

## Mucins and associated O-glycans based immunoprofile for stratification of colorectal polyps: clinical implication for improved colon surveillance

Shiv Ram Krishn<sup>1</sup>, Sukhwinder Kaur<sup>1</sup>, Yuri M. Sheinin<sup>2</sup>, Lynette M. Smith<sup>3</sup>, Shailendra K. Gautam<sup>1</sup>, Asish Patel<sup>4</sup>, Maneesh Jain<sup>1</sup>, Vasthala Juvvigunta<sup>5</sup>, Priya Pai<sup>1</sup>, Audrey J. Lazenby<sup>2</sup>, Hemant K. Roy<sup>7</sup> and Surinder K. Batra<sup>1,2,6</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>2</sup> Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>3</sup> Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>4</sup> Department of Surgery, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>5</sup> Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>6</sup> Fred and Pamela Buffett Cancer Center, Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>7</sup> Department of Gastroenterology, Boston Medical Center, Boston, Massachusetts, USA

**Correspondence to:** Surinder K. Batra, **email:** sbatra@unmc.edu

Sukhwinder Kaur, **email:** skaur@unmc.edu

**Keywords:** mucins, hyperplastic polyps, sessile serrated adenoma/polyps, tubular adenomas, O-glycans

**Received:** August 12, 2016

**Accepted:** September 20, 2016

**Published:** September 29, 2016

### ABSTRACT

**Sessile serrated adenoma/polyps (SSA/P) are premalignant lesions of colorectal cancer that are difficult to distinguish histologically from hyperplastic polyps (HP) of minimal to no malignant potential. Specific markers for differentiating SSA/P from HP can aid clinicians for optimizing colon surveillance intervals. The present study investigates the potential of mucins and associated O-glycans to distinguish SSA/P from HP. Expression of colonic mucins (MUC1, MUC4, MUC17, MUC2, and MUC5AC) and O-glycans [Sialyl Lewis<sup>A</sup> (CA19-9) and Tn/Sialyl-Tn on MUC1] were analyzed in HP (n=33), SSA/P (n=39), and tubular adenoma (TA) (n=36) samples by immunohistochemistry. A significantly reduced expression of MUC4 (p=0.0066), elevated expression of MUC17 (p=0.0002), and MUC5AC (p<0.0001) was observed in SSA/P cases in comparison to HP cases. Interestingly, significantly higher number of SSA/P cases (p<0.0001) exhibited MUC5AC expression in the goblet cells as well as filled the crypt lumen compared to only goblet cells in majority of the HP cases. Improved diagnostic potential was revealed by multivariate logistic regression analysis where combinatorial panel of MUC5AC/MUC17 discriminated SSA/P from HP (SN/SP=85/82%). Finally, the decision tree model based marker panel (CA19-9/MUC17/MUC5AC) predicted HP, SSA/P and TA with SN/SP of 58%/95%, 79%/90% and 97%/83%, respectively. Overall, the mucin and associated O-glycan based panel defined in the present study could aid in discriminating SSA/P from HP to devise better colon surveillance strategies.**

### INTRODUCTION

Colorectal cancer (CRC) is a molecularly and histologically heterogeneous, biologically diverse public health problem with over one million new cases diagnosed worldwide every year [1]. The multistep

process of colorectal carcinogenesis is associated with the accumulation of genetic and epigenetic alterations in epithelial cells *via* distinct molecular pathways including the most common “adenoma-carcinoma” pathway and the recently identified “serrated polyp neoplasia” pathway [2-4]. Conventional adenomas [tubular adenoma (TA),

villous adenoma (VA), and tubulovillous adenoma (TVA)] exhibit characteristic cytological dysplasia and are widely regarded as premalignant lesions accounting for 65-70% of CRCs [5]. In contrast to this, the serrated neoplasia pathway is associated with serrated polyps and accounts for 15-30% of CRCs, and potentially the majority of the interval cancers [6, 7]. According to the latest World Health Organization criteria, the heterogeneous group of serrated lesions is pathologically classified as hyperplastic polyps (HP), sessile serrated adenoma/polyps (SSA/P) and traditional serrated adenomas (TSA) [8]. Traditionally, the small HP found mainly in the distal colon are considered to be benign and non-neoplastic [9], while the larger sessile serrated polyps found mainly in the proximal colon are associated with synchronous and metachronous advanced adenocarcinomas [10, 11]. Epidemiology studies of large cohorts suggest that a stepwise progression of dysplasia and carcinoma from SSA/P may take 10 to 15 years [12] however; there are case reports that suggest faster neoplastic progression [13] or progression taking place in as little as eight months [10, 14, 15].

The differential risk of HP and SSA/P for malignant progression mandates accurate discrimination of these premalignant lesions for improved screening and surveillance. However, due to high cost and low specificity, testing of genetic events like KRAS and BRAF mutations is impractical [16], while histologic evaluation is often compromised because of shared morphological features. As a result, a diagnostic “gray zone” exists between HP and SSA/P, with a wide interobserver and biomarker variability [17-19]. As current surveillance recommendations are dictated by the histologic type of the polyps, therefore, inaccurate classification of these premalignant precursors may lead to an inappropriate follow-up that may potentially contribute to interval carcinomas or unnecessary intervention [20]. Therefore, sensitive and specific markers that could effectively distinguish SSA/P from other polyp subtypes are urgently needed.

Secreted and transmembrane mucins are the most abundant components and major building blocks of the dynamic stratified gastrointestinal mucus layer including that of the colon [21-23]. Alterations in mucins and their associated carbohydrate structures play a role in various gastrointestinal disorders including inflammatory bowel diseases and CRC [24]. Recent analyses of a combinatorial panel of apomucin backbones and associated O-glycans by our group demonstrated their potential to discriminate premalignant and malignant lesions of the colon from normal colon, inflamed colon and benign HP [22]. Further, a recent study identified the transmembrane mucin MUC17 as a potential immunohistochemical marker for SSA/P [25]. In another study, Bartley *et al.* found the relatively high specificity (82%) but low sensitivity (54%) of MUC6 for SSA/P [26], while Renauld *et al.* observed MUC5AC hypomethylation as an early marker of polyps

with malignant potential [27]. However, most of these studies were conducted on single or few mucins, and did not consider alterations in mucin-associated O-glycans. Further, there is no study that investigated mucins and associated O-glycans as a combinatorial biomarker panel for the stratification of polyp subtypes. Considering the high clinical utility of the appropriate polyp classification, in this study, we evaluated the potential of colonic mucins (MUC1, MUC4, MUC17, MUC2, MUC5AC, and MUC6) and associated glycans (Tn/STn-MUC1 and CA19-9) for differentiating SSA/P from HP and TA.

