confirmed cancer patients in some of the analyses.

2.5. Image Analysis

An initial set of features to be used for linear discriminant analysis (LDA) was generated from the outputs of Gray Level Co-Occurrence Matrix (GLCM) analysis, two-dimensional fast Fourier transform (2D-FFT) methods, and the curvelet transform combined with the FIRE extraction algorithm (CT-FIRE) [29]. FIJI with the Texture Analyzer plugin was used to calculate five GLCM texture features associated with the similarities and differences between adjacent pixels: Angular Second Moment (ASM), Entropy, Inverse Difference Moment (IDM), Contrast, and Correlation.

The 2D-FFT analysis was performed in FIJI using the Radial Profile Extended and the Oval Profile Plot plugins to characterize the alignment and radial exponential decay of the image power spectrum. We previously described the radial- and azimuthal-averaging procedures used in this approach [17]. All curve fitting of this Data was done in Origin 2018 (OriginLab, Northampton, MA, USA).

CT-FIRE was utilized to perform curvelet transform and fiber extraction to characterize individual fiber morphology features (fiber length, width, and straightness). As there was insufficient signal from the weaker backward channel for the CT-FIRE analysis, only the Data readouts from the forward channel were analyzed. Lastly, the image coverage (packing coefficient) was quantified by creating a binary mask over a dynamic lower threshold and calculating the fraction of the resulting non-vanishing pixels. The methods and features used are summarized in Table 1.

Table 1. Summary of different metrics used for Linear Discriminant Analysis (LDA).

Method	Feature	Measurement Output
GLCM—encodes the intensity differences of adjacent pixel pairs into a matrix	ASM	Homogeneity of image texture
	Entropy	Randomness of image intensity distribution; inversely correlated to ASM
	IDM	Homogeneity of image intensity
	Contrast	Local intensity variations; inversely correlated to IDM
	Correlation	Correlation of intensity between adjacent pixels
2D-FFT—converts an image into its power spectrum prior to analysis	Alignment	Overall alignment of image by examining signal at different spatial frequencies
	Amplitude	Signal strength at lowest spatial frequencies
	Time Constant (TC)	Time constant of radial power spectrum, inversely proportional to exponential decay rate (k) and related to the incidence of high frequency features
CT-FIRE—identifies discrete fibers in a 2D image	Fiber length	Average length of fibers in image
	Fiber width	Average width of fibers in image
	Fiber straightness	Average straightness of fibers in image
Packing coefficient		Image coverage; related to the density of the collagen network in the image

The SAS software (SAS Institute Inc., Cary, NC, USA) was used to reduce the full feature set to the most significant metrics via forward selection at a significance level of $\alpha = 0.35$ (STEPDISC procedure). After the features were selected, the truncated Dataset was inputted to a linear discriminant analysis (LDA) performed with singular value decomposition and N-weighted priors. For this portion, 37 crops as normal, 6 as p53 signatures, and 37 as STIC were analyzed.

The accuracies and F_1 scores of the trained model were calculated for each class, where the latter quantity is the harmonic mean of precision (TP/(TP + FP)) and recall (TP/(TP + FN)), where TP, FP, and FN are true positives, false positives, and false negatives, respectively. An F_1 score is an additional metric of classifier accuracy which accounts for class imbalance. Receiver Operating Characteristic (ROC) curves were generated under five-fold cross-validation to assess the binary classifier performance and to quantify the trade-off between the true positive rate (TPR) and false positive rate (FPR).

3. Results

3.1. Mass Spectrometry Analysis

We explored the changes in ECM and ECM-related proteins, with a focus on collagen isoform expression in human normal ovarian and tumor tissues via mass spectrometry-based proteomics approaches. In total, we identified 233 ECM proteins, where 25 were

single α-chains comprising several collagen isoforms. We observed numerous proteins that were only found in either tumor or normal samples (Figure S1a,b). For instance, COL8A1, COL8A2, and COL21A1 were only detected in normal tissues whereas COL7A1 and MUC1 were exclusively identified in tumor samples (Table S2). Importantly, even with a low number of replicates, we were able to identify and quantify many differentially expressed (especially down-regulated) ECM proteins (Figure S1) in the tumor group.

