

Differential extracellular matrix proteomic signatures in colorectal tumors from Appalachian and non-Appalachian patients

ALEXANDER T. SOUGIANNIS¹, HARRISON B. TAYLOR², STEPHEN C. ZAMBRZYCKI²,
LINDSEY CONROY³, RACHEL STRUBLER², CHRISTIN EDGE², RICHARD R. DRAKE²,
KRISTIN WALLACE⁴, DEREK ALLISON⁵, EUN LEE⁵, RAMON C. SUN^{3,6} and PEGGI M. ANGEL²

¹College of Medicine, Medical University of South Carolina, Charleston, SC 29403, USA;

²Department of Cell and Molecular Pharmacology and Experimental Therapeutics, Medical University of South Carolina, Charleston, SC 29403, USA; ³Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, KY 40506, USA;

⁴Department of Public Health Services, College of Medicine, Medical University of South Carolina, Charleston, SC 29403, USA;

⁵Department of Pathology and Laboratory Medicine, Markey Cancer Center, University of Kentucky, Lexington, KY 40506, USA;

⁶Department of Biochemistry, University of Florida, Gainesville, FL 32610, USA

Received February 6, 2024; Accepted September 25, 2025

DOI: 10.3892/ol.2025.15159

Abstract. Emerging evidence reports that regulation of the extracellular matrix influences the progression of colorectal cancer (CRC). The present study investigated regulation of the extracellular matrix proteome in colorectal malignancy within a high-risk Appalachian population compared with non-Appalachian populations. A targeted mass spectrometry imaging proteomic method directed at collagen regulation was used. Tissue microarrays (TMAs) comprising of matched CRC with adjacent normal to tumor (NAT) from 45 patients were constructed into 86 samples to evaluate the extracellular matrix proteome (ECM). A total of five specific peaks were discovered to differ between NAT and tumor with high sensitivity and specificity by receiver operating characteristic (AUROC) ≥ 0.7 , Wilson/Brown $P < 0.0002$. Evaluation of patient TMA cores showed increased levels of combined ECM peptides in advanced stage Appalachian CRC (III + IV) compared with early staged CRC (I + II) (AUROC 0.8595; 95% confidence interval, 0.8190-0.8999; Wilson/Brown $P < 1.0 \times 10^{-15}$), contrasting with the non-Appalachian tumors, which showed a decreased ability to discriminate between early and late stage (AUROC 0.6618; 95% confidence interval, 0.6126-0.7110; Wilson/Brown $P < 1.0 \times 10^{-9}$). Comparison of advanced stage CRCs between Appalachian and non-Appalachian populations showed high sensitivity and specificity in distinguishing the populations (AUROC 0.7612; 95% confidence interval,

0.7109-0.8114; Wilson/Brown $P < 3.0 \times 10^{-15}$). History of smoking, sex and tumor origin location did not show significant ability to distinguish by AUROC. A combination of high mass resolution, high mass accuracy spatial proteomics and sequencing proteomics by liquid chromatography coupled to tandem mass spectrometry revealed that fibrillar collagens were spatially regulated within the CRC tumor microenvironment. Fibrillar collagen post-translational modifications of hydroxylated proline revealed distinct spatial separation based on the presence of a number of hydroxylated proline sites. The present study highlighted that the targeted mass spectrometry imaging of the ECM proteome may provide new insight and novel predictive tools for understanding CRC, particularly among Appalachian patients.

Introduction

Colorectal cancer (CRC) is the third most common malignancy for men and women. Treatment options for advanced stage CRC are limited and survival remains poor relative to early stage CRC (i.e. $<20\%$ to $>90\%$) (1-9). Underserved and geographically isolated populations, such as Appalachian and African Americans, often present with at diagnosis with more aggressive CRCs (relative to other groups) and have higher mortality (9-12). Appalachian populations are diagnosed with CRC at a later stage of disease which is associated with poor survival (13-15). Although the causes of the disparities are not fully understood, health services factors (e.g., access to care), and income related variables contribute to poorer outcomes (13,14,16,17). Further, lifestyle differences such as smoking, obesity, and diabetes have been shown to contribute to increased cancer burden and poor prognosis (13,14,16,17). Nevertheless, after adjustment for health services, income-related variables (18,19), and lifestyle behavioral characteristics, appreciable differences in outcomes remain, highlighting the need to uncover other factors that contribute to the differences in outcomes in underserved populations.

Correspondence to: Dr Peggi M. Angel, Department of Cell and Molecular Pharmacology and Experimental Therapeutics, Medical University of South Carolina, 173 Ashley Avenue BSB 358, Charleston, SC 29403, USA
E-mail: angelp@musc.edu

Key words: imaging, spatial proteomics, collagen, extracellular matrix, cancer, mass spectrometry imaging, colorectal cancer

Understanding molecular differences within CRCs within this population may influence clinical decision making and improve prognosis of all high-risk CRC patients.

Pathological progression of CRC has a significant involvement of collagen alteration during initiation, progression, and metastasis (20-23). Collagen interfacing pathways such as discoidin domain signaling, integrin signaling, metalloproteinases, and cross-linking lysyl hydroxylase, and prolyl hydroxylase enzymes are systematically altered and predictive of prognosis and survival (24-28). Translational regulation of specific collagen types contributes to expansion of the tumor microenvironment, promotes cell migration, alters cell adhesion, and induces endothelial mesenchymal transition (EMT) (29-31). Biomechanical properties of the extracellular matrix (ECM) microenvironment are altered and premalignant CRC tumor tissues have been reported to have increased stiffness of stromal ECM compared to normal tissues (32-34). Products of active collagen regulation are detected from early stages in both tissue and serum (21,35,36) with high levels of collagen α -1(I) detected in urine during metastatic stages (37,38). However, the spatial regulation collagen expression as defined by pathological cell feature and related to specific population disparities within the tumor microenvironment remains poorly defined.

Several tools exist to evaluate the ECM in CRC. *Ex-vivo* spatial imaging tools such as multiphoton microscopy (MPM) (39,40), second harmonic imaging (SHI) (41) or MPM-SHI combined (42,43) have been used to study the organization of collagen fibers in colon cancer. These MPM and SHI studies have shown collagen degradation in CRC tumor formation and as the tumor becomes invasive, collagen fibers become straighter and denser. Collagen signatures of the CRC tumor microenvironment have also been measured by MPM and SHI and are predictive of recurrence and survival (40). Imaging techniques using Raman or infrared spectroscopy elucidate basic chemistries to distinguish cancerous CRC tissue *ex-vivo* (44,45). *In vivo* tools such as magnetic resonance imaging or Raman & SHG endoscopy may be used as a CRC measurement to spatially image the collagen structure during a fibrotic process (46-48). Very few reports exist towards understanding the molecular details of collagen translational and post-translational regulation within the CRC microenvironment.

In this study, we examine specific collagen translational and post-translational changes that could distinguish normal and CRC tissues and show unique spatial contributions to the CRC tumor microenvironment. Special attention is paid to comparing Appalachian and non-Appalachian residents of eastern Kentucky and southern Ohio. We previously reported a strategy for targeted spatial proteomic imaging of the collagen type structures and other extracellular matrix (ECM) proteins from clinically archived formalin-fixed, paraffin-embedded tissues by mass spectrometry imaging (MSI) (49-51). The strategy can uniquely be used for tissue microarray mass spectrometry imaging (TMA-MSI) to allow assessment of larger patient numbers within highly defined pathologies and for tissue sections towards investigating molecular gradients across the tissue microenvironment. This study specifically reports differences in collagen peptides between CRC and normal adjacent to tumor (NAT) with high predictive power,

particularly for late-stage CRC within the Appalachian population. Further examination of select CRC resections demonstrates complex translational and post-translational spatial regulation of the ECM microenvironment within CRC. This proof-of-concept study advances our understanding of collagen regulation in CRC pathologies towards finding new key diagnostic or therapeutic targets in populations, such as Appalachian, that are at increased risk for CRC.

Materials and methods

Tissue microarray construction. Colon tissues were collected in Y2002-2009 from residual surgical tissue by the Markey BPTP-SRF with approval by the University of Kentucky IRB for use under a waiver of consent (IRB13-0608). Cases were screened and selected by Dr. Eun Lee with the UK Dept. of Pathology and Laboratory Medicine and representative areas of tumor and normal selected for coring. Cases selected were annotated by the Markey Cancer Research Informatics Facility and the tissue microarray (TMA) was constructed by the Markey Cancer Center BPTP-SRF. Inclusion criteria were age 40-85 as the target risk population, female and male, and smoking or nonsmoking status data. Exclusion criteria were pregnant female. Appalachian vs. non-Appalachian origin was determined by patient's county of residence. All patients self-identified as white. Clinical data included history of smoking and location of tumor (proximal, distal, rectal) as influence on disease prognosis. Slides of the TMA and associated data were de-identified before dispensing for research. Resections of colorectal cancer were obtained through the Hollings Cancer Center Biorepository. Resections that demonstrated malignant polyps of male or female age similar in age were selected as representatives of collagen gradients across the colon tissue. Analysis of tissue microarrays and tissues were approved as exemption #4 by the Medical University of South Carolina IRB. Maps of counties were created through the publicly available USGS National Map National Boundaries deposited in the public domain and free for public use; Data on counties are available from U.S. Geological Survey, National Geospatial Program.

TMA and CRC resection proteomic imaging analysis. TMAs and tissues were prepared and analyzed for extracellular matrix proteomic signatures as described previously (49,50,52-54) and included an internal standard, (GluFib peptide m/z 1570.6768) spiked into the matrix. TMAs were analyzed by MALDI QTOF (timsTOF fleX, Bruker) in positive ion mode, collecting 300 laser shots per pixel with stepsizes of 80 μ m between pixels. Data were acquired over m/z range 600-2,500. Focus PreTOF transfer time was 75 μ s, Pre Pulse Storage was 20 μ s, quadrupole ion energy was set to 15.0 eV with a low mass of 500 m/z. Data were analyzed using SCiLS^(TM) Lab software 2021C (Bruker Scientific, LLC, Bremen, Germany). After normalization to the internal standard, extracted peak intensities were exported from SCiLS for statistical analysis. Hierarchical clustering analysis of TMA data was performed using the Cluster and TreeView software tool programs originally developed for analyzing cDNA microarray data (55).