## RESULTS

Our earlier findings indicated a differential expression of mucins and their associated O-glycans during the progression of CRC and we demonstrated their utility in differentiating benign lesions from precancerous (polyps with malignant potential-SSA/P and TA) and cancerous ones [22]. Owing to the lack of reliable markers for distinguishing SSA/P from other polyp subtypes, we studied the expression of transmembrane mucins (MUC1, MUC4, MUC17), secreted mucins (MUC2, MUC5AC, MUC6), and O-glycans (Tn/STn-MUC1 and CA19-9) to identify a panel of markers that could efficiently discriminate SSA/P from other polyp subtypes. The clinical and endoscopic findings of the patients and their polyp samples included in this study were retrieved from the electronic medical records. These are summarized in Table 1. There were no significant differences among polyps with regard to age, sex ratio, ethnicity, family history of colon cancer, tobacco consumption, polyp size, and prior or subsequent colonoscopies (Table 1).

### Association of mucins and associated O-glycans expression in colorectal polyps with clinicopathological features

As summarized in Table 2, expression of mucins and associated O-glycans was independent of age, gender, family history of colon cancer, and polyp size of the patient. Interestingly, expression of MUC5AC ( $p = 0.027$ ) and CA19-9 ( $p = 0.021$ ) was significantly higher in the polyps from Caucasian patients in comparison to those from other ethnicities. Of all mucins examined, the expression of MUC4 was significantly altered by polyp site ( $p = 0.037$ ) and tobacco consumption ( $p = 0.021$ ). Expression of MUC4 was higher in polyps resected from descending colon compared to all other sites and in polyps from patients with a history of tobacco use. Levels of MUC1 were significantly high ( $p = 0.034$ ) in cases having  $\geq 2$  polyps in number compared to  $< 2$  polyps. MUC2 levels were significantly high ( $p = 0.024$ ) in the polyps of individuals who had undergone  $\geq 2$  prior or subsequent colonoscopies.

**Table 1: Demographic and clinicopathological characteristics of patients included in the study**

		HP (n = 33)	SSA/P (n = 39)	TA (n = 36)	p-value
<b>Gender</b>	Female	17 (52%)	22 (56%)	18 (50%)	0.93
	Male	11 (33%)	15 (39%)	14 (39%)	
	Unknown	5 (15%)	2 (5%)	4 (11%)	
<b>Age</b>	Median (Range)	59 (50-91)	60 (28-77)	61 (47-84)	0.9
<b>Ethnicity</b>	African American	6 (18%)	2 (5%)	3 (8%)	0.19
	Caucasian	20 (61%)	30 (77%)	24 (67%)	
	Hispanic	0 (0%)	0 (0%)	1 (3%)	
	Asian	0 (0%)	1 (3%)	0 (0%)	
	Unknown	7 (21%)	6 (15%)	8 (22%)	
<b>Polyp site</b>	Cecum	0 (0%)	4 (10%)	4 (11%)	<b>0.025</b>
	Ascending	2 (6%)	10 (26%)	10 (28%)	
	Transverse	1 (3%)	6 (15%)	6 (17%)	
	Descending	4 (12%)	5 (13%)	3 (8%)	
	Sigmoid	10 (30%)	7 (18%)	4 (11%)	
	Rectum	9 (27%)	5 (13%)	5 (14%)	
	Unknown	7 (21%)	2 (5%)	4 (11%)	
<b>Family History of colon cancer</b>	No	18 (55%)	29 (74%)	27 (75%)	0.34
	Yes	7 (21%)	6 (15%)	4 (11%)	
	Unknown	8 (24%)	4 (10%)	5 (14%)	
<b>Tobacco</b>	No	6 (18%)	13 (33%)	10 (28%)	0.5
	Yes	20 (61%)	22 (57%)	21 (58%)	
	Unknown	7 (21%)	4 (10%)	5 (14%)	
<b>Polyp size (cm<sup>3</sup>)</b>	Mean (SD)	0.048 (0.084)	0.101 (0.168)	0.088 (0.174)	0.41
<b>Number of polyps</b>	Median (range)	2 (1-7)	2 (1-7)	3 (1-7)	<b>0.018</b>
<b>Prior or subsequent colonoscopies with polyps</b>	Median (range)	1 (1-3)	3 (0-5)	1 (0-3)	0.28

HP, hyperplastic polyp; SSA/Ps, sessile serrated adenoma/polyp; TA, tubular adenoma

### Expression of mucins and associated O-glycans in colorectal polyps and adjacent normal colon mucosa

The expression of MUC1, MUC5AC, Tn/STn-MUC1, and CA19-9 were significantly increased in all types of polyp lesions ( $p < 0.0005$ ) compared to adjacent normal colon (NC) (Table 3). Expression of MUC17 was significantly high in HP ( $p < 0.0005$ ) and SSA/P ( $p < 0.0005$ ) with statistically non-significant alterations in TA compared to NC (Table 3). Transmembrane mucin MUC4 expression was significantly decreased in SSA/P ( $p < 0.0005$ ) with statistically non-significant alterations in HP and TA compared to NC (Table 3). In comparison to NC, the expression of secretory mucin MUC2 was significantly decreased in both SSA/P ( $p < 0.005$ ) and TA ( $p < 0.0005$ ) and non-significantly altered in HP (Table 3).

### Differential expression of mucins and associated O-glycans among colorectal polyp subtypes

Looking at the aberrant expression of mucins and associated O-glycans in different colorectal polyps compared to NC, we next investigated whether mucins and associated O-glycans are differentially expressed between benign (HP) and pre-cancerous lesions of the colon (SSA/P and TA).

#### SSA/P vs. HP

In comparison to HP, SSA/P exhibited reduced expression of MUC4 ( $p = 0.0066$ ), elevated expression of MUC17 ( $p = 0.0002$ ) and MUC5AC ( $p < 0.0001$ ) (Figure 1A, 1B; Table 3). Interestingly, in SSA/P, the strong reactivity of MUC4 (H-Score  $> 2$ ) was restricted to 10% of cases in comparison to the 45% of HP cases. While moderate to the strong reactivity (H-Score  $> 1$ ) of MUC17, MUC5AC was observed in 33% and 62% of SSA/P cases, respectively, compared to only 6% and 15% of HP cases respectively.