As shown in the cluster map in Figure 1, 15 single α-chains from different collagen isoforms were present in all samples, and statistical proteomic landscape differences were found between the two groups. Interestingly, we observed decreased expression levels of many of these chains in the tumor samples (e.g., those for COL1, COL3, COL5, COL6, COL12, and COL14). Moreover, differences in multiple single chains of the same collagen isoform were detected (e.g., COL1A1 and COL1A2), improving the confidence of our observations. These findings support the existence of unique matrisome features in each group, where there were larger intergroup differences than intragroup variations. This was also borne out by analysis of the principal components (not shown). A full list of identified and quantified proteins can be found in the supplemental spreadsheet.
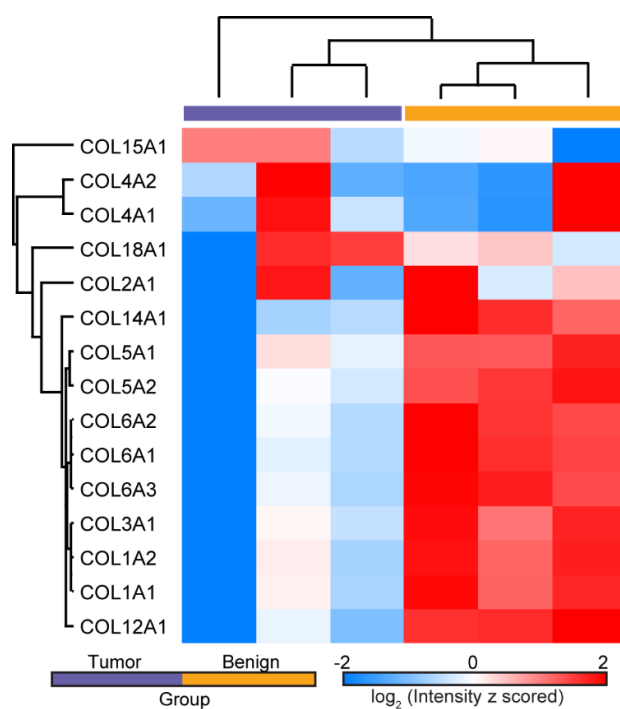
Figure 1. Hierarchical clustering of LFQ ion intensities showing the differential expression of single α-chains from several collagen isoforms in normal ovarian and HGSOC tissues. Several chains were down-regulated in HGSOC.

3.2. SHG Imaging and Analysis

Locating and mapping pure p53 signatures and STIC lesions. To obtain tissues with p53 signatures and STIC lesions with no concurrent tumors, we focused on FT tissues obtained from gynecologic surgeries not related to HGSOC. An archival text-based survey of patients with STIC over a 5-year period (2013–2018) revealed 12 patient cases that met our criteria. Only 2 of the 12 had concurrent p53 signature lesions along with the STIC. None of these cases had any pathological characteristics for HGSOC. The low number of cases with pure p53 signature and STIC lesions is consistent with their reported low incidence in these cohorts.

To confirm that these were pure precursors without the presence of cancer, routine histological stains (H&E and p53) were completed. Figure 2 provides an example of

one such precursor along with an example of HGSOC. In addition to the normal H&E distribution, the weak p53 reactivity is consistent with the absence of cancer.
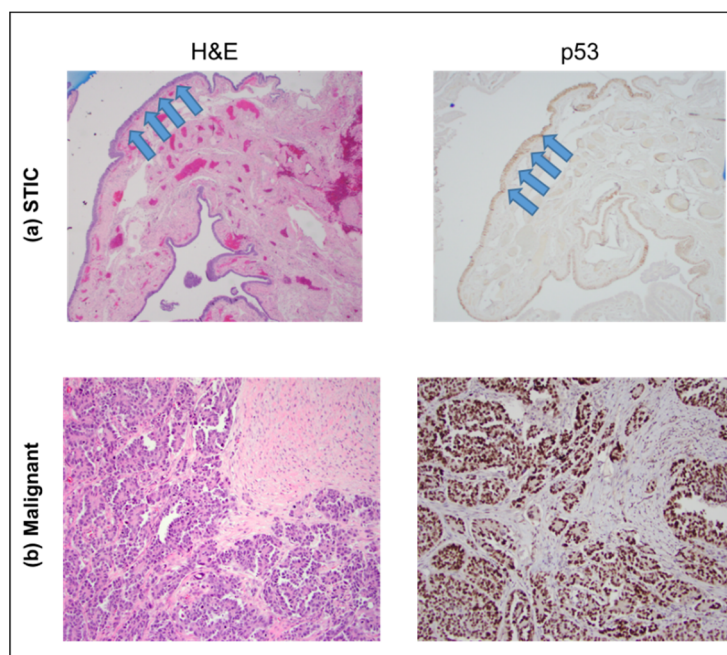

Figure 2. H&E (**left**) and p53 (**right**) staining of sections showing a STIC lesion (top, (**a**) images) and HGSOC (bottom, (**b**) images), respectively. The STIC lesion is confined to the fallopian tube epithelial surface (arrows), whereas the HGSOC forms a destructive, expansile lesion within the stroma of the involved tissue. The images were acquired at $40\times$.

The fallopian tube tissue slices for SHG imaging were unstained and, therefore, posed a challenge to accurately identify the normal regions, p53 signatures, and STICs. This issue was addressed by the H&E and p53 staining of an adjacent section of the same tissue, where these slides were used as a template to manually map and score the normal areas and precursors on the slides used for SHG imaging. This workflow was completed for all FT samples and is outlined in Figure 3.

A trained gynecologic pathologist (P.W.) first identified the specific areas of normal tissue (green rectangle), p53 signature (red rectangle), and STIC (blue rectangle) (Figure 3a). For orientation purposes, bright field images at $4\times$, $20\times$, and $40\times$ at each spot were also taken (not shown). The correlated unstained tissue (Figure 3b) was mapped according to the designated areas and imaged. The SHG images of collagen in areas corresponding to normal tissue, the p53 signature, and STIC are shown in Figure 3c, followed by the corresponding CT-FIRE and 2D-FFT analysis, which is quantified in Figure 4.