Resections of CRC were analyzed on a Fourier Transform Ion Cyclotron Resonance mass spectrometer (Scimax, Bruker)

equipped with a dual source matrix assisted laser desorption/ionization (MALDI and electrospray ionization (ESI). Data was acquired in broadband positive ion mode with 200 shots per pixel with transient lengths of 1.3282 over a m/z range of 544-2,500. Image data were analyzed using SCiLS Lab software 2023A (Bruker Scientific, LLC, Bremen, Germany). Principal components analysis was performed on individual spectra derived from pathologies annotated as muscularis, submucosa and tumor using unit variance scaling on 5 components within SCiLS software. Hierarchical image clustering on colorectal resections utilized a bisecting k-means algorithm with the Manhattan metric.

Sequencing proteomics on CRC resections. Data were collected on an Orbitrap Eclipse mass spectrometer (ThermoFisher Scientific) coupled to a Dionex Ultimate 300 RSLCnano system (ThermoFisher Scientific). One microgram of the recovered peptides from each sample was injected onto a 5 mm nanoviper μ -Precolumn (i.d.300 μ m, C18 PepMap 100, 5.0 μ m, 100 Å) from ThermoFisher Scientific at 5 μ l/min in formic acid/H₂O 0.1/99.9 (v/v) for 5 min to desalt and concentrate the samples. For the chromatographic separation of peptides, the trap-column was switched at 5 min to align with the EASY-Spray column PepMap RSLC C18 with a 150 mm column (i.d. 75 μ m, C18, 3.0 μ m, 100 Å). The peptides were eluted using a variable mobile phase (MP) gradient from 98% phase A (Formic acid/H₂O 0.1/99.9, v/v) to 32% phase B (Formic Acid/Acetonitrile 0.1/99.9, v/v) over 60 min (from 5-65 min) at 300 nl/min followed by a high organic wash up to 90%B at 66 min, hold for 10 min and return to initial conditions at 77 min to re-equilibrate at 90 min. MS1 data were collected in the Orbitrap (120,000 resolution; maximum injection time 50 ms; AGC 4×10^5). Charge states between 2 and 6 were acquired for MS2 analysis, and a 20 sec dynamic exclusion window was used. Cycle time was set at 2.5 sec. MS2 scans were performed in the ion trap with HCD fragmentation (isolation window 0.8 Da; NCE 30%; maximum injection time 40 ms; AGC 5×10^4). The data was recorded using Thermo Scientific Xcalibur 4.5 software. Data was searched through MaxQuant as previously done(49,50,52-54) and filtered to peptides with scores ≥ 70 with reporting of site modification probabilities for hydroxylated proline. Putative peptide identities were assigned to TMA image data by accurate mass matching under 10 mDa and to resection image data by accurate mass matching ≤ 5 ppm. Previous databases of collagen peptides found by the method (49,50,52-54,56) were used to further support putative peptide identifications.

Statistical analysis. All samples were included in statistical analysis. Peak height analysis was performed using commercial software (SigmaStat V3.5, SPSS, Chicago, IL). Clinical characteristics were compared by Chi-square P-value. The selection of clinically important peaks was determined by one-way ANOVA or two-way ANOVA when appropriate. Individual peptides were evaluated by Mann-Whitney U test on natural log normalized peak intensities. Statistical significance was set with an α value of $P < 0.05$. Data are represented as mean $\pm$ standard error of the mean (SEM). Peaks that yielded statistically significant differences based on ANOVA and multiple comparisons were utilized for meta-analyses and

analysis for clinical correlations. Clinically significant peaks were determined based on individual and then combined peak scores based on receiver operating characteristic (ROC) curve analysis (Prism 9.2.0, La Jolla, CA). This analysis generated Area Under the Receive Operating Curve (AUROC) values which were utilized to determine clinical significance from each analysis (≥ 0.70 ; Wilson/Brown P-value ≤ 0.001). Predictive individual peaks were combined into a single predictive reading to explore AUROC values as a combined evaluation. Principal components analysis within SCiLS software (Bruker Scientific, LLC, Bremen, Germany) was completed on colorectal tissue pathologies used unit variance scaling, 5 components and evaluated 2,453 peaks from individual spectra originating from pathologically defined region of tumor (18,948 spectra), muscularis (12,599 spectra) and submucosa (19,157 spectra).

Results

Overview of the study. A goal in this study was to understand the potential contribution of collagen proteomic regulation in normal adjacent to tumor (NAT) and tumor and the relationship to colorectal cancer disparities observed between United States Appalachian (App) region compared to Non-Appalachian (N-App) region (Fig. 1A). Consented cases collected from 2014-2019 were annotated as App (n=19) or N-App (n=25) based on the patient's county of residence within the Kentucky Appalachian or Non-Appalachian regions (Fig. 1B). There was no significant difference in clinical data included self-reported race, sex, body-mass-index (BMI), smoking, location (proximal, distal, rectum) and tumor stage; Appalachian patients were significantly older at diagnosis compared to Non-Appalachian patients (Table I). TMA block and core selection was performed by a board-certified pathologist and up to three cores were selected per patient dependent on the availability of the tissue (Fig. 1B). Data includes pathology hematoxylin and eosin stain and peptide intensity by collagen targeted mass spectrometry imaging (Fig. S1). Collagen peptide regulation is systematically examined in NAT compared and tumor stage, evaluating for Appalachian specific regulation based on clinical characteristics. A subset of colorectal cancer tissue resections was further analyzed for overall tissue gradient patterns and sequence information by chromatography coupled to tandem mass spectrometry. Putative peptide identifications (49,50,52-54,56) to peaks found in the tissue microarray data are presented as a compilation of the peptide database from proteomics in this study summarized with comparison to previously published databases (Table II) (57,58). This study suggests that spatially localized translational and post-translational regulation of collagen contributes to increased tumor burden and changes with malignancy and increased risk factors.

Specific collagen peptides differentiate tumor independent of appalachian status & tumor stage. Collagen remodeling is a hallmark of colorectal cancer with significant roles in conversion from precancerous and cancerous lesions (1,2,22,57,58). TMA cores were initially compared based on NAT (n=57) and tumor (n=52) status independent of Appalachian status or tumor stage (Fig. 2). A total of 311 manually curated

Table I. Clinical characteristics of the study.

Variable	Appalachian, n ^a	Non-appalachian, n ^a	P-value
Region by county			
Total patients	25	29	
Total cores	47	54	
Sex (cores)			0.4351
Male	12 (23)	17 (31)	
Female	13 (24)	12 (23)	
Age at diagnosis, n (95% CI)	69.0 (64.4,7)	59.2 (54.1,64.4)	0.0025
BMI at diagnosis, n (95% CI)	26.4 (24.9,27.9)	25.9 (23.5,28.2)	0.1661
Lifestyle (cores)			0.1179
Smoking	13 (23)	9 (18)	
Non-smoking/unknown	12 (24)	20 (36)	
Tumor location (cores)			0.9305
Proximal	6 (12)	8 (15)	
Distal	12 (21)	14 (22)	
Rectum	7 (14)	7 (17)	
Tumor stage (cores)			0.2725
Stage I	6 (11)	11 (21)	
Stage II	7 (13)	2 (4)	
Stage III	4 (8)	7 (14)	
Stage IV	2 (4)	5 (8)	
Unknown	6 (11)	4 (7)	

^aNumbers in parenthesis indicates multiple cores from within the same patient population.

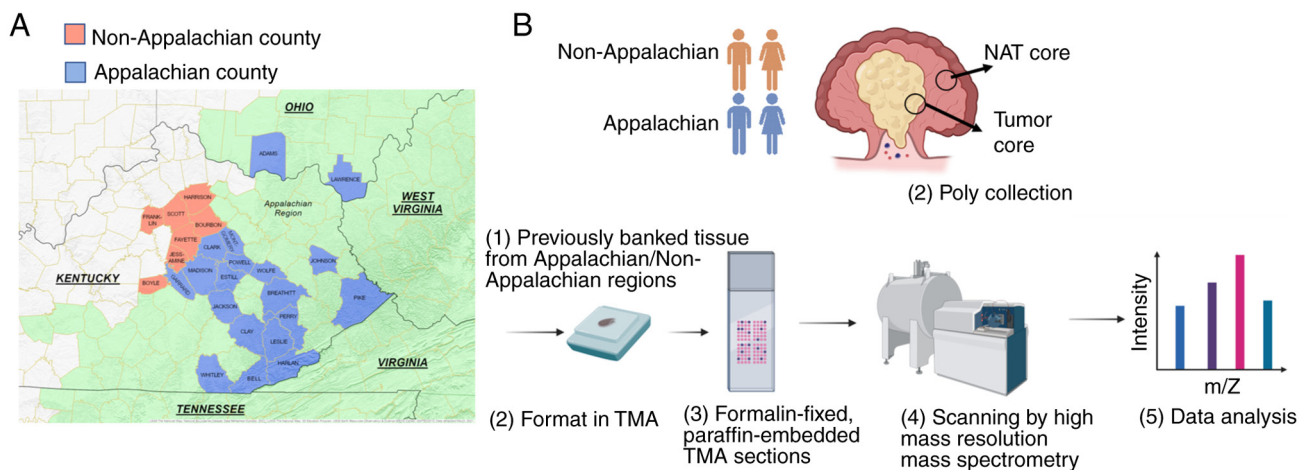


Figure 1. Study design and workflow. (A) Regional Appalachian and non-Appalachian counties inform which samples were collected in the study. Map was constructed from the publicly available U.S. Geological Survey, National Geospatial Program (<https://www.usgs.gov/programs/national-geospatial-program>). (B) Previously banked patient matched tumors and NAT from men and women were used in the study (panel 1). Cores were selected by a pathologist from the tumor or NAT and formatted into a TMA (panel 2). TMAs were sectioned for proteomic imaging analysis. Tumors were further annotated by stage. Pathologist-designated cores were used to create a formalin-fixed, paraffin-embedded tissue microarray (panel 3). Tissue sections of the TMA prepared for collagen targeted proteomic analysis were scanned by mass spectrometry; Each core was sampled ~225 times (panel 4). Data analysis used peak intensities as a total score per core (panel 5). Images created using Biorender (Biorender.com). NAT, normal adjacent to tumors; TMA, tissue microarray.

monoisotopic peptide peaks showed distinct clusters based on peptide intensity (Fig. 2A). When evaluated based on benign or malignant status, five peptide peaks showed significant alteration with area under the receiver operating curve (AUROC) demonstrating sensitive and specific discrimination per NAT or

tumor status (Fig. 2B-F). Putative peptide identifications from previous published accounts and from proteomics within this study on collagen regulation in colorectal resections reported that the majority of the peptides are from the fibrillar collagens collagen α -1(I), collagen α -2(I), collagen α -1(III) (Table II).