**Table 2: Association of mucins and associated O-glycans expression with clinicopathological characteristics**

		N	MUC1	MUC4	MUC17	MUC2	MUC5AC	CA19-9	Tn/STn-MUC1
			Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE
Age	< 60 years	47	0.77 ± 0.09	1.31 ± 0.11	0.5 ± 0.07	1.71 ± 0.11	0.88 ± 0.09	1.21 ± 0.15	1.34 ± 0.07
	≥60 years	50	0.72 ± 0.08	1.27 ± 0.11	0.64 ± 0.06	1.8 ± 0.1	0.68 ± 0.09	0.98 ± 0.14	1.2 ± 0.07
	p-value		0.76	0.79	0.093	0.54	0.1	0.25	0.37
Gender	Female	57	0.74 ± 0.07	1.29 ± 0.1	0.57 ± 0.06	1.7 ± 0.1	0.71 ± 0.08	1.12 ± 0.14	1.28 ± 0.06
	Male	40	0.75 ± 0.1	1.29 ± 0.12	0.58 ± 0.08	1.84 ± 0.12	0.87 ± 0.11	1.04 ± 0.15	1.25 ± 0.09
	p-value		0.67	0.96	0.68	0.25	0.32	0.41	0.61
Ethnicity	Caucasian	74	0.73 ± 0.06	1.31 ± 0.09	0.56 ± 0.05	1.73 ± 0.09	0.83 ± 0.08	1.18 ± 0.12	1.22 ± 0.06
	Other	13	0.87 ± 0.18	1.31 ± .21	0.48 ± 0.12	1.85 ± 0.17	0.4 ± 0.12	0.45 ± 0.21	1.45 ± 0.15
	p-value		0.51	0.95	0.58	0.79	<b>0.027</b>	<b>0.021</b>	0.15
Polyp site	Cecum	8	0.65 ± 0.1	1.03 ± .14	0.75 ± 0.1	1.6 ± 0.16	0.75 ± 0.15	1.19 ± 0.24	1.15 ± 0.1
	Ascending	22	0.73 ± 0.12	1.28 ± .22	0.56 ± 0.19	1.96 ± 0.12	0.93 ± 0.24	1.01 ± 0.28	1.28 ± 0.15
	Transverse	13	0.5 ± 0.15	1.11 ± 0.27	0.55 ± 0.16	1.43 ± 0.32	0.54 ± 0.25	0.84 ± 0.36	1.05 ± 0.14
	Descending	12	0.68 ± 0.17	1.73 ± 0.19	0.55 ± 0.1	1.57 ± 0.19	0.59 ± 0.09	0.95 ± 0.21	1.35 ± 0.13
	Sigmoid	21	0.74 ± 0.12	1.42 ± 0.16	0.43 ± 0.09	1.98 ± 0.15	0.79 ± 0.11	1.17 ± 0.22	1.3 ± 0.11
	Rectum	19	1.02 ± 0.15	0.93 ± 0.2	0.59 ± 0.11	1.88 ± 0.19	1.06 ± 0.24	1.03 ± 0.28	1.39 ± 0.13
	p-value		0.16	<b>0.037</b>	0.33	0.33	0.47	0.93	0.42
FH of colon cancer	no	74	0.77 ± 0.07	1.27 ± 0.09	0.56 ± 0.05	1.72 ± 0.09	0.77 ± 0.08	0.97 ± 0.11	1.28 ± 0.06
	yes	17	0.62 ± 0.09	1.38 ± 0.14	0.49 ± 0.1	1.92 ± 0.11	0.76 ± 0.18	1.44 ± 0.25	1.13 ± 0.1
	p-value		0.58	0.66	0.64	0.65	0.9	0.064	0.15
Tobacco	no	29	0.63 ± 0.09	1.02 ± 0.13	0.51 ± 0.07	1.65 ± 0.15	0.89 ± 0.14	1.1 ± 0.18	1.28 ± 0.1
	yes	63	0.79 ± 0.08	1.43 ± 0.1	0.56 ± 0.06	1.8 ± 0.09	0.7 ± 0.08	1.02 ± 0.13	1.24 ± 0.06
	p-value		0.3	<b>0.021</b>	0.78	0.52	0.27	0.51	0.84
Polyp size	<0.05 cm <sup>3</sup>	66	0.7 ± 0.07	1.3 ± 0.09	0.53 ± 0.06	1.78 ± 0.08	0.72 ± 0.08	0.98 ± 0.12	1.25 ± 0.07
	≥0.05 cm <sup>3</sup>	24	0.83 ± 0.14	1.31 ± 0.16	0.68 ± 0.09	1.75 ± 0.17	0.98 ± 0.13	1.29 ± 0.21	1.32 ± 0.08
	p-value		0.51	0.9	0.11	0.84	0.087	0.17	0.27
Number of polyps	<2	21	0.58 ± 0.14	1.3 ± 0.17	0.62 ± 0.11	1.91 ± 0.12	0.97 ± 0.17	1.28 ± 0.22	1.33 ± 0.13
	≥2	75	0.79 ± 0.06	1.29 ± 0.09	0.57 ± 0.05	1.71 ± 0.09	0.72 ± 0.07	1.02 ± 0.12	1.25 ± 0.06
	p-value		<b>0.034</b>	0.99	0.76	0.44	0.2	0.35	0.41
Prior or subsequent colonoscopies with polyps	<2	45	0.7 ± 0.08	1.05 ± 0.13	0.6 ± 0.08	1.48 ± 0.11	0.71 ± 0.1	0.99 ± 0.15	1.19 ± 0.07
	≥2	21	0.69 ± 0.11	1.25 ± 0.14	0.6 ± 0.1	1.91 ± 0.14	0.84 ± 0.15	0.95 ± 0.2	1.35 ± 0.11
	p-value		0.94	0.31	0.78	<b>0.024</b>	0.48	0.83	0.25

FH, family history; SE, standard error

### SSA/P vs. TA

TA exhibited elevated expression of MUC1 ( $p = 0.0005$ ) and decreased expression of MUC17 ( $p = 0.0054$ ), MUC2 ( $p = 0.02$ ), MUC5AC ( $p < 0.0001$ ), and CA19-9 ( $p < 0.0001$ ) in comparison to SSA/P (Figure 1A, 1B; Table 3). Moderate to strong reactivity (H-Score > 1) of MUC1 was observed in 47% of TA compared to 13% of SSA/P. Further, moderate to strong reactivity (H-Score > 1) of MUC17, MUC2, MUC5AC, and CA19-9 was observed in a greater proportion of SSA/P cases (33%, 100%, 62%, and 82% respectively) compared to the TA cases (11%, 58%, 8%, and 0% respectively).