Our previous study indicated that a collagen coverage of less than 70% significantly altered the accuracy of the image analysis techniques, and many of these regions had sparser coverage. As a solution, representative image stacks for each channel were duplicated and cropped to 45×45 microns field of view (FOV), and we selected regions of interest (ROIs) with sufficient coverage. The p53 signatures and STIC lesions were small in spatial extent and localized in their respective tissue and yielded z-stacks comprised of 10 or fewer optical sections. Cropped images of each group (normal, $N = 37$; p53 signature, $N = 6$; STIC, $N = 37$) were analyzed using the image analysis protocols from our previous study [17].
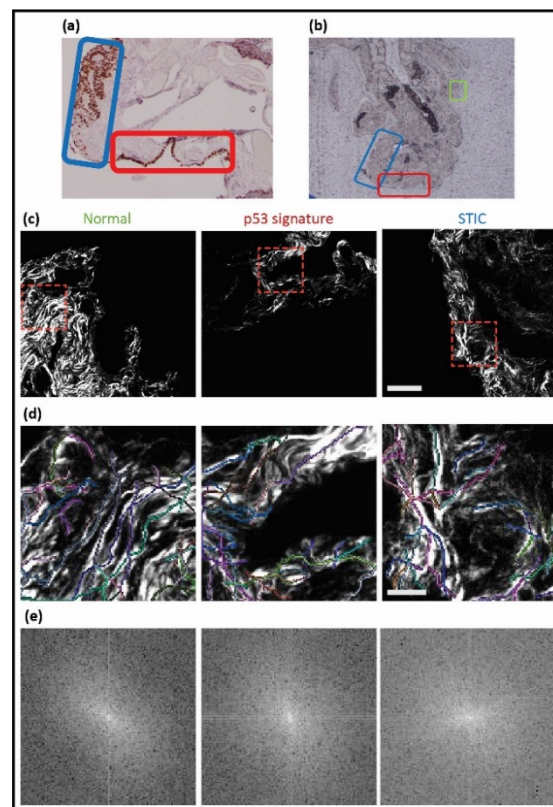
Figure 3. Workflow for identifying p53 and STIC precursors. A pathologist defined these tissues in mutant p53 stained images ((**a**), $4\times$), where green, red, and blue correspond to normal regions, p53, and STIC, respectively. The correlated unstained regions ((**b**), $4\times$) were identified and then used for SHG imaging, where the corresponding images are shown in (**c**) (field size = 180×180 microns). (**d**,**e**) CT-FIRE overlays and 2D-FFT power spectra of the cropped SHG images (red-dashed boxes) are shown.

Texture and other image features. For our analysis, we included features from GLCM, 2D-FFT methods, and CT-FIRE, where these techniques were applied to both the forward and backward channels. Unlike fluorescence where the emission is isotropic, SHG has an emission directionality that is related to the underlying structure [27]. Specifically, smaller and more disorganized features can appear in the backward channel, where these are often obscured in the forward collected signal. We have shown that these images are sufficiently different through the structural similarity index [17] to justify the inclusion of both signal pathways as independent features. However, the backward signal is intrinsically weaker, and insufficient signal was present for use in the CT-FIRE analysis. The Data from all these analyses are summarized in Figure 4.

Although we are specifically focused on discrimination between the distal normal and precursors (p53 and STIC), we included HGSOC as a point of comparison. In good agreement with a previous study [17], we found that images from HGSOC were associated with higher entropy and correlation, as well as lower contrast with respect to other groups (Figure 4a). In this context, lower contrast refers to the similarity of pairwise pixels rather than low signal and is also consistent with high correlation. However, we did not find any significant differences in the GLCM metrics between either of the precursors and distal normal regions.
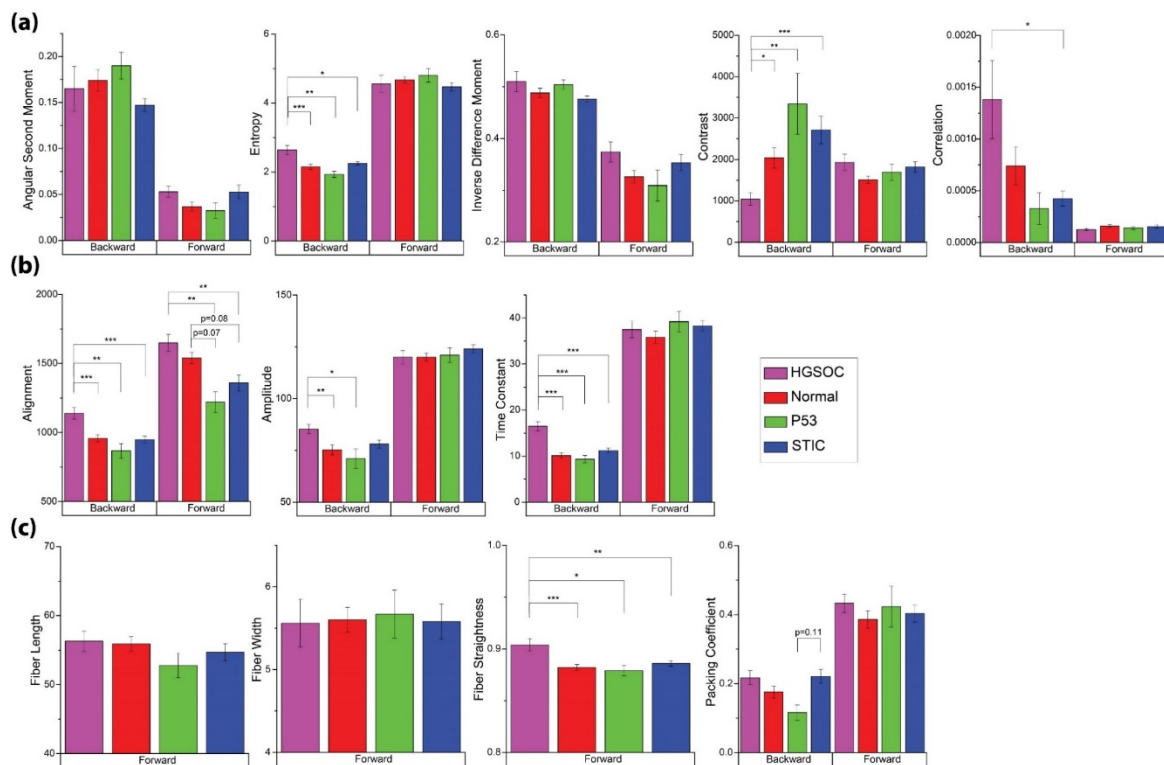
Figure 4. Image and texture features from the analyses given in Table 1 for the four tissue classes of normal, HGSOC, p53, and STIC. (**a**) GLCM features. (**b**) 2D-FFT parameters. (**c**) CT-FIRE outputs and Packing Coefficient. Significant *p*-values are denoted as follows: * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$.