Table II. Putative peptide identifications to tissue microarray data.

Obs. Mass	Theor. Mass	Mass Error (mDa)	Sequence	Amino acid no.	No. HYP	Gene	Protein description	Accession	Database reference (Refs.)
843.394	843.396	2.0	GPAGERGSP	506-514	1	COL1A1	Collagen α -1(I) chain	P02452	(53)
870.404	870.406	2.0	GPAGQDGRP	557-565	1	COL1A1	Collagen α -1(I) chain	P02452	(53)
897.438	897.443	5.0	GVRGEPGPP	368-376	2	COL1A1	Collagen α -1(I) chain	P02452	(53,56) ^a
1034.585	1034.574	-11.0	GPVGPVVGARGPA	1,076-1,087	0	COL1A1	Collagen α -1(I) chain	P02452	(53,56)
1041.540	1041.532	-8.0	GAVGAKGEAGPQ	347-358	0	COL1A1	Collagen α -1(I) chain	P02452	(53) ^a
1082.634	1082.632	-2.0	IAGQGVVGLP	954-964	1	COL1A1	Collagen α -1(I) chain	P02452	(53) ^a
	1082.632	-2.0	GPLGIAGLTGAR	946-956	0	COL3A1	Collagen α -1(III) chain	P02461	(53) ^a
1242.574	1242.582	8.0	GPAGARGNDGATGAA	317-331	0	COL1A1	Collagen α -1(I) chain	P02452	(53,56) ^a
1343.675	1343.670	-5.0	GIAGAPGFPARGPS	404-418	2	COL1A1	Collagen α -1(I) chain	P02452	(53,56) ^a
801.388	801.385	-3.0	GPAGATGDR	685-693	0	COL1A2	Collagen α -2(I) chain	P08123	(57)
827.431	827.426	-5.0	GPAGEVGKP	562-570	1	COL1A2	Collagen α -2(I) chain	P02461	(57) ^a
	827.437	6.0	GPIGSRGPS	604-612	0	COL1A2	Collagen α -2(I) chain	P08123	(53)
875.439	875.426	-1.3	GARGFP(1)GTP	163-171	1	COL1A2;	Collagen α -2(I) chain;	P08123;	(53,54,56) ^a
				273-281		COL2A1	Collagen α -1 (II)	P02458	
758.382	758.379	-3.0	P(1)GERGGLG	728-735	1	COL3A1	Collagen α -1(III) chain	P02461	(53) ^a
843.429	843.432	3.0	GLPGAAGER	477-485	1	COL3A1	Collagen α -1(III) chain	P02461	(53,56) ^a
1172.520	1172.518	-2.0	GSPGERGETGPP	795-806	2	COL3A1	Collagen α -1(III) chain	P02461	(53) ^a
1223.619	1223.627	8.0	GEANTLPGPVGVP	1,232-1,244	1	COL4A2	Collagen α -2(IV) chain	P08572	(53,56)
1491.681	1491.712	31.0	LANMTEQIRQDR	138-149	0	COL6A2	Collagen α -2(VI) chain	P12110	(54)

^aIndicates main peptide form sequenced by present study. DB-database references: Obs, observed mass; Theor, theoretical mass; HYP, hydroxylated proline.

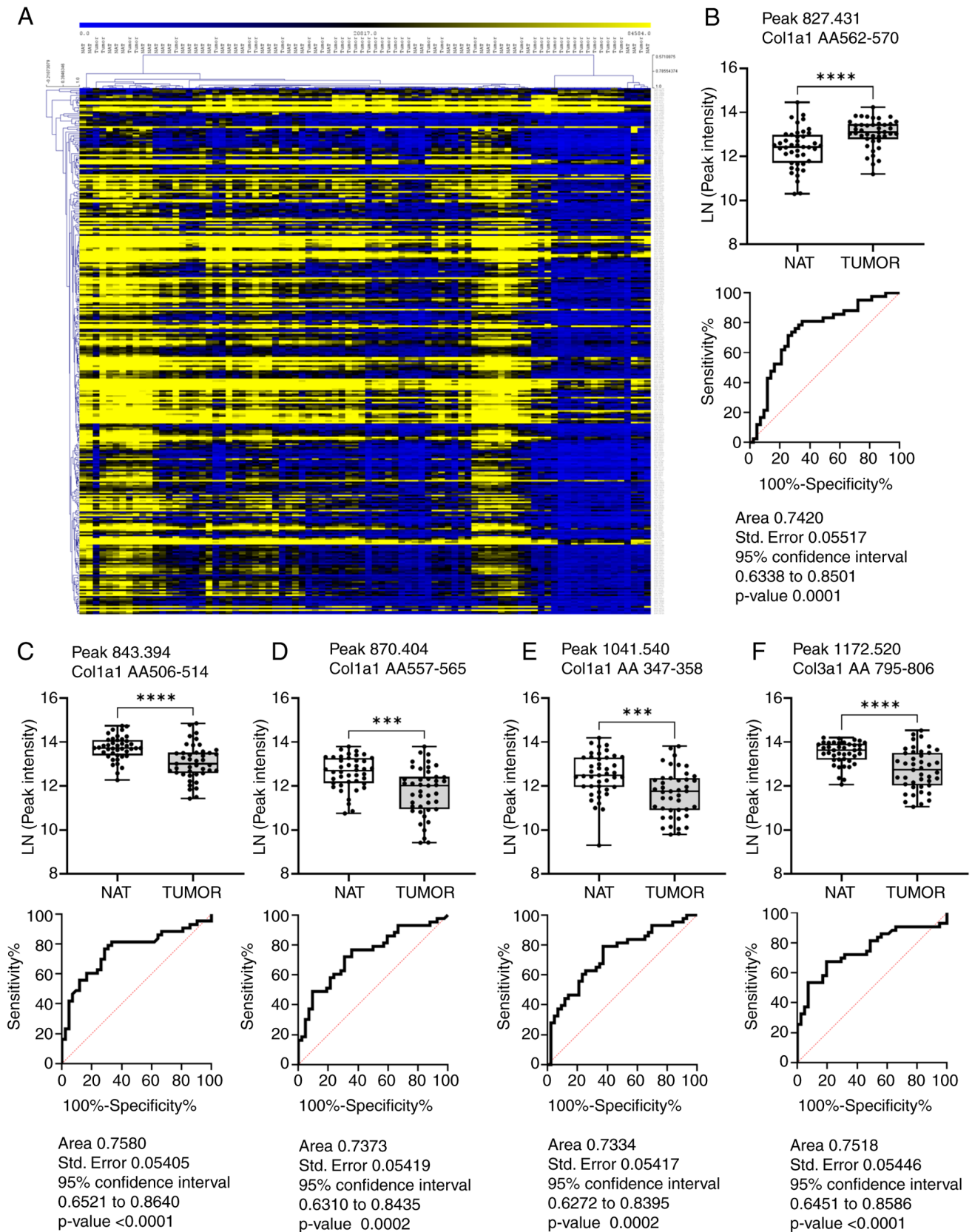


Figure 2. Evaluation of NAT versus tumor independent of Appalachian status and independent of tumor stage. (A) Peptide peaks showed distinct clusters based on peptide intensity. (B-F) Five peptide peaks were altered in NAT vs. tumor per $P \leq 0.001$. (B) Peptide 827.431, collagen α -1(I) chain amino acid domain 562-570, showed significant increases in tumor with AUROC of 0.7420 ($P=0.00001$); (C) Peptide 843.394, collagen α -1(I) chain amino acid domain 506-514, showed significant decreases in tumor with an AUROC of 0.7580 ($P=0.0001$); (D) Peptide 870.404, collagen α -1(I) chain amino acid domain 557-565, decreased in tumor with an AUROC of 0.7373 ($P=0.0002$); (E) Peptide 1041.540, collagen α -1(I) chain amino acid domain 347-358, decreased in tumor with an AUROC of 0.7334, ($P=0.0002$); (F) Peptide 1172.520, collagen α -1(III) chain amino acid domain 795-806 decreased in tumor with an AUROC of 0.7518 ($P<0.0001$). AA designates amino acid domain per each collagen. Peptides showed sensitive and specific discrimination between NAT and tumor by area under the receiver operating curve (AUROC) ≥ 0.7 and Wilson/Brown $P \leq 0.0001$. *** $P < 0.001$ and **** $P < 0.0001$. NAT, normal adjacent to tumors; Col, collagen; LN, natural log.

Sites were primarily from the collagen triple helical region, and many contained post-translational modifications with variable sites of hydroxylated proline. To summarize, specific collagen peptides were found that appeared to differentiate between normal adjacent to tumor or tumor status.