### Alteration in mucins and associated O-glycans localization

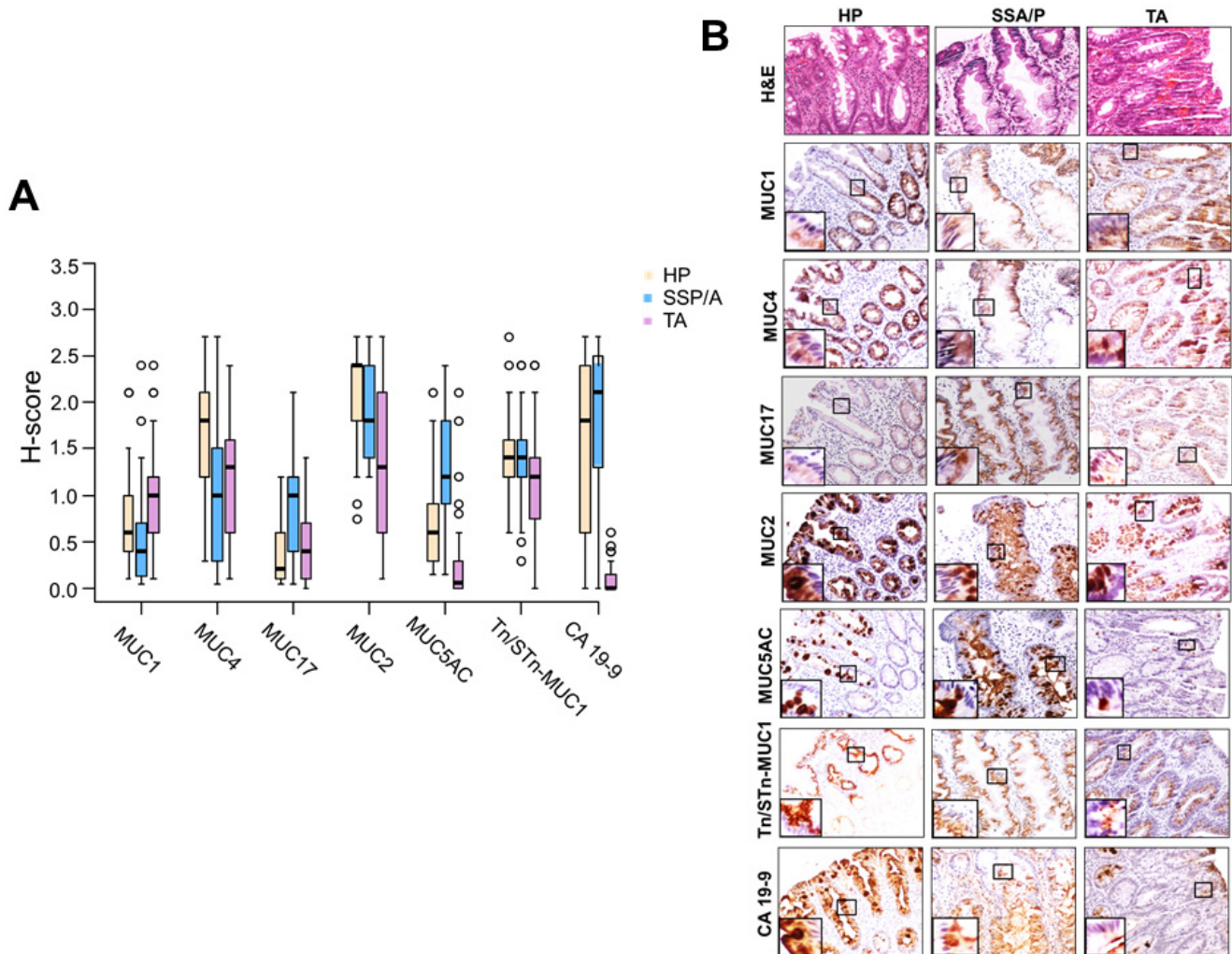
Interestingly, mucins and associated O-glycans exhibited a distinct expression pattern in HP, SSA/P, TA lesions in comparison to adjacent normal colon. In both adjacent normal colon and polyp subtypes, cytoplasmic expression of MUC17 and Tn/STn-MUC1 was observed in crypt epithelial cells. CA19-9 expression was absent in adjacent normal colon while in the polyp lesions its expression was observed to be in goblet cells and cytoplasm of crypt epithelial cells. In adjacent normal colon, MUC2 expression was observed solely in the goblet cells, while in the polyp lesions, it was expressed in both the goblet cells as well as the in cytoplasm of crypt epithelial cells. MUC1 was uniformly expressed in the



cytoplasm of the crypt epithelial cells in adjacent normal colon as well as in polyp lesions.

While secretory mucin MUC5AC expression was undetectable in the adjacent normal colon, it was overexpressed in all HP and SSA/P cases. Notably, a

stark difference in the MUC5AC expression pattern was observed within the colon crypts among the study groups. In HP, MUC5AC expression was limited to goblet cells in 61% cases, while the remaining 39% exhibited overexpression in the goblet cells as well as



**Figure 1: Immunohistochemical expression of mucins and associated O-glycans in colorectal polyps.** **A.** Box plots representing the H-score for mucins (MUC1, MUC4, MUC17, MUC2, and MUC5AC) and O-glycans (CA19-9, Tn/STn-MUC1) expression in HP, SSA/P and TA. A significant downregulation of MUC4 ( $p = 0.0066$ ), significant upregulation of MUC17 ( $p = 0.0002$ ) and MUC5AC ( $p < 0.0001$ ) was observed in SSA/P cases in comparison to HP cases. The box plots further show that compared to SSA/P in TA cases, MUC1 expression was significantly increased ( $p = 0.0005$ ) while there was a significant decrease in expression of MUC17 ( $p = 0.0054$ ), MUC2 ( $p = 0.02$ ), MUC5AC ( $p < 0.0001$ ) and CA19-9 ( $p < 0.0001$ ). **B.** The HP, SSA/P, and TA tissue specimens were probed with MUC1 (mouse monoclonal-HMFG2), MUC4 (mouse monoclonal-8G7), MUC17 (rabbit polyclonal-SN-1139-2), MUC2 (rabbit monoclonal-EPR6145), MUC5AC (mouse monoclonal-45M1), CA19-9 (mouse monoclonal-NS 19.9), and Tn/STn-MUC1 (mouse monoclonal-5E5) antibodies after non-specific blocking with horse serum. All the tissue sections were examined under a microscope, and the expression was evaluated on basis of reddish brown staining by a pathologist. Representative photomicrographs (20X) of Hematoxylin & Eosin staining, mucins (MUC1, MUC4, MUC17, MUC2, and MUC5AC), and O-glycans (Tn/STn-MUC1 and CA19-9) stained tissues of HP, SSA/Ps, and TA were taken. For uniformity, images for all the mucins and O-glycans staining were taken from the same area in a sample of HP, SSA/P and TA respectively. Magnified images in inset shows expression of respective mucins in different cellular compartments. MUC1 expression was significantly high ( $p = 0.0005$ ) in TA in comparison to HP and SSA/P. Strong expression of MUC4 was observed in adjacent normal colon. Expression of MUC4 was significantly reduced in SSA/P ( $p = 0.0066$ ) compared with HP. Significantly high immunoreactivity of MUC17 was observed in SSA/P lesions compared to HP and TA. Compared to HP and SSA/P, MUC2 expression was significantly reduced in TA ( $p = 0.0018$ , and  $0.020$  respectively). MUC5AC expression was significantly high in HP and SSA/P compared to TA ( $p < 0.0001$ ). HP, SSA/P, and TA expressed Tn/STn-MUC1 at similar levels. CA19-9 expression was significantly high in HP ( $p < 0.0001$ ) and SSA/P ( $p < 0.0001$ ) compared to TA. NC: adjacent normal colon; HP: hyperplastic polyps; SSA/P: sessile serrated adenoma/polyps; TA: tubular adenoma.