2D-FFT methods were able to distinguish HGSOC as having higher alignment in the forward and backward channels, where both readouts of this metric should trend in the same direction. There was also a greater relative occurrence of high frequency (smaller) features in HGSOC, which is indicated by the higher time constant in its radial power spectrum (Figure 4b). Conversely, the STIC and p53 signature groups were characterized by lower alignment in the forward channel, although this was not significantly different. Lastly, while HGSOC was associated with straighter fibers, no differences were found between the two precursors and normal (Figure 4c).

Linear Discriminant Analysis (LDA). While none of the individual metrics from Figure 4 showed differences between the two precursors, we can attempt to obtain discrimination through the development of a linear discriminant (LD) model. This process can provide improved classification even if the individual components are not themselves statistically different, and we have used this process previously [17,30]. Since the collagen morphology of HGSOC in the FT is markedly distinct and already characterized [8,17], we limited our discrimination analysis to the two precursors and distal normal regions.

In order to better differentiate between the precursor groups (and to prevent overfitting in the trained LD model), the feature space generated by the GLCM, 2D-FFT methods, CT-FIRE, and the packing coefficient was limited via forward selection up to a significance level of $\alpha = 0.35$ (SAS/STEPDISC procedure). The most significant variables for discriminating between these groups (Packing Coefficient [B], ASM [B], IDM [F], Entropy [F], Entropy [B], Alignment [F], Alignment [B], Time Constant [B], and Fiber Straightness) were then used to construct canonical variables and visualize the classes via a scatter plot (Figure 5).
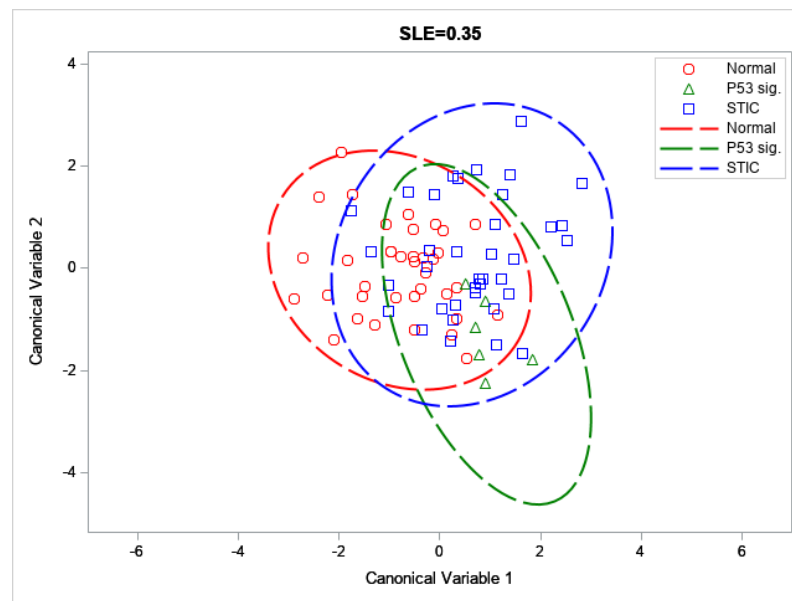
Figure 5. Scatter plot of the distal normal, p53 signature, and STIC across the first two canonical variables; with 95% confidence ellipses depicted by Dashed lines.