Collagen peptides discriminate malignant stage III + IV appalachian tumor from non-appalachian tumors. Tumor cores were investigated comparing between geographical Appalachian (App) or Non-Appalachian (N-App) region and by stage. A total of sixteen peptides were found altered between combined early-stage CRC (I+II) and late stage (III+IV) (Fig. 3). Putative peptide identifications linked altered peptides primarily to fibrillar collagens (Fig. 3A, Table II). Higher stages showed increased levels of collagen peptides, corresponding with previous literature reports that processes of fibrosis and deposition of collagen α -1(I) are associated with CRC progression (21,57,58). In most cases, collagen peptide levels were increased in Appalachian patient cores compared to Non-Appalachian patient cores. Combined peptide peaks that were altered between stage I+II Appalachian compared to Non-Appalachian showed no discriminatory power by AUROC (Fig. 3B, AUROC 0.532; 95% CI, 0.4809-0.5947; Wilson/Brown P-value 2.165E-1). Comparison within the Appalachian population by early and late-stage CRCs reported high discriminatory power of the combined peptide signature in differentiating tumor stage (Fig. 3C, AUROC 0.859; 95% CI, 0.819-0.899; Wilson/Brown P-value 1.0E-15). Contrasting with this, Non-Appalachian patient cores reported that the same combined peptide signature reported lower discriminating power between stages I=II and Stage III=IV (Fig. 3D, AUROC 0.662; 95% CI, 0.613-0.711; Wilson/Brown P-value 1.08E-9). In later stage III + IV tumor, comparison of Appalachian to Non-Appalachian patient cores presented high discriminatory values (Fig. 3E, AUROC 0.761; 95% CI, 0.719-0.811, Wilson/Brown P-value 3.0E-15). There were no discriminating differences in peptide signatures by sex, smoking status or origin location of the tumor, compared within Appalachian or compared to non-Appalachian patient cores (Figs. S2-S4). Overall, the data suggests that very specific collagen peptide increases may be associated with later stage disease in Appalachian CRC patients.

Colorectal resections display complex pathological gradients of peptides by collagen-targeted mass spectrometry imaging. To further understand the pathological role of specific collagen peptide regulation in CRC, tissue sections of CRC were analyzed by high mass resolution, high mass accuracy mass spectrometry imaging. Imaging experiments were followed by chromatography coupled to tandem mass spectrometry to sequence collagen peptides from CRC tissue (Fig. 4A). CRC resections used in the study were 3-8 cm (Fig. 4B). Total ion current demonstrated reported 4,696 features above a threshold absolute intensity of 1,000, including isotopes (Fig. 4C). Spectra were highly complex with near-isobaric peaks (Fig. 4D) having substantially different distribution patterns (Fig. 4E). Heuristic clustering (Figs. 4F and S5) showed a complex peptide population following pathological annotation (Fig. 4F) across the CRC resections. Principal Component Analysis of the 4,696 peak set showed separation of tumor, muscularis, and submucosa spectra

with location of spectra from pathology accounting for 25.6% of the variability in component 1 (Figs. 4F and S6; Table SI). Single and combined visualization of PCA components further highlighted distinct pathological distribution patterns of peptide populations (Fig. 4G). A conclusion is that by collagen targeted mass spectrometry imaging reports significant and complex pathological distribution patterns within the colorectal tumor microenvironment.

Collagen post-translational modifications align with specific CRC pathologies. To further investigate pathological variation associated with collagen post-translational modifications, we evaluated images from peptides significant in tumor compared to normal adjacent to tumor found in the tissue microarray data. An important distinction is that in the TMA data, the core was chosen from the center of the tumor whereas in the CRC resections, the entire resection (3-8 cm) is shown. This allows visualization of molecular gradients across several pathologies. Each CRC showed a diversity of spatial regulation related to collagen α -1(I) chain and collagen α -2(I) chain mapping same m/z previously found in the TMA found as differential in tumor vs. NAT or in Appalachian vs. Non-Appalachian (Fig. 5A). Interestingly, higher intensity of identified collagen peaks was found adjacent to the primary tumor. Evaluation of peak intensity based on proline hydroxylation status varied in each tissue with hydroxylated sites being more intense (Fig. 5B). Mapping each collagen peptide domain by site variant showed differential spatial patterns [Fig. 5C, example peptide GPIGSRGPS amino acids #604-612 from collagen α -2(I)]. In certain cases, overlaying modified and unmodified versions demonstrated complementary patterns related to collagen post-translational regulation (Figs. 5D and S7-S9). In summary, collagen post-translational modification of hydroxylated proline appears spatially regulated in CRC with increased intensity adjacent to tumor.

Discussion

Colorectal cancer (CRC) begins as a submucosal invasion of tumor cells that forms the initial tumor and subsequently advances into the muscularis propria and then perirectal adventitia (58-61). High stromal content is a poor prognosticator as well as advanced stage, higher grade, and invasion of the perineural and lymphovascular regions (62,63). Re-organization of the stromal extracellular matrix remodeling plays a significant role throughout colon cancer including metastasis (20,23,64-67). In CRC, collagen systematically changes from irregular wavy structures to highly dense and linearly packed fibers that have been shown to increase migration of cancer activated fibroblasts (68,69). Fibrillar collagens contain protein domains that are a rich systems biology (70) facilitating communication through the cell-matrix interactions. Alterations in collagen organization reflect differential exposure of the collagen domains involved in cell binding, signaling, and protease interaction (38,71-77). In this study we leverage mass spectrometry imaging to examine changes in collagen peptide variation from staged CRC as a tissue microarray and as tissue sections. This study included tumors from the Appalachian population that have been shown to have higher risk of developing more aggressive cancers including

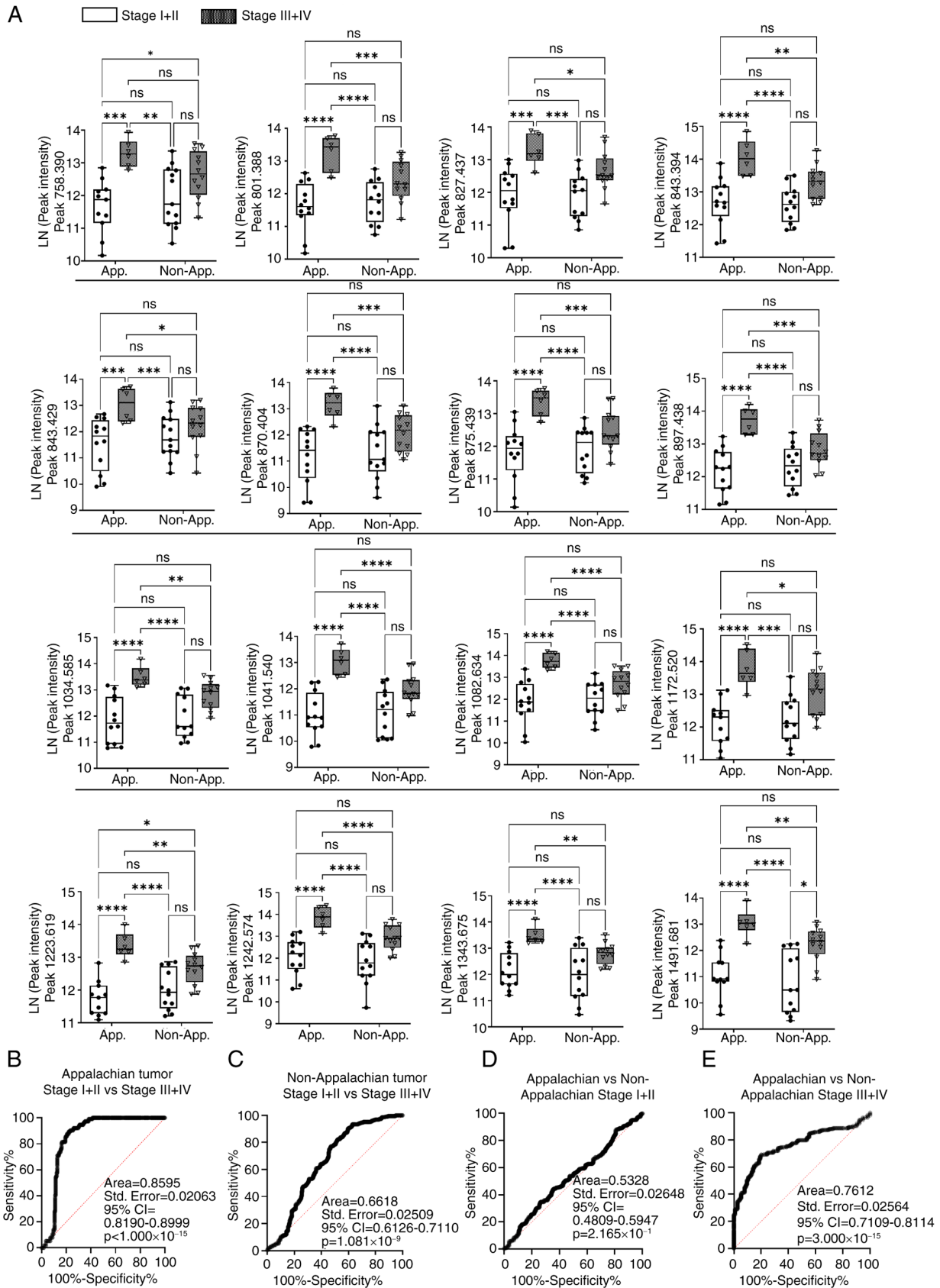


Figure 3. CRC tumors show alterations based on stage and regionalized Appalachian or non-Appalachian county origins. Stages are combined as Stage I + II (early) and Stage II + IV (late). (A) A total of sixteen peptides reported potential differences compared between Appalachian/non-Appalachian early/late stage within the same population group. Patient data is an average from a minimum of two cores per patient. AA designates amino acid domain per each collagen. (B) Combined results for area under the receiver operating curve showing significant discriminatory difference in Appalachian stage I + II compared with stage III + IV CRC. (C) Lower discriminatory difference in non-Appalachian stage I + II compared with stage III + IV CRC. (D) Combined results for area under the receiver operating curve showing no discriminatory difference in Appalachian compared with non-Appalachian in stage I/II CRC. (E) Significant discriminatory differences when comparing between Appalachian and non-Appalachian stage III + IV CRC. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$. CRC, colorectal cancer; App., Appalachian; non-App., non-Appalachian; ns, not significant; Col, collagen; LN, natural log.