hypersecretion in lumen of the colon crypts. In contrast, in 89% SSA/P cases ( $p < 0.0001$ ) significantly increased MUC5AC expression was observed in both goblet cells and secreted in the lumen of colon crypts (Figure 2) while in the remaining 11% it was restricted to goblet cells only. Unlike MUC5AC, the normal colon expressed MUC4, which was localized in cytoplasmic compartments throughout the crypt epithelial cells but absent from luminal surface epithelium. In HP and TA, expression of MUC4 was observed throughout the crypt as well as on luminal surface epithelium. Interestingly, in approximately 30% SSA/P cases, MUC4 expression was limited to lower one-third of the crypts only, while in the remaining 70%, its expression was similar to HP and TA (Figure 2).

### Mucins and associated O-glycans as potential molecular marker(s) for differentiating colon polyp subtypes

Considering the significant differences in the levels and expression patterns of several mucins and associated O-glycans, we next evaluated their individual or combinatorial utility to discriminate the polyp-subtypes. To investigate this, we utilized the univariate logistic regression analysis along with receiver-operator characteristics (ROC) curves analysis.

#### SSA/P vs. HP

Low expression of MUC4 and increased expression of MUC17 and MUC5AC significantly differentiated

SSA/P polyps from HP with the sensitivity and specificity (SN/SP) of 61.5%/70%, 67%/91%, and 77%/73% respectively (Supplementary Figure 1A; Table 4).

#### SSA/P vs. TA

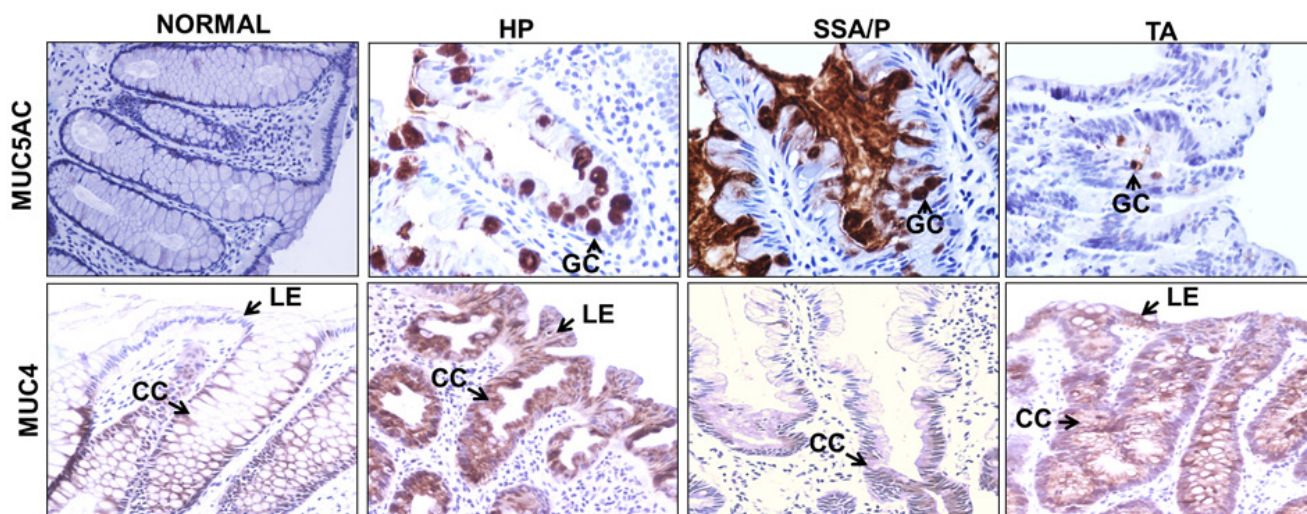
Low expression of MUC1, higher levels of MUC17, MUC2, MUC5AC, Tn/STn-MUC1, and CA19-9 significantly differentiated SSA/P from TA with the sensitivity and specificity (SN/SP) of 74.4%/72.2%, 67%/78%, 72%/58.3%, 92.3%/81%, 79.5%/44.4%, and 87.2%/100% respectively (Supplementary Figure 1B; Table 4).

### Combinatorial panels of mucins and associated O-glycans for differentiating colorectal polyp subtypes

Looking at some of the individual markers that exhibited better sensitivity while others exhibited better specificity to differentiate colorectal polyp subtypes, we evaluated the combined utility of significantly altered mucins and O-glycans as markers for differential identification of polyp subtypes. For this, multivariate regression models were created using the most promising markers from the univariate logistic regression analysis, in conjunction with ROC curve based analysis.

#### SSA/P vs. HP

The combination of MUC17/MUC5AC effectively discriminated SSA/P from HP with SN/SP of 85%/82%



**Figure 2: Differential localization of mucins in colorectal polyp subtypes.** Differential localization pattern for MUC5AC and MUC4 were observed for polyp subtypes. MUC5AC expression was absent in normal colon, restricted to the goblet cells (GC) in HP, and present in the goblet cells and lumen of colon crypts in SSA/P. In case of TA, the MUC5AC expression was observed in goblet cells but at lower intensity. In normal colon, MUC4 expression was observed throughout the colon crypts (CC) while it was absent on luminal surface epithelium (LE). In the HP and TA, MUC4 expression was observed throughout CC including luminal epithelium. In contrast to this, in 30% SSA/P cases, MUC4 expression was limited to lower one-third of colon crypts. The differential localization pattern of MUC4 and MUC5AC could provide useful molecular adjunct to histological classification of colorectal polyps. CC: colon crypts; GC: goblet cells; LE: luminal surface epithelium; HP: hyperplastic polyps; SSA/P: sessile serrated adenoma/polyps; TA: tubular adenoma.