These forward selected metrics were then used to train an LD model capable of distinguishing between the three tissue groups by a set of binary classifiers (One-vs-Rest or OvR). Through this analysis, we were able to achieve accuracies and F_1 scores between ~65 to 91% and 9.1–66.0, respectively (Table 2). In particular, we achieved good discrimination for the distal normal and STIC lesion groups, despite low sample sizes (N = 37 each). The corresponding AUROCs for these classifiers were somewhat low (0.71 and 0.62 for the distal normal and STIC lesion, respectively; see Figure 6) but this may be improved upon by increasing the size of the training set.
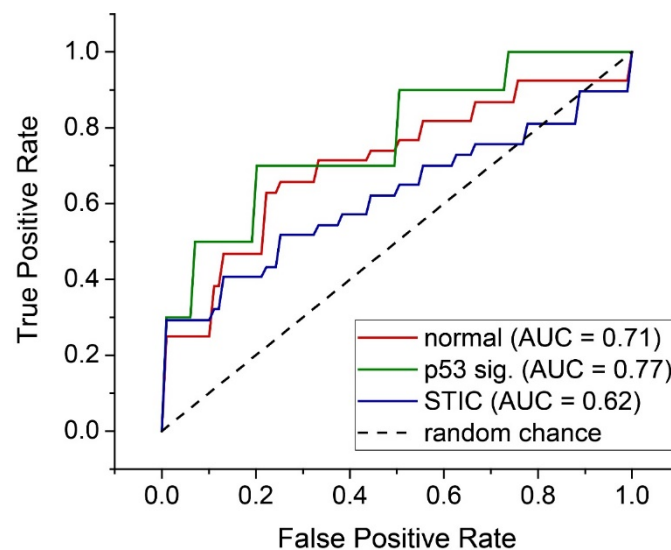
Figure 6. Individual ROC curves for each OvR classifier for the three tissue classes.

Table 2. The accuracies and F_1 scores for the OvR classifier.

Group	Accuracy %	F_1 Score
Distal normal	69.6	66.0
p53 signature	90.8	9.1
STIC lesion	65.4	61.0

Despite the high overall accuracy (~91%) for p53 signature classification, the corresponding F_1 score of 9.1 indicates a low number of true positives. To overcome the limitation of low *N* for the p53 signature group, we trained a classifier to distinguish between the distal normal and precursor regions, where the p53 signature and STIC were aggregated into a more general precursor group. Through a similar model using slightly different metrics, we achieved a high accuracy and F_1 score (74.7 and 77.8, respectively), as well as an AUROC of 0.68. The scatter plot and ROC curve for this model are included in the Supporting Information (see Figures S2 and S3).

4. Discussion

Using SHG microscopy, we have previously shown that there are significant changes in the collagen fibrillar morphology in HGSOC in the ovary itself, as well as in the fallopian tubes [8,17]. More subtle differences were observed in STIC regions that were co-existent in tissues with HGSOC. It is important to examine the collagen morphology in pure precursor tissues (p53 and STIC) to determine when characteristic collagen fiber alterations can be detected by SHG, as these could be used as unique diagnostic biomarkers of early-stage disease.

It is further important to perform these investigations in tissues without HGSOC, as these lesions can induce collagen remodeling in distant regions through a field effect mechanism [19]. However, the acquisition of these pure precursors is clinically rare and it took extensive time and effort to identify even the relatively low number of suitable banked tissues used here. Specifically, an experienced gyn/onc pathologist (P.W.) scanned patient cases from over a 5-year period for this study.

We also sought to determine if there were differences in the collagen expression patterns between normal ovarian tissues and HGSOC, and further, if these were related to the collagen morphology changes visualized by SHG microscopy. This is important as, while previous studies have suggested the up-regulation of several collagen isoforms in HGSOC (e.g., Col III and VI) [31,32], these studies used immunostaining and the results were not verified by quantitative molecular techniques.

This is potentially problematic as most available antibodies lack a high level of specificity for different isoforms (specifically Col I vs. Col III) [33] as the same epitope at the end of the helix is often tagged. As a consequence, the in vivo isoform composition in HGSOC is not yet definitively known and this represents a large gap in knowledge. We note that we have used self-assembled in vitro models of known composition [33,34] to show how the incorporation of Col III and Col V affects the fibrillar structure, but it has not yet been possible to create a direct link between collagen proteomic changes and fiber alterations in tumors. This is because SHG is not sensitive to non-collagen components, and more generally applicable techniques, such as mass spectrometry (MS), are required for this purpose.

To begin making this connection, we utilized MS and found the decreased expression of numerous single α-chains from several different collagen isoforms in HGSOC. Of particular note, the expression of Col I and Col III chains were both downregulated, where the latter is not consistent with previous immunostaining Data [31]. This is likely due to the qualitative nature of immunostaining and the lack of specificity of the available Col III antibodies.

Interestingly, we previously used detailed SHG analysis to show that there were large structural changes in the α-chains; however, these were not consistent with an increase in Col III expression [11]. Moreover, the triple helical structure was found to be more disorga-

nized in HGSOC. This disorganization was also validated by our wavelength dependent optical scattering measurements, which probed size scales over the range of ~50 nm–1μm [30]. Collectively, the MS Data and our previous macro/supramolecular SHG analyses suggest there may be transcriptomic and/or post-translational modifications in HGSOC coinciding with characteristic, pronounced changes in the collagen fiber morphology.