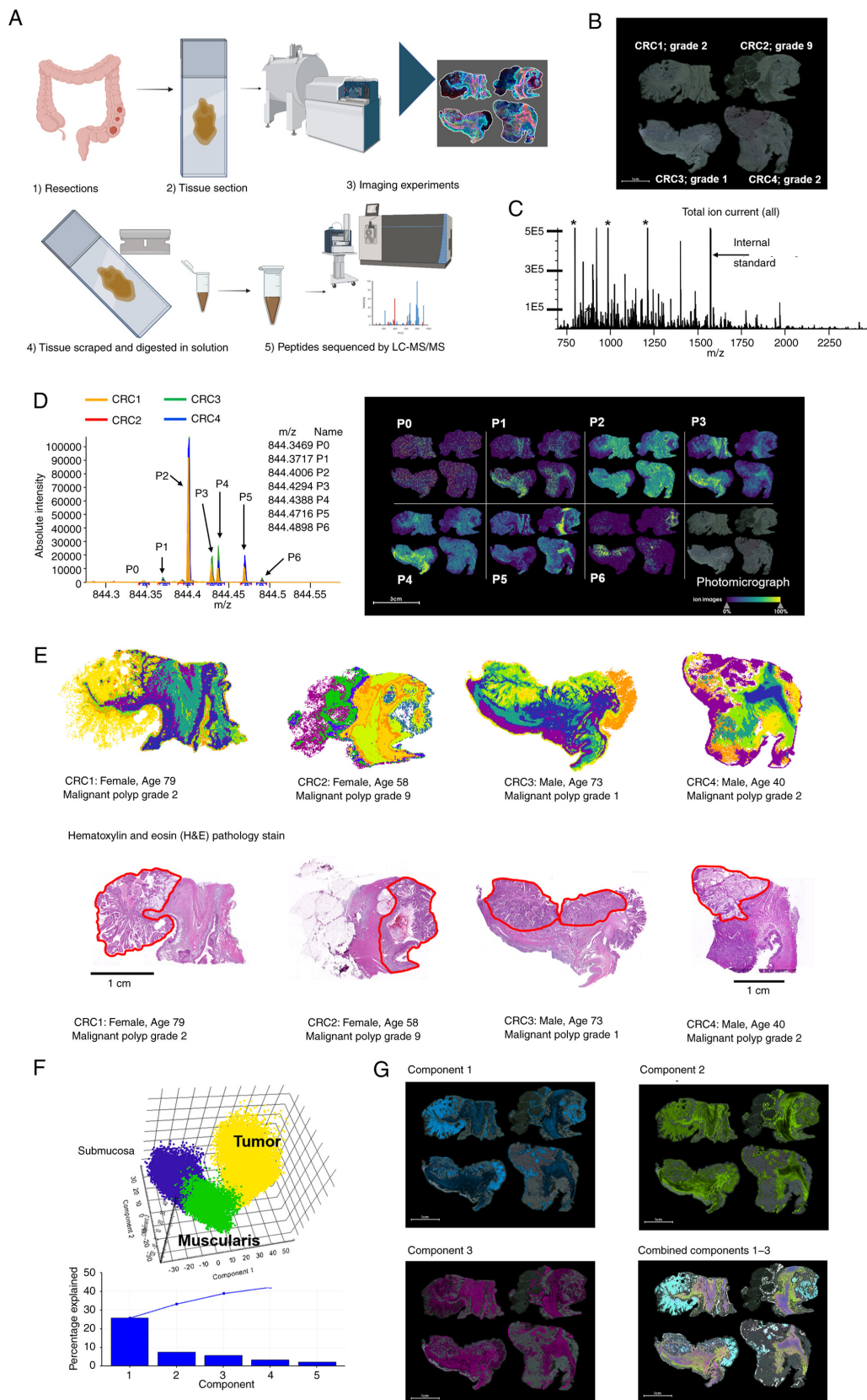


Figure 4. Complex peptide gradient patterns in CRC resections by collagen targeted mass spectrometry imaging. (A) Workflow took tissue sections from colorectal resections for experiments by high mass accuracy, high mass resolution imaging mass spectrometry. Image created using Biorender (biorender.com). (B) Photomicrograph of sections from four colorectal resections demonstrating different CRC features. (C) Total ion current over all four sections showed 4,696 peptides. (D) A high level of peptide complexity was found within a narrow mass window representing unique image patterns within the four sections (scale bar, 3 cm). (E) Heuristic peptide clustering of the 4,696 peak set by image segmentation shows high definition of pathologically defined regions found by hematoxylin and eosin pathology staining (scale bar, 1 cm). (F) Principal components analysis of all spectra (4,696 peak set) from pathological regions of tumor, mucosa and muscularis demonstrates separation based on pathological region. Component 1 represents 25.6% of variance derived from pathological location. (G) Spatial mapping of components 1 through 3 further confirmed distinct pathological regions defined by collagen peptide regulation (scale bar, 1 cm). CRC, colorectal cancer; LC-MS/MS, liquid chromatography-mass spectrometry.

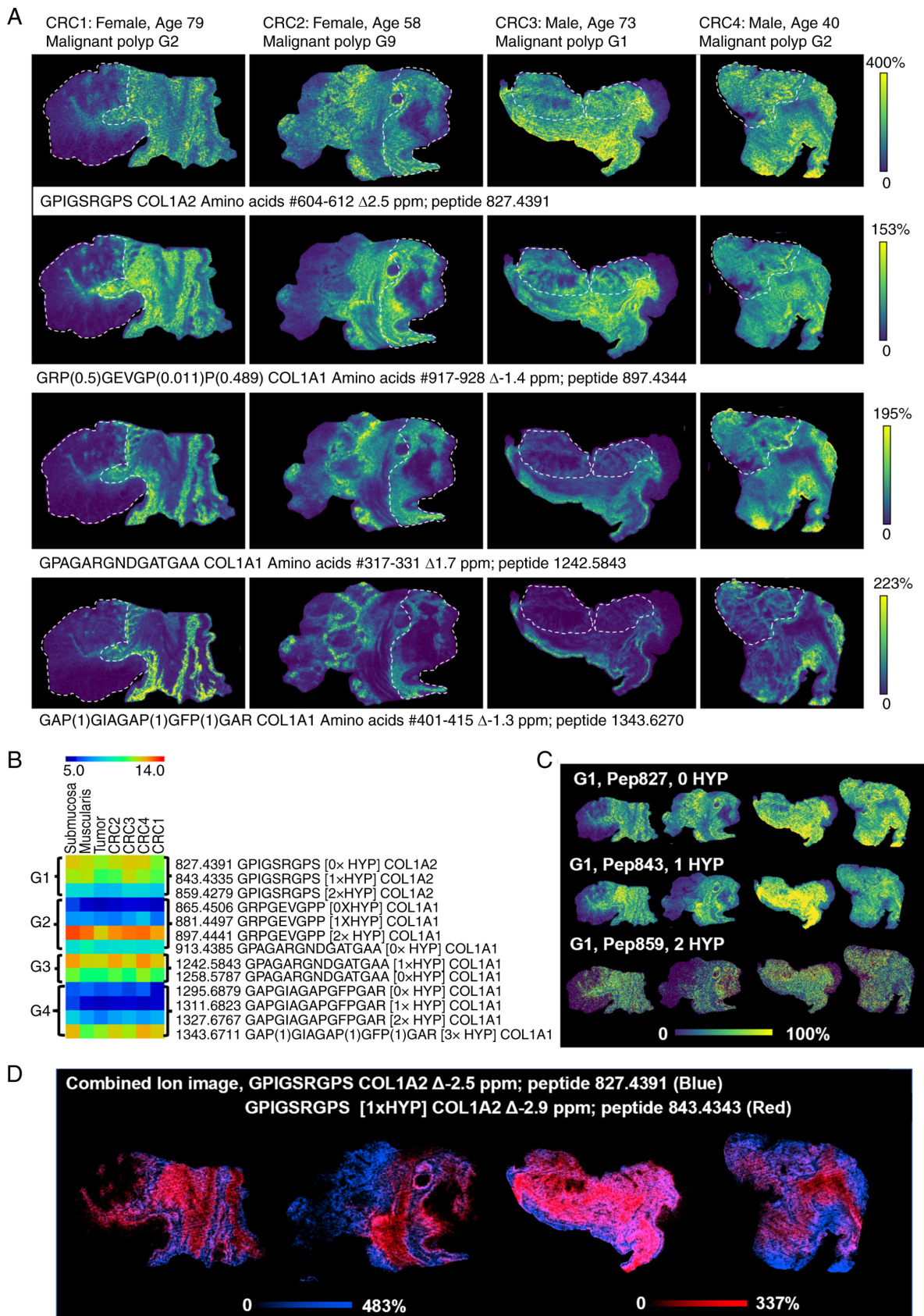


Figure 5. Variation in CRC pathology dependent on collagen peptide status of hydroxylated proline modification. (A) Case studies of example peptides identified by sequencing proteomics on the same tissue sections (G=grade). Each peptide reports unique spatial patterns following tissue pathology. Images are from high mass accuracy, high mass resolution imaging experiments by Fourier Transform Ion Cyclotron Resonance mass spectrometer. Parenthesis refers to site probability found by high mass accuracy, high mass resolution sequencing proteomics. (B) Peptide peak intensity varying by potential proline status in pathological regions or across each section. Peaks are defined by difference in hydroxylated proline status within each peptide. Window for peak selection from image data was $\pm$ 5 ppm. (C) Example mapping of peaks defined by potential differences in proline status defined by high mass accuracy. (D) Combined ion image showing complementary peak distribution of a Col peptide with unmodified status or hydroxylated proline status. CRC, colorectal cancer; Col, collagen; HYP, hydroxylated proline.