**Table 3: Expression of mucins and associated O-glycans in different colorectal polyp subtypes and corresponding adjacent normal colon**

		NC	HP	NC	SSA/P	NC	TA
MUC1	Mean ± SE	0.25±0.07 <sup>e</sup>	0.76 ± 0.09 <sup>b</sup>	0.09±0.02 <sup>l</sup>	0.53 ± 0.09 <sup>c,h</sup>	0.28±0.05 <sup>i</sup>	1.02 ± 0.09 <sup>c,k</sup>
	Median (Min, Max)	0.1(0,1.4)	0.6(0.1, 2.1)	0.05(0, 0.3)	0.4(0.05, 2.4)	0.2(0, 1)	1(0.1, 2.4)
MUC4	Mean ± SE	1.87±0.11	1.66 ± 0.12 <sup>k</sup>	1.63±0.15 <sup>l</sup>	1.05 ± 0.12 <sup>c,e</sup>	1.25±0.12	1.21 ± 0.12
	Median (Min, Max)	2.1(0.2, 2.55)	1.8(0.3, 2.7)	1.8(0.1, 2.4)	1(0.05, 2.7)	1.2(0.05, 2.4)	1.3(0.1, 2.4)
MUC17	Mean ± SE	0.11±0.03 <sup>f</sup>	0.37 ± 0.05 <sup>c,l</sup>	0.06±0.02 <sup>l</sup>	0.85 ± 0.08 <sup>c,f,g</sup>	0.24±0.05	0.47 ± 0.07 <sup>j</sup>
	Median (Min, Max)	0.05(0, 0.7)	0.2(0.05, 1.2)	0(0, 0.3)	1(0.05, 2.1)	0.1(0, 1.2)	0.4(0, 1.4)
MUC2	Mean ± SE	2.31±0.09	2.09 ± 0.09 <sup>h</sup>	2.26±0.09 <sup>k</sup>	1.92 ± 0.08 <sup>b,g</sup>	2.21±0.06 <sup>i</sup>	1.35 ± 0.14 <sup>c,e,j</sup>
	Median (Min, Max)	2.4(1.2, 2.7)	2.4(0.75, 2.7)	2.4(0.9, 2.7)	1.8(1.2, 2.7)	2.4(1, 2.7)	1.3(0.1, 2.7)
MUC5AC	Mean ± SE	0±0 <sup>f</sup>	0.65 ± 0.08 <sup>c,l,i</sup>	0±0 <sup>l</sup>	1.24 ± 0.09 <sup>c,f,i</sup>	0.03±0.02 <sup>i</sup>	0.3 ± 0.08 <sup>c,f,l</sup>
	Median (Min, Max)	0(0, 0)	0.6(0.15, 2.1)	0(0, 0)	1.2(0.15, 2.4)	0(0, 0.6)	0.06(0, 2.1)
Tn/STn-MUC1	Mean ± SE	0.23±0.05 <sup>f</sup>	1.42 ± 0.07 <sup>e</sup>	0.17±0.05 <sup>l</sup>	1.39 ± 0.08 <sup>e</sup>	0.39±0.05 <sup>i</sup>	1.13 ± 0.1 <sup>e</sup>
	Median (Min, Max)	0.1(0, 0.8)	1.4(0.6, 2.7)	0(0, 0.8)	1.4(0.3, 2.4)	0.3(0, 1.6)	1.2(0, 2.4)
CA19-9	Mean ± SE	0.1±0.05 <sup>f</sup>	1.53 ± 0.17 <sup>c,i</sup>	0.04±0.02 <sup>l</sup>	1.72 ± 0.13 <sup>c,i</sup>	0.02±0.01 <sup>i</sup>	0.1 ± 0.03 <sup>c,f,l</sup>
	Median (Min, Max)	0(0, 1.2)	1.8(0, 2.7)	0(0, 0.4)	2.1(0, 2.7)	0(0, 0.3)	0(0, 0.6)

<sup>a</sup>*p* < 0.05 vs. adjacent normal; <sup>b</sup>*p* < 0.005 vs. adjacent normal; <sup>c</sup>*p* < 0.0005 vs. adjacent normal; <sup>d</sup>*p* < 0.05 vs. HP; <sup>e</sup>*p* < 0.005 vs. HP; <sup>f</sup>*p* < 0.0005 vs. HP; <sup>g</sup>*p* < 0.05 vs. TA; <sup>h</sup>*p* < 0.005 vs. TA; <sup>i</sup>*p* < 0.0005 vs. TA; <sup>j</sup>*p* < 0.05 vs. SSA/P; <sup>k</sup>*p* < 0.005 vs. SSA/P; <sup>l</sup>*p* < 0.0005 vs. SSA/P

NC, adjacent normal colon; HP, hyperplastic polyps; SSA/P, sessile serrated adenoma/polyps; TA, tubular adenoma; SE, standard error; Min, minimum; Max, maximum

(Figure 3A; Table 4). The optimal cut point value based on the H-Score derived from the model [ $P(\text{SSA/P}) \geq 0.49$ ] suggests that, if the predicted probability from the combined MUC17/MUC5AC model is  $\geq 0.49$  then the model predicts SSA/P, and if it is  $< 0.49$  then model predicts HP. The model has 83% accuracy (60/72) in predicting SSA/P vs. HP.

#### SSA/P vs. TA

The combination of MUC1/MUC17/CA19-9 effectively discriminated SSA/P from TA with SN/SP of 92.3%/97.2% (Figure 3B; Table 4). Further, optimal cutpoint [ $P(\text{SSA/P}) \geq 0.385$ ] derived from the model suggests that if the predicted probability from the combined MUC1/MUC17/CA19-9 model is  $\geq 0.385$  then the model predicts SSA/P and if it is  $< 0.385$  then model predicts TA. The model has 95% accuracy (71/75) in predicting SSA/P vs. TA.