We did not expand the proteomic analysis to the p53 and STIC precursors as the tissues were insufficient in volume. However, given the current paradigm in HGSOC that the ovary is the metastatic site from the FT [16,35], the genetic modifications giving rise to different isoform expression are expected to be similar in both sites. Indeed, we showed that the collagen fiber morphology in HGSOC was highly similar in the FT and ovary [17], suggesting that similar proteomic modifications occur in both tissues.

Given the congruence in the SHG and the proteomic Data suggesting differential collagen isoform expression in normal and HGSOC tissues, these observations further support the validity of performing the MS analysis on more available ovarian stromal tissues. We further suggest that the SHG analysis of the collagen in precursors is a true reflection of the biochemical alterations occurring during carcinogenesis and is an important surrogate measurement that allows for non-destructive analysis, preserving the rare precursors for IHC and transcriptomic analysis.

Based on our prior work [17], we expected that the collagen organization differences between the two precursors (p53 only and STICs) would be less pronounced relative to that of high grade disease. However, our linear discriminant model still achieved good accuracies, F_1 scores, and ROC curves, which were further improved by grouping p53 signatures and STIC lesions together. Importantly, even with the small sample size, our analysis supports our hypothesis that collagen alterations in the FT occur prior to frank HGSOC.

We note that, since the collagen alterations of early-stage disease are subtle, the classifier performance will be sensitive to under-sampling. We suggest that with a much larger specimen number, the performance could be improved to a standard that is suitable for future clinical applications. The sample thickness and coverage and limited collagen density leading to relatively weak SHG signal intensities were major limitations in this study. We suggest that higher accuracies should be achievable on thicker sections (~50–100 μm), which are readily imaged by SHG.

Another limitation of the study at the moment is that there is no clear definition of the biochemical factors that affect the change in the collagen structure in HGSOC or in precursor lesions. The differential expression of collagen isoforms combined with specific post-translational processing may contribute to the change in collagen structure. The successful implementation of mass spectroscopy and transcriptomic analysis will be necessary to obtain the biochemical understanding of changes occurring in the collagen and the surrounding ECM. For example, other markers, such as fibronectin, laminin, and secreted MMPs, have been suggested to have altered regulation in HGSOC and could be added to the analysis.

We foresee both long and short-term applications based on our observations and methodology. For the former, it may be possible to construct an SHG laser-scanning micro-endoscope to be used in conjunction with laparoscopy or hysteroscopy [36,37]. Analogous fiber optic-based scanning approaches are currently under development for other pathologies [38,39], and these should be adaptable for FT imaging.

The scheme may be feasible for the in vivo detection of precursor lesions, especially since the majority of the significant variables for class discrimination came from the SHG backward channel, which would be the usable direction in an endoscopic configuration. In the shorter term, the ex vivo analysis of resected fallopian tube tissue from risk reduction surgery can be used either as a pre-screen or to complement the histology and also to identify MSbased proteomic correlations.

5. Conclusions

The collagen fiber morphology is highly remodeled in HGSOC in the ovary or the fallopian tubes, where the fibers become more aligned relative to normal tissues and other tumor sub-types. However, for clinical applications, it is important to investigate these alterations in the precursors (p53 and STICs) due to the early metastasis of HGSOC. Unfortunately, there is a limited availability of pure precursor lesions without the presence of malignant tissue.

Still, due to the high specificity and sensitivity to collagen morphology afforded by SHG microscopy, sufficient discrimination between distal normal regions and p53 and STIC precursors was attained in this limited study to demonstrate proof of concept. Moreover, mass spectrometry analyses showed concurrent proteomic changes in normal and HGSOC tissues, where many collagen single α-chains were downregulated. These results suggest that the combined use of MS proteomic and SHG microscopy analyses forms a basis for further in vivo and ex vivo explorations of HGSOC and its precursor lesions.