CRC (13,14). To our knowledge, this is the first study to assess ECM proteomic regulation in clinically archived tissues that includes potential collagen hydroxyproline changes within patient-matched NAT and colorectal tumors and is the first to consider a role in spatial distribution of collagen expression as a factor in the CRC tumor microenvironment.

Collagen and stromal extracellular matrix have been shown to have predictive values colorectal cancer (23,38,57,77). These studies have been done at the qualitative level with limited information on collagen protein structure changes that may influence cancer progression. In the present study we discovered five peaks that distinguish tumor compared to NAT with high sensitivity and specificity. Peptides belonging to fibrillar collagens collagen α -1(I) chain (COL1A1), collagen α -2(I) chain (COL1A2), and collagen α -1(III) chain (COL3A1) showed specific sites of hydroxylated proline (HYP) that were altered in NAT vs. tumor. Collagen HYP sites form hydrogen bonds to glycine residues that are the basis for the triple helical collagen structure in common with fibrillar collagens (75). Differences in sites of proline hydroxylation has been previously established to modulate collagen signaling within the tissue microenvironment by exposing or covering domains involved in cellular function and signaling (23,51,53). Prolyl hydroxylases, the enzymes that facilitate HYP modification alters tumor progression across many cancers (71). Previous proteomic studies report that prolyl 4-hydroxylase 1 (P4HA1) was increased relative to benign colon mucosa and was an independent prognosticator showing poor outcome for patients with early stage CRC (27). Furthermore, inhibition of P4HA1 decreased tumor cell growth and metastasis to liver and bone (78). Few studies have reported the dynamics of hydroxylated proline site changes due to the activity of prolyl 4-hydroxylases. A single proteomic study found a lower degree of collagen hydroxylation within the triple helical region when CRC metastasizes to liver (66). Collectively, the evidence supports that site specific modulation of the collagen protein structure alters with changes in malignancy.

A specific set of peptides was found to distinguish late-stage CRC in the Appalachian population compared to Non-Appalachian population. The Appalachian population has multiple factors that contribute to disparities: a high level of poverty and unemployment, lower levels of education, geographic isolation, and an inadequate health care infrastructure of healthcare resources and professionals (13,16,17). Along with these societal aspects, generational lifestyle behaviors also influence cancer progression. High levels of obesity, smoking, and a sedentary lifestyle are characteristic of the Appalachian population (13,14,16,17), and these characteristics correlate to high prevalence and severity of colorectal cancer (15,79), not only in this specific population, but in the general population as well. Our group and others have previously explored linking molecular factors to lifestyle influences as a contribution to disease (80-82). The present study, while done on a small cohort, provides initial evidence that the microenvironment of the extracellular matrix proteome (outside of the cells) is altered in a population with specific risk factors for colorectal cancer. The increased expression in late-stage colorectal cancer from the Appalachian population corresponds with previous reports that collagen deposition and pathways increase in CRC (20-23). Here, very specific peptides found increased above that of the

non-Appalachian populations, suggestive of a collagen structure-function regulation that facilitates a more aggressive cancer microenvironment (Figs. S6-S8). The peaks selected showed similar trends in expression patterns. This is expected as past studies using the method (49-51) show that a majority of peaks come from the same protein population consisting of collagens and extracellular matrix proteins that provide collagen fiber regulation. Proteomic modulations to the extracellular matrix microenvironment contribute to colorectal signaling in this high-risk population and may provide a new avenue to stratify patients, provide new therapeutic targets, and direct treatment from an earlier stage in diagnosis.

A new finding presented in this manuscript is the significant spatial regulation of collagen structure within the colorectal tumor microenvironment. In the tissue microarrays, cores were selected by a pathologist and show relative patterns from tumor tissue or normal adjacent tissue. Mass spectrometry imaging of the tissue from colorectal resections provided a view of the molecular gradients contributing to colorectal cancer. Surprisingly, while the resections were from diverse patients, the samples all show that collagen was relatively decreased in the primary tumor when compared to the adjacent pathologies or margins. Spatial clustering of multiple peptides and principal components analysis strengthened the primary structural features of tumor, submucosa and muscularis all show very distinct spatial regulation of the extracellular matrix proteomes. Sequencing data further highlighted that collagen peptides that altered by post-translational status aligned differentially with pathological features of colorectal resections. Previous literature has described that collagen fibers form interconnective 'highways' that influence tumor cell migration (83) and this may be why expression increased at the tumor boundaries. However, the molecular content of these fibers and the regulation of informational domains that participate in cell function and signaling has yet to be defined. This work suggests that across the colorectal tumor microenvironment, fibrillar collagens are discretely regulated at the post-translational level dependent on cell content and alter with primary colorectal structure (Figs. S6-S8). Further, these regulations may be exacerbated by environmental factors specific to the patient which can influence the patient's prognosis.

There are limitations to this study. patients self-identified as white and it is possible that the unique tri-racial European-white, Sub-Saharan, and Indigenous American ancestry of some Appalachians (84) may contribute to higher cancer risk and lead to development of more aggressive CRC. Genetically defined ancestry would aid in determining hereditary CRC traits. African-Americans also show significant disparities in progression and outcome of colorectal cancer (12). We hypothesize that the significant molecular differences by extracellular protein modulation discovered in this study may also contribute to CRC progression within African-American populations. Lower socioeconomic factors contribute to lower frequencies of physician visits, and this may be why the Appalachian population showed a later diagnosis age compared to non-Appalachians (11). We recognize that this study was conducted as a single-center retrospective study on a small cohort, which might have led to potential selection bias. Future investigations will expand patient number and patient diversity in order to fully

understand the influence of collagen proline hydroxylation in CRC pathology in underserved, high risk populations.

In conclusion, increased collagen deposition has been established as a hallmark of colorectal cancer progression. Past studies have reported on pathological collagen regulation by chemical staining and collagen fiber measurements (20-23). We show here that molecular profiles of the collagen protein structure change with malignancy and differentiate high risk populations. Additionally, we show that spatial regulation of collagen occurs across the colorectal tumor microenvironment, with post-translational regulation by hydroxylated proline differentiating the CRC histopathological region. The collagen protein structure has discrete domains that contribute to cell and protein binding, modulating cell function within the tumor microenvironment. These structural domains are modulated for cell use through alteration in hydroxylated proline status. Our collective data suggests that cell-matrix specific interactions contribute to structure-function changes within the CRC tumor microenvironment. Previous work by many groups has shown that collagen regulation is a predictive tool in clinical outcomes in many cancers, including CRC. Future work expands on the findings of this study to further define how collagen structure regulation associates with clinical outcomes in underserved populations. Studies linking cell signaling to spatial collagen regulation in the colorectal cancer microenvironment will be essential to understanding mechanisms and effects of collagen regulation in colorectal cancer progression.

Acknowledgements

The authors would like to thank Mr. William H. Angel (Consultant, Charleston, SC, USA) for their assistance in creating county maps.

Funding

This study was supported by the National Institutes of Health/National Cancer Institute (grant nos. R01CA253460, R21CA240148 and R01CA226086); the Biorepository and Tissue Analysis Shared Resource, Hollings Cancer Center, Medical University of South Carolina; the MUSC Digestive Disease Research Core (grant no. P30DK123704); the South Carolina Centers of Economic Excellence SmartState program; the Biospecimen Procurement and Translational Pathology Shared Resource Facility of the University of Kentucky Markey Cancer Center (grant no. P30CA177558). In addition, mass spectrometry data were collected in the University of Cincinnati Proteomics Laboratory under the direction of KDG on a Thermo Orbitrap Eclipse instrument purchased in part through an NIH instrumentation (grant no. 1S10OD026717). The MUSC Mass Spectrometry Facility and Redox Proteomics Core is supported by the Medical University of South Carolina and grant no. P20GM103542 (from NIH/NIGMS) with shared instrumentation S10OD010731, S10OD025126 and S10 OD030212 (from NIH/OD).

Availability of data and materials

The data generated in the present study may be found as a MassIVE dataset, a member of the Proteome Exchange

Consortium, under accession number (MSV000094873) or at the following URL: <https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000094873>.

Authors' contributions

CE performed the data collection from tissue microarrays. LC, RS, and CE completed histopathological staining and curation of de-identified clinical annotations. ATS performed initial data analysis with the assistance of PMA. SCZ and HBT performed data collection on colorectal tumor resections. RS, DA, RCS, KW and PMA assisted in the data analysis of resections. DA and EL performed histopathological analysis of TMAs and resections. RRD contributed to acquisition resources, participated in data interpretation, and reviewed versions of the manuscript. RCS and PMA supervised the research. ATS and PMA wrote the manuscript. ATS and PMA confirm the authenticity of all the raw data. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participation

All tissue samples were blinded to investigators and were delivered from the tissue bank at The University of Kentucky. No patient identifiers were disclosed in the completion of this study. The study was reviewed and approved by the University of Kentucky IRB with a waiver of informed patient consent (under IRB approval no. 63918, documented as 63918_401538 MALDI MSI_IRB). The Medical University of South Carolina IRB review for imaging mass spectrometry analysis waved the requirements for ethics approval and informed consent based on deidentified samples previously collected for banking and not for sole purposes of this research.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Wu X, Cai J, Zuo Z and Li J: Collagen facilitates the colorectal cancer stemness and metastasis through an integrin/PI3K/AKT/Snail signaling pathway. *Biomed Pharmacother* 114: 108708, 2019.
2. Nissen NI, Kehlet S, Boisen MK, Liljefors M, Jensen C, Johansen AZ, Johansen JS, Erler JT, Karsdal M, Mortensen JH, *et al*: Prognostic value of blood-based fibrosis biomarkers in patients with metastatic colorectal cancer receiving chemotherapy and bevacizumab. *Sci Rep* 11: 1-12, 2021.
3. Longley DB, Harkin DP and Johnston PG: 5-fluorouracil: Mechanisms of action and clinical strategies. *Nat Rev Cancer* 3: 330-338, 2003.
4. Iacovelli R, Pietrantonio F, Maggi C, de Braud F and Di Bartolomeo M: Combination or single-agent chemotherapy as adjuvant treatment of gastric cancer: A systematic review and meta-analysis of published trials. *Crit Rev Oncol Hematol* 98: 24-28, 2016.
5. Iacovelli R, Pietrantonio F, Palazzo A, Maggi C, Ricchini F, de Braud F and Di Bartolomeo M: Incidence and relative risk of grade 3 and 4 diarrhoea in patients treated with capecitabine or 5-fluorouracil: A meta-analysis of published trials. *Br J Clin Pharmacol* 78: 1228-1237, 2014.