#### CA19-9/MUC17/MUC5AC: a novel biomarker combination for differentiating polyp subtypes

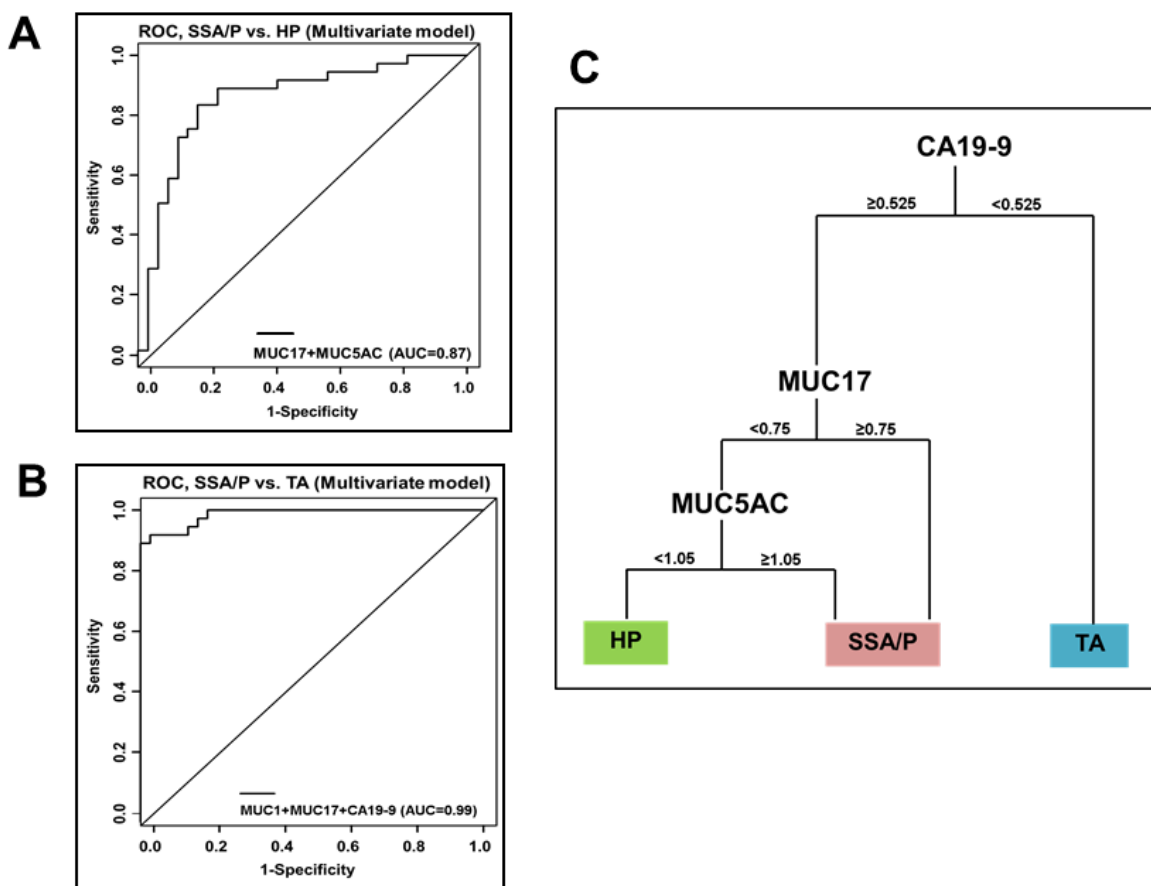
Next, we evaluated clinical utility of the significantly varying markers obtained from multivariate logistic regression results utilizing the tree based modeling to predict the diagnostic power for differentiating various lesions (HP, SSA/P, or TA) (Figure 3C). The combined utility of three markers (CA19-9, MUC17 and MUC5AC in order of preference) was identified for differential diagnosis of precancerous polyps (polyps with malignant potential-SSA/P and TA) from benign polyps (HP) (Figure 3C). According to this model, if the observed CA19-9 H-Score is  $< 0.525$ , the polyp under observation is predicted to be TA. If CA19-9 H-Score is  $\geq 0.525$  then the next marker considered is MUC17 (Figure 3C). If MUC17 H-Score is  $\geq 0.75$ , lesion is likely to be SSA/P. However, if MUC17 H-Score is  $< 0.75$ , then next marker considered is MUC5AC. If H-Score of MUC5AC is  $\geq 1.05$ , then SSA/P is predicted and if MUC5AC H-Score  $< 1.05$ , then HP is predicted (Figure 3C). Overall, the tree based model predicted HP, SSA/P and TA with SN/SP of

58%/95%, 79%/90% and 97%/83% respectively. The tree based model has 79% accuracy (85/108) in predicting all 3 groups simultaneously. Thus, we demonstrated the utility of differentially expressed mucins and associated O-glycans for effective discrimination of colorectal polyp subtypes at histological level.

## DISCUSSION

Colonoscopy has been a gold standard colon screening option. However, a substantial increase in CRCs arising during intervals between surveillance colonoscopy has been reported [28]. Interestingly, the precursor lesions

for these interval cancers were hypothesized to be SSA/Ps. Despite association of serrated polyps with CRC [11], the identification for serrated polyps at the morphological, histopathological, and molecular levels remains challenging. This confusion and inconsistency translates in misdiagnoses of serrated lesions due to interobserver variability and ultimately inappropriate follow-up care and therapy. Studies have demonstrated differential expression of mucins in colorectal polyp subtypes. Unfortunately, most of these studies were either limited to one or few mucins, did not consider modifications of associated O-glycan alterations, and focused only on detecting expression levels without emphasizing diagnostic utility. In addition, the previous studies did not explore the



**Figure 3: Combined performance of mucins and associated O-glycans for discriminating colorectal polyp subtypes.**

The combined utility of significantly altered mucins and O-glycans as markers for differential identification of polyp subtypes was evaluated utilizing multivariate regression models in conjunction with ROC curve based analysis. **A.** ROC curves representing diagnostic potential of combinatorial panels of mucins and associated O-glycans for differentiating SSA/P from HP. MUC17+MUC5AC combination differentiated SSA/P from HP with the AUC of 0.87. The model has 83% accuracy (60/72) in predicting SSA/P vs. HP. **B.** MUC1 and MUC17+CA19-9 combination differentiated SSA/P from TA with the AUC of 0.99. the model has 95% accuracy (71/75) in predicting SSA/P vs. TA. **C.** Decision tree based model to accurately define the combined diagnostic efficacy of mucins and associated O-glycans for polyp stratification. The panel of three markers CA19-9/MUC17/MUC5AC differentiated precancerous polyps (SSA/P and TA) from benign polyps (HP). Polyp sections having H-Score < 0.525 for CA19-9 predicts the presence of TA. For tissues sections having H-Score  $\geq 0.525$  for CA19-9, and H-score  $\geq 0.75$  for transmembrane mucin MUC17 the lesion is likely to be SSA/P. In cases where MUC17 H-score is < 0.75, and MUC5AC H-score is  $\geq 1.05$ , then SSA/P is predicted and if MUC5AC H-score < 1.05, then HP is predicted. The tree based model has 79% accuracy (85/108) in predicting all 3 groups simultaneously. HP: hyperplastic polyps; SSA/P: sessile serrated adenoma/polyps; TA: tubular adenoma.