References

1. Bowtell, D.D. The genesis and evolution of high-grade serous ovarian cancer. *Nat. Rev. Cancer* **2010**, *10*, 803–808. [CrossRef] [PubMed]
2. Qayyum, A.; Coakley, F.V.; Westphalen, A.C.; Hricak, H.; Okuno, W.T.; Powell, B. Role of CT and MR imaging in predicting optimal cytoreduction of newly diagnosed primary epithelial ovarian cancer. *Gynecol. Oncol.* **2005**, *96*, 301–306. [CrossRef]
3. Karlan, B.Y. The status of ultrasound and color Doppler imaging for the early detection of ovarian carcinoma. *Cancer Investig.* **1997**, *15*, 265–269. [CrossRef] [PubMed]
4. Bristow, R.E.; Giuntoli, R.L., 2nd; Pannu, H.K.; Schulick, R.D.; Fishman, E.K.; Wahl, R.L. Combined PET/CT for detecting recurrent ovarian cancer limited to retroperitoneal lymph nodes. *Gynecol. Oncol.* **2005**, *99*, 294–300. [CrossRef] [PubMed]
5. Karlan, B.Y.; Thorpe, J.; Watabayashi, K.; Drescher, C.W.; Palomares, M.; Daly, M.B.; Paley, P.; Hillard, P.; Andersen, M.R.; Anderson, G.; et al. Use of CA125 and HE4 serum markers to predict ovarian cancer in elevated-risk women. *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 1383–1393. [CrossRef]
6. Bast, R.C., Jr.; Brewer, M.; Zou, C.; Hernandez, M.A.; Daley, M.; Ozols, R.; Lu, K.; Lu, Z.; Badgwell, D.; Mills, G.B.; et al. Prevention and early detection of ovarian cancer: Mission impossible? *Recent Results Cancer Res.* **2007**, *174*, 91–100. [PubMed]
7. Blyuss, O.; Burnell, M.; Ryan, A.; Gentry-Maharaj, A.; Marino, I.P.; Kalsi, J.; Manchanda, R.; Timms, J.F.; Parmar, M.; Skates, S.J.; et al. Comparison of Longitudinal CA125 Algorithms as a First-Line Screen for Ovarian Cancer in the General Population. *Clin. Cancer Res.* **2018**, *24*, 4726–4733. [CrossRef]
8. Wen, B.; Campbell, K.R.; Tilbury, K.; Nadiarnykh, O.; Brewer, M.A.; Patankar, M.; Singh, V.; Eliceiri, K.W.; Campagnola, P.J. 3D texture analysis for classification of second harmonic generation images of human ovarian cancer. *Sci. Rep.* **2016**, *6*, 1–9. [CrossRef]
9. Wen, B.L.; Brewer, M.A.; Nadiarnykh, O.; Hocker, J.; Singh, V.; Mackie, T.R.; Campagnola, P.J. Texture analysis applied to second harmonic generation image Data for ovarian cancer classification. *J. Biomed. Opt.* **2014**, *19*, 096007. [CrossRef]
10. Campbell, K.R.; Campagnola, P.J. Assessing local stromal alterations in human ovarian cancer subtypes via second harmonic generation microscopy and analysis. *J. Biomed. Opt.* **2017**, *22*, 116008. [CrossRef]
11. Campbell, K.R.; Chaudhary, R.; Handel, J.M.; Patankar, M.S.; Campagnola, P.J. Polarization-resolved second harmonic generation imaging of human ovarian cancer. *J. Biomed. Opt.* **2018**, *23*, 066501. [CrossRef]
12. Crum, C.P.; Drapkin, R.; Miron, A.; Ince, T.A.; Muto, M.; Kindelberger, D.W.; Lee, Y. The distal fallopian tube: A new model for pelvic serous carcinogenesis. *Curr. Opin. Obstet. Gynecol.* **2007**, *19*, 3–9. [CrossRef]
13. Labidi-Galy, S.I.; Papp, E.; Hallberg, D.; Niknafs, N.; Adleff, V.; Noe, M.; Bhattacharya, R.; Novak, M.; Jones, S.; Phallen, J.; et al. High grade serous ovarian carcinomas originate in the fallopian tube. *Nat. Commun.* **2017**, *8*, 1–11. [CrossRef] [PubMed]
14. Piek, J.M.; van Diest, P.J.; Verheijen, R.H. Ovarian carcinogenesis: An alternative hypothesis. *Adv. Exp. Med. Biol.* **2008**, *622*, 79–87. [CrossRef] [PubMed]
15. Mehra, K.; Mehrad, M.; Ning, G.; Drapkin, R.; McKeon, F.D.; Xian, W.; Crum, C.P. STICS, SCOUTs and p53 signatures; a new language for pelvic serous carcinogenesis. *Front. Biosci. Elite Ed.* **2011**, *3*, 625–634. [CrossRef] [PubMed]
16. Folkins, A.K.; Jarboe, E.A.; Saleemuddin, A.; Lee, Y.; Callahan, M.J.; Drapkin, R.; Garber, J.E.; Muto, M.G.; Tworoger, S.; Crum, C.P. A candidate precursor to pelvic serous cancer (p53 signature) and its prevalence in ovaries and fallopian tubes from women with BRCA mutations. *Gynecol. Oncol.* **2008**, *109*, 168–173. [CrossRef]
17. Rentchler, E.C.; Gant, K.L.; Drapkin, R.; Patankar, M.; Campagnola, P.J. Imaging Collagen Alterations in STICs and High Grade Ovarian Cancers in the Fallopian Tubes by Second Harmonic Generation Microscopy. *Cancers* **2019**, *11*, 1805. [CrossRef] [PubMed]
18. Backman, V.; Roy, H.K. Advances in biophotonics detection of field carcinogenesis for colon cancer risk stratification. *J. Cancer* **2013**, *4*, 251–261. [CrossRef]
19. Damania, D.; Roy, H.K.; Kunte, D.; Hurteau, J.A.; Subramanian, H.; Cherkezyan, L.; Krosnjar, N.; Shah, M.; Backman, V. Insights into the field carcinogenesis of ovarian cancer based on the nanocytology of endocervical and endometrial epithelial cells. *Int. J. Cancer* **2013**, *133*, 1143–1152. [CrossRef]
20. Radosevich, A.J.; Mutyal, N.N.; Yi, J.; Stypula-Cyrus, Y.; Rogers, J.D.; Goldberg, M.J.; Bianchi, L.K.; Bajaj, S.; Roy, H.K.; Backman, V. Ultrastructural alterations in field carcinogenesis measured by enhanced backscattering spectroscopy. *J. Biomed. Opt.* **2013**, *18*, 097002. [CrossRef]
21. Wang, J.; Lin, L.; Parkash, V.; Schwartz, P.E.; Lauchlan, S.C.; Zheng, W. Quantitative analysis of follicle-stimulating hormone receptor in ovarian epithelial tumors: A novel approach to explain the field effect of ovarian cancer development in secondary mullerian systems. *Int. J. Cancer* **2003**, *103*, 328–334. [CrossRef] [PubMed]
22. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef] [PubMed]
23. Naba, A.; Clauser, K.R.; Hoersch, S.; Liu, H.; Carr, S.A.; Hynes, R.O. The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteom.* **2012**, *11*, M111.014647. [CrossRef] [PubMed]
24. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics Data. *Nat. Methods* **2016**, *13*, 731–740. [CrossRef] [PubMed]
25. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and related tools and resources in 2019: Improving support for quantification Data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [CrossRef]