6. Jang BS, Park HJ, Kim K, Jang JY, Kim SW, Oh DY and Chie EK: Role of adjuvant chemoradiotherapy for duodenal cancer: An updated analysis of long-term follow-up from single institution. *World J Surg* 42: 3294-3301, 2018.
7. Lee JJ, Beumer JH and Chu E: Therapeutic drug monitoring of 5-fluorouracil. *Cancer Chemother Pharmacol* 78: 447-464, 2016.
8. Sougiannis AT, VanderVeen BN, Enos RT, Velazquez KT, Bader JE, Carson M, Chatzistamou I, Walla M, Pena MM, Kubinak JL, *et al*: Impact of 5 fluorouracil chemotherapy on gut inflammation, functional parameters, and gut microbiota. *Brain Behav Immun* 80: 44-55, 2019.
9. Siegel R, DeSantis C and Jemal A: Colorectal cancer statistics, 2014. *CA Cancer J Clin* 64: 104-117, 2014.
10. Lengerich EJ, Wyatt SW, Rubio A, Beaulieu JE, Coyne CA, Fleisher L, Ward AJ and Brown PK: The appalachia cancer network: Cancer control research among a rural, medically underserved population. *J Rural Health* 20: 181-187, 2004.
11. Blackley D, Behringer B and Zheng S: Cancer mortality rates in Appalachia: Descriptive epidemiology and an approach to explaining differences in outcomes. *J Community Health* 37: 804-813, 2012.
12. Augustus GJ and Ellis NA: Colorectal cancer disparity in African Americans: Risk factors and carcinogenic mechanisms. *Am J Pathol* 188: 291-303, 2018.
13. Lengerich EJ, Tucker TC, Powell RK, Colsher P, Lehman E, Ward AJ, Siedlecki JC and Wyatt SW: Cancer incidence in Kentucky, Pennsylvania, and West Virginia: Disparities in appalachia. *J Rural Health* 21: 39-47, 2005.
14. Wingo PA, Tucker TC, Jamison PM, Martin H, McLaughlin C, Bayakly R, Bolick-Aldrich S, Colsher P, Indian R, Knight K, *et al*: Cancer in Appalachia, 2001-2003. *Cancer* 112: 181-192, 2008.
15. Doubeni CA, Major JM, Laiyemo AO, Schootman M, Zauber AG, Hollenbeck AR, Sinha R and Allison J: Contribution of behavioral risk factors and obesity to socioeconomic differences in colorectal cancer incidence. *J Natl Cancer Inst* 104: 1353-1362, 2012.
16. Fleming ST, Mackley HB, Camacho F, Seiber EE, Gusani NJ, Matthews SA, Liao J, Yang TC, Hwang W and Yao N: Clinical, sociodemographic, and service provider determinants of guideline concordant colorectal cancer care for Appalachian residents. *J Rural Health* 30: 27-39, 2014.
17. Paskett ED, Fisher JL, Lengerich EJ, Schoenberg NE, Kennedy SK, Conn ME, Roberto KA, Dwyer SK, Fickle D and Dignan M: Disparities in underserved white populations: The case of cancer-related disparities in Appalachia. *Oncologist* 16: 1072-1081, 2011.
18. Alexander DD, Waterbor J, Hughes T, Funkhouser E, Grizzle W and Manne U: African-American and Caucasian disparities in colorectal cancer mortality and survival by data source: An epidemiologic review. *Cancer Biomark* 3: 301-313, 2007.
19. Tapan U, Lee SY, Weinberg J, Kolachalama VB, Francis J, Charlot M, Hartshorn K and Chitalia V: Racial differences in colorectal cancer survival at a safety net hospital. *Cancer Epidemiol* 49: 30-37, 2017.
20. Le CC, Bennisroune A, Langlois B, Salesse S, Boulagnon-Rombi C, Morjani H, Dedieu S and Appert-Collin A: Functional interplay between collagen network and cell behavior within tumor microenvironment in colorectal cancer. *Front Oncol* 10: 527, 2020.
21. Zou X, Feng B, Dong T, Yan G, Tan B, Shen H, Huang A, Zhang X, Zhang M, Yang P, *et al*: Up-regulation of type I collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis. *J Proteomics* 94: 473-485, 2013.
22. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, van de Velde CJ and Watanabe T: Colorectal cancer. *Nat Rev Dis Primers* 1: 15065, 2015.
23. Shi R, Gao S, Zhang J, Xu J, Graham LM, Yang X and Li C: Collagen prolyl 4-hydroxylases modify tumor progression. *Acta Biochim Biophys Sin (Shanghai)* 53: 805-814, 2021.
24. Oh BY, Kim KH, Chung SS, Hong KS and Lee RA: Role of β 1-integrin in colorectal cancer: Case-control study. *Ann Coloproctol* 30: 61-70, 2014.
25. Said AH, Raufman JP and Xie G: The role of matrix metalloproteinases in colorectal cancer. *Cancers (Basel)* 6: 366-375, 2014.
26. Jeitany M, Leroy C, Tosti P, Lafitte M, Le Guet J, Simon V, Bonenfant D, Robert B, Grillet F, Mollevi C, *et al*: Inhibition of DDR 1-BCR signalling by nilotinib as a new therapeutic strategy for metastatic colorectal cancer. *EMBO Mol Med* 10: e7918, 2018.
27. Tanaka A, Zhou Y, Shia J, Ginty F, Ogawa M, Klimstra DS, Hendrickson RC, Wang JY and Roehrl MH: Prolyl 4-hydroxylase alpha 1 protein expression risk-stratifies early stage colorectal cancer. *Oncotarget* 11: 813-824, 2020.
28. Murray GI, Duncan ME, O'Neil P, Melvin WT and Fothergill JE: Matrix metalloproteinase-1 is associated with poor prognosis in colorectal cancer. *Nat Med* 2: 461-462, 1996.
29. Kirkland SC: Type I collagen inhibits differentiation and promotes a stem cell-like phenotype in human colorectal carcinoma cells. *Br J Cancer* 101: 320-326, 2009.
30. Miyazaki K, Togo S, Okamoto R, Idris A, Kumagai H and Miyagi Y: Collective cancer cell invasion in contact with fibroblasts through integrin- α 5 β 1/fibronectin interaction in collagen matrix. *Cancer Sci* 111: 4381-4392, 2020.
31. Karagiannis GS, Petraki C, Prassas I, Saraon P, Musrap N, Dimitromanolakis A and Diamandis EP: Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget* 3: 267-285, 2012.
32. Brauchle E, Kasper J, Daum R, Schierbaum N, Falch C, Kirschniak A, Schäffer TE and Schenke-Layland K: Biomechanical and biomolecular characterization of extracellular matrix structures in human colon carcinomas. *Matrix Biol* 68-69: 180-193, 2018.
33. Egeblad M, Rasch MG and Weaver VM: Dynamic interplay between the collagen scaffold and tumor evolution. *Curr Opin Cell Biol* 22: 697-706, 2010.
34. Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, Fong SF, Csiszar K, Giaccia A, Weninger W, *et al*: Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* 139: 891-906, 2009.
35. Solé X, Crous-Bou M, Cordero D, Olivares D, Guinó E, Sanz-Pamplona R, Rodriguez-Moranta F, Sanjuan X, de Oca J, Salazar R and Moreno V: Discovery and validation of new potential biomarkers for early detection of colon cancer. *PLoS One* 9: e106748, 2014.
36. Bonde A, Smith DA, Kikano E, Yoest JM, Tirumani SH and Ramaiya NH: Overview of serum and tissue markers in colorectal cancer: A primer for radiologists. *Abdom Radiol (NY)* 46: 5521-5535, 2021.
37. Lalmahomed ZS, Bröker ME, van Huizen NA, van den Braak RRJC, Dekker LJ, Rizopoulos D, Verhoef C, Steyerberg EW, Luider TM and Ijzermans JNM: Hydroxylated collagen peptide in urine as biomarker for detecting colorectal liver metastases. *Am J Cancer Res* 6: 321-330, 2016.
38. Kehlet SN, Sanz-Pamplona R, Brix S, Leeming DJ, Karsdal MA and Moreno V: Excessive collagen turnover products are released during colorectal cancer progression and elevated in serum from metastatic colorectal cancer patients. *Sci Rep* 6: 30599, 2016.
39. Qiu J, Jiang W, Yang Y, Feng C, Chen Z, Guan G, Zhuo S and Chen J: Monitoring changes of tumor microenvironment in colorectal submucosa using multiphoton microscopy. *Scanning* 37: 17-22, 2015.
40. Chen W, Dong S, Liu X, Wang G, Xu S, Lei S, Zhuo S and Yan J: Association of the collagen signature in the tumor microenvironment with recurrence and survival of patients with T4N0M0 colon cancer. *Dis Colon Rectum* 64: 563-575, 2021.
41. Zhuo S, Yan J, Chen G, Shi H, Zhu X, Lu J, Chen J and Xie S: Label-free imaging of basement membranes differentiates normal, precancerous, and cancerous colonic tissues by second-harmonic generation microscopy. *PLoS One* 7: e38655, 2012.
42. Liu N, Chen J, Xu R, Jiang S, Xu J and Chen R: Label-free imaging characteristics of colonic mucinous adenocarcinoma using multiphoton microscopy. *Scanning* 35: 277-282, 2013.
43. Xia G, Zhi W, Zou Y, Wang L, Wang C, Peng R and Hu X: Non-linear optical imaging and quantitative analysis of the pathological changes in normal and carcinomatous human colorectal muscularis. *Pathology* 49: 627-632, 2017.
44. Nallala J, Jaynes C, Saunders S, Smart N, Lloyd G, Riley L, Salmon D and Stone N: Characterization of colorectal mucus using infrared spectroscopy: A potential target for bowel cancer screening and diagnosis. *Lab Invest* 100: 1102-1110, 2020.
45. Brozek-Pluska B, Musial J, Kordek R and Abramczyk H: Analysis of human colon by raman spectroscopy and imaging-elucidation of biochemical changes in carcinogenesis. *Int J Mol Sci* 20: 3398, 2019.
46. Farrar CT, Gale EM, Kennan R, Ramsay I, Masia R, Arora G, Looby K, Wei L, Kalpathy-Cramer J, Bunzel MM, *et al*: CM-101: Type I collagen-targeted MR imaging probe for detection of liver fibrosis. *Radiology* 287: 581-589, 2018.