26. Chen, X.; Nadiarynkh, O.; Plotnikov, S.; Campagnola, P.J. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nat. Protoc.* **2012**, *7*, 654–669. [CrossRef] [PubMed]

27. Lacomb, R.; Nadiarnykh, O.; Townsend, S.S.; Campagnola, P.J. Phase Matching considerations in Second Harmonic Generation from tissues: Effects on emission directionality, conversion efficiency and observed morphology. *Opt. Comm.* **2008**, *281*, 1823–1832. [CrossRef]

28. Lacomb, R.; Nadiarnykh, O.; Campagnola, P.J. Quantitative SHG imaging of the diseased state Osteogenesis Imperfecta: Experiment and Simulation. *Biophys. J.* **2008**, *94*, 4504–4514. [CrossRef]

29. Bredfeldt, J.S.; Liu, Y.; Pehlke, C.A.; Conklin, M.W.; Szulczewski, J.M.; Inman, D.R.; Keely, P.J.; Nowak, R.D.; Mackie, T.R.; Eliceiri, K.W. Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer. *J. Biomed. Opt.* **2014**, *19*, 016007. [CrossRef]

30. Tilbury, K.B.; Campbell, K.R.; Eliceiri, K.W.; Salih, S.M.; Patankar, M.; Campagnola, P.J. Stromal alterations in ovarian cancers via wavelength dependent Second Harmonic Generation microscopy and optical scattering. *BMC Cancer* **2017**, *17*, 1–11. [CrossRef]

31. Ricciardelli, C.; Rodgers, R.J. Extracellular matrix of ovarian tumors. *Semin. Reprod. Med.* **2006**, *24*, 270–282. [CrossRef] [PubMed]

32. Kauppila, S.; Bode, M.K.; Stenback, F.; Risteli, L.; Risteli, J. Cross-linked telopeptides of type I and III collagens in malignant ovarian tumours in vivo. *Br. J. Cancer* **1999**, *81*, 654–661. [CrossRef] [PubMed]

33. Tilbury, K.; Lien, C.H.; Chen, S.J.; Campagnola, P.J. Differentiation of Col I and Col III Isoforms in Stromal Models of Ovarian Cancer by Analysis of Second Harmonic Generation Polarization and Emission Directionality. *Biophys. J.* **2014**, *106*, 354–365. [CrossRef] [PubMed]

34. Ajeti, V.; Nadiarnykh, O.; Ponik, S.M.; Keely, P.J.; Eliceiri, K.W.; Campagnola, P.J. Structural changes in mixed Col I/Col V collagen gels probed by SHG microscopy: Implications for probing stromal alterations in human breast cancer. *Biomed. Opt. Express* **2011**, *2*, 2307–2316. [CrossRef]

35. Piek, J.M.; Kenemans, P.; Zweemer, R.P.; van Diest, P.J.; Verheijen, R.H. Ovarian carcinogenesis, an alternative theory. *Gynecol. Oncol.* **2007**, *107*, 355. [CrossRef]

36. Keenan, M.; Tate, T.H.; Kieu, K.; Black, J.F.; Utzinger, U.; Barton, J.K. Design and characterization of a combined OCT and wide field imaging falloposcope for ovarian cancer detection. *Biomed. Opt. Express* **2017**, *8*, 124–136. [CrossRef] [PubMed]

37. Akhoundi, F.; Qin, Y.; Peyghambarian, N.; Barton, J.K.; Kieu, K. Compact fiber-based multi-photon endoscope working at 1700 nm. *Biomed. Opt. Express* **2018**, *9*, 2326–2335. [CrossRef]

38. Wu, Y.C.; Leng, Y.X.; Xi, J.F.; Li, X.D. Scanning all-fiber-optic endomicroscopy system for 3D nonlinear optical imaging of biological tissues. *Opt. Express* **2009**, *17*, 7907–7915. [CrossRef]

39. Williams, J.C.; Campagnola, P.J. Wearable Second Harmonic Generation Imaging: The Sarcomeric Bridge to the Clinic. *Neuron* **2015**, *88*, 1067–1069. [CrossRef]