47. Caravan P, Das B, Dumas S, Epstein FH, Helm PA, Jacques V, Koerner S, Kolodziej A, Shen L, Sun WC and Zhang Z: Collagen-targeted MRI contrast agent for molecular imaging of fibrosis. *Angew Chem Int Ed Engl* 46: 8171-8173, 2007.
48. Sarri B, Canonge R, Audier X, Simon E, Wojak J, Caillol F, Cadot C, Marguet D, Poizat F, Giovannini M and Rigneault H: Fast stimulated Raman and second harmonic generation imaging for intraoperative gastro-intestinal cancer detection. *Sci Rep* 9: 10052, 2019.
49. Angel PM, Comte-Walters S, Ball LE, Talbot K, Mehta A, Brockbank KGM and Drake RR: Mapping extracellular matrix proteins in formalin-fixed, paraffin-embedded tissues by MALDI imaging mass spectrometry. *J Proteome Res* 17: 635-646, 2018.
50. Angel PM, Schwamborn K, Comte-Walters S, Clift C, Ball LE, Mehta AS and Drake RR: Extracellular matrix imaging of breast tissue pathologies by MALDI imaging mass spectrometry. *Proteomics Clin Appl* 13: e1700152, 2019.
51. Angel PM, Bruner E, Bethard J, Clift CL, Ball L, Drake RR and Feghali-Bostwick C: Extracellular matrix alterations in low-grade lung adenocarcinoma compared with normal lung tissue by imaging mass spectrometry. *J Mass Spectrom* 55: e4450, 2020.
52. Angel PM, Bruner E, Clift CL, Bethard JB, Ball LE, Drake RR and Feghali-Bostwick C: Extracellular matrix alterations in low grade lung adenocarcinoma compared to normal lung tissue by imaging mass spectrometry. *J Mass Spectrom* 55: e4450, 2019.
53. Angel PM, Spruill L, Jefferson M, Bethard JR, Ball LE, Hughes-Halbert C and Drake RR: Zonal regulation of collagen-type proteins and posttranslational modifications in prostatic benign and cancer tissues by imaging mass spectrometry. *Prostate* 80: 1071-1086, 2020.
54. Clift CL, Su YR, Bichell D, Smith HCJ, Bethard JR, Norris-Caneda K, Comte-Walters S, Ball LE, Hollingsworth MA, Mehta AS, *et al*: Collagen fiber regulation in human pediatric aortic valve development and disease. *Sci Rep* 11: 1-17, 2021.
55. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA and Quackenbush J: TM4 microarray software suite. *Methods Enzymol* 411: 134-193, 2006.
56. Rujchanarong D, Lefler J, Saunders JE, Pippin S, Spruill L, Bethard JR, Ball LE, Mehta AS, Drake RR, Ostrowski MC and Angel PM: Defining the tumor microenvironment by integration of immunohistochemistry and extracellular matrix targeted imaging mass spectrometry. *Cancers (Basel)* 13: 4419, 2021.
57. Angel PM and Zambrzycki SC: Predictive value of collagen in cancer. *Adv Cancer Res* 154: 15-45, 2022.
58. Compton CC, Byrd DR, Garcia-Aguilar J, Kurtzman SH, Olawaiye A and Washington MK: AJCC cancer staging atlas: A companion to the seventh editions of the AJCC cancer staging manual and handbook. Springer Science & Business Media, 2012.
59. Pai RK, Bettington M, Srivastava A and Rosty C: An update on the morphology and molecular pathology of serrated colorectal polyps and associated carcinomas. *Mod Pathol* 32: 1390-1415, 2019.
60. Gimon TI, Dykstra MA, Chezar K, Buie WD and MacLean A: Malignant colorectal polyp pathology: Are we getting sufficient information to make decisions? *Dis Colon Rectum* 63: 135-142, 2020.
61. Freeman HJ: Early stage colon cancer. *World J Gastroenterol* 19: 8468-8473, 2013.
62. Johncilla M and Yantiss RK: Histology of colorectal carcinoma: Proven and purported prognostic factors. *Surg Pathol Clin* 13: 503-520, 2020.
63. van Pelt GW, Sandberg TP, Morreau H, Gelderblom H, van Krieken JHJM, Tollenaar RAEM and Mesker WE: The tumour-stroma ratio in colon cancer: The biological role and its prognostic impact. *Histopathology* 73: 197-206, 2018.
64. Nyström H: Extracellular matrix proteins in metastases to the liver-composition, function and potential applications. *Semin Cancer Biol* 71: 134-142, 2021.
65. Winslow S, Lindquist KE, Edsjö A and Larsson C: The expression pattern of matrix-producing tumor stroma is of prognostic importance in breast cancer. *BMC Cancer* 16: 841, 2016.
66. van Huizen NA, Burgers PC, van Rosmalen J, Doukas M, IJzermans JNM and Luijckx TM: Down-regulation of collagen hydroxylation in colorectal liver metastasis. *Front Oncol* 10: 557737, 2020.
67. Gjaltema RA and Bank RA: Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. *Crit Rev Biochem Mol Biol* 52: 74-95, 2017.
68. Tanjore H and Kalluri R: The role of type IV collagen and basement membranes in cancer progression and metastasis. *Am J Pathol* 168: 715-717, 2006.
69. Ikeda K, Iyama KI, Ishikawa N, Egami H, Nakao M, Sado Y, Ninomiya Y and Baba H: Loss of expression of type IV collagen $\alpha 5$ and $\alpha 6$ chains in colorectal cancer associated with the hypermethylation of their promoter region. *Am J Pathol* 168: 856-865, 2006.
70. Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, *et al*: Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J Biol Chem* 283: 21187-21197, 2008.
71. Emsley J, Knight CG, Farndale RW, Barnes MJ and Liddington RC: Structural basis of collagen recognition by integrin $\alpha 2 \beta 1$. *Cell* 101: 47-56, 2000.
72. Zhao Y, Marcink TC, Gari RRS, Marsh BP, King GM, Stawikowska R, Fields GB and Van Doren SR: Transient collagen triple helix binding to a key metalloproteinase in invasion and development. *Structure* 23: 257-269, 2015.
73. Coelho NM, Wang A and McCulloch CA: Discoidin domain receptor 1 interactions with myosin motors contribute to collagen remodeling and tissue fibrosis. *Biochim Biophys Acta Mol Cell Res* 1866: 118510, 2019.
74. Ichikawa O, Osawa M, Nishida N, Goshima N, Nomura N and Shimada I: Structural basis of the collagen-binding mode of discoidin domain receptor 2. *EMBO J* 26: 4168-4176, 2007.
75. Shoulders MD and Raines RT: Collagen structure and stability. *Ann Rev Biochem* 78: 929-958, 2009.
76. Gorres KL and Raines RT: Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol* 45: 106-124, 2010.
77. Liang Y, Lv Z, Huang G, Qin J, Li H, Nong F and Wen B: Prognostic significance of abnormal matrix collagen remodeling in colorectal cancer based on histologic and bioinformatics analysis. *Oncol Rep* 44: 1671-1685, 2020.
78. Agarwal S, Behring M, Kim HG, Bajpai P, Chakravarthi BVSK, Gupta N, Elkholy A, Al Diffalha S, Varambally S and Manne U: Targeting P4HA1 with a small molecule inhibitor in a colorectal cancer PDX model. *Transl Oncol* 13: 100754, 2020.
79. Hannan LM, Jacobs EJ and Thun MJ: The association between cigarette smoking and risk of colorectal cancer in a large prospective cohort from the United States. *Cancer Epidemiol Biomarkers Prev* 18: 3362-3367, 2009.
80. Rujchanarong D, Scott D, Park Y, Brown S, Mehta AS, Drake R, Sandusky GE, Nakshatri H and Angel PM: Metabolic links to socioeconomic stresses uniquely affecting ancestry in normal breast tissue at risk for breast cancer. *Front Oncol* 12: 876651, 2022.
81. Karlamangla AS, Merkin SS, Crimmins EM and Seeman TE: Socioeconomic and ethnic disparities in cardiovascular risk in the United States, 2001-2006. *Ann Epidemiol* 20: 617-628, 2010.
82. Robinson O, Carter AR, Ala-Korpela M, Casas JP, Chaturvedi N, Engmann J, Howe LD, Hughes AD, Jarvelin MR, Kähönen M, *et al*: Metabolic profiles of socio-economic position: A multi-cohort analysis. *Int J Epidemiol* 50: 768-782, 2021.
83. Wyckoff JB, Wang Y, Lin EY, Li JF, Goswami S, Stanley ER, Segall JE, Pollard JW and Condeelis J: Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. *Cancer Res* 67: 2649-2656, 2007.
84. Estes RJ, Goins JH, Ferguson P and Crain JL: Melungeons, a multi-ethnic population. *J Genetic Genealogy* 7: 1-108, 2011.



Copyright © 2025 Sougiannis et al. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.