**Table 4: Univariate and Multivariate logistic regression and ROC curve analysis of mucins and associated O-glycans for colorectal polyps classification**

Univariate Analysis							
Marker	Diagnosis	OR (95% CI)	p-value	AUC (95% CI)	Cutpoint	Sensitivity	Specificity
MUC1	SSA/P vs. HP	0.411 (0.156, 1.084)	0.072	0.6772 (0.5531, 0.8012)	≤0.30	0.487	0.848
	SSA/P vs. TA	0.188 (0.069, 0.509)	0.001	0.7639 (0.6545, 0.8733)	≤0.60	0.744	0.722
MUC4	SSA/P vs. HP	0.324 (0.160, 0.657)	0.0018	0.7257 (0.6089, 0.8425)	≤1.2	0.615	0.697
	SSA/P vs. TA	0.745 (0.395, 1.405)	0.36	0.5630 (0.4319, 0.6942)	≤1.2	0.615	0.5
MUC17	SSA/P vs. HP	14.376 (3.664, 56.41)	0.0001	0.7739 (0.6623, 0.8855)	≥0.8	0.667	0.909
	SSA/P vs. TA	5.990 (1.955, 18.346)	0.0017	0.7222 (0.6047, 0.8397)	≥0.80	0.667	0.778
MUC2	SSA/P vs. HP	0.530 (0.210, 1.337)	0.18	0.5960 (0.4640, 0.7279)	≤1.8	0.513	0.727
	SSA/P vs. TA	3.288 (1.573, 6.873)	0.0016	0.6991 (0.5765, 0.8216)	≥1.60	0.718	0.583
MUC5AC	SSA/P vs. HP	9.112 (2.912, 28.51)	0.0001	0.7995 (0.6947, 0.9043)	≥0.9	0.769	0.727
	SSA/P vs. TA	21.755 (5.985, 79.078)	<0.0001	0.9081 (0.8320, 0.9842)	≥0.60	0.923	0.806
Tn/STn-MUC1	SSA/P vs. HP	0.870 (0.317, 2.390)	0.79	0.5085 (0.3756, 0.6415)	≤0.8	0.179	0.939
	SSA/P vs. TA	2.472 (1.016, 6.016)	0.046	0.6353 (0.5095, 0.7611)	≥1.2	0.795	0.444
CA19-9	SSA/P vs. HP	1.279 (0.746, 2.194)	0.37	0.5287 (0.3907, 0.6668)	>2.1	0.333	0.606
	SSA/P vs. TA	54.383 (6.489, 455.7)	0.0002	0.9551 (0.9095, 1.0000)	≥0.80	0.872	1
Multivariate Analysis							
Diagnosis	Comparison	AUC	Lower	Upper	Threshold	Sensitivity	Specificity
SSA/P vs. HP	MUC17+MUC5AC	0.865	0.776	0.954	P(SP)≥0.4928	0.846	0.818
SSA/P vs. TA	MUC1+MUC17+CA19-9	0.9868	0.97	1	P(SPT)≥0.385	0.923	0.972

HP, hyperplastic polyps; SSA/P, sessile serrated adenoma/polyps; TA, tubular adenoma; OR, odds ratio; AUC, area under curve; CI, confidence interval

combinatorial power of mucins and associated O-glycans. We recently demonstrated the combined utility of mucins and associated O-glycans as markers to discriminate benign polyps from premalignant and malignant lesions [22]. Building on these observations the current study for the first time investigated expression of the major mucins and associated O-glycans that are significantly altered during colon carcinogenesis for discriminating SSA/P from HP and TA.

Previous studies have suggested that MUC1 is either absent or is weakly expressed in the normal colon [29, 30]. Focal to mild reactivity in HP and its inability to discriminate them from SSA/P has previously been described [29, 31-33]. In contrast, increased MUC1 positivity has been reported in conventional adenomas [29]. Consistent with these studies, we observed focal to mild immunoreactivity of MUC1 in SSA/P and significantly higher expression in TA. Based on these observations and a previous report demonstrating MUC1 expression to be independent of microsatellite status [34], it can be suggested that MUC1 may not be a prominent player in the serrated pathway of neoplasia. However, higher levels of MUC1 in TA along with reports of its even higher expression in CRC [30], and association with invasion risk [33] argue for the potential applicability of MUC1 expression as a predictor of malignant transformation in the adenoma-carcinoma sequence.

Congruent with previous studies [29], MUC4 appeared to be a prominent mucin gene expressed in the normal colon. Loss of MUC4 in CRC cases has previously been demonstrated [22, 35]. Interestingly, we observed a significant loss of MUC4 in SSA/P compared to HP, suggesting that loss of MUC4 might play an important role during serrated pathway of neoplasia. MUC4 was differentially expressed in SSA/P and TA *versus* benign lesions (HP) however; it did not appear as a significant contributor in the combinatorial panels of markers that discriminated these polyp subtypes. Interestingly, MUC4 was the only mucin whose expression was significantly impacted by the site of the polyp, suggesting the role of colonic microenvironment in the regulation of this mucin. Interestingly, we also found that MUC4 expression was significantly higher in polyps of the patients with a history of tobacco use. Long-term cigarette smoking is associated with colorectal polyps and increased risk of CRC mortality [36, 37]. Recently, smoke exposure was reported to result in increased Muc4 expression in the distal colon of mice [38]. Similarly, reports have shown that cigarette smoke upregulates expression of MUC4 [39]. The presence of MUC4 confers worse prognosis to early stage CRC patients [35, 40]. Therefore, in future it would be interesting to investigate the role of cigarette smoke-mediated upregulation of MUC4 during CRC progression.

MUC17 expression is significantly increased in CRC cases compared to colon inflammation cases [22]. In this study, we not only conducted comprehensive analysis of MUC17 expression in different polyp subtypes, but also observed its exceptional utility as a biomarker to discriminate SSA/P from HP and TA. MUC17 has previously been shown to be a top differentially expressed gene in SSA/P *versus* TA [41]. RNA-seq analysis also indicated 82-fold increase in MUC17 in SSA/P compared to adenomatous polyps, uninvolved and normal colon samples [25]. Interestingly, none of these studies considered HP to compare differential expression of genes in SSA/P, thus limiting the overall interpretations in terms of molecular characterization of SSA/P for their efficient diagnosis. We observed significantly increased MUC17 expression in HP and SSA/P compared to NC. Most importantly, MUC17 was significantly overexpressed in SSA/P in comparison to HP and TA cases, and emerged as a significant independent marker of SSA/P *versus* HP and TA with high specificity. In multivariate models, the combination of MUC17 and MUC5AC improved the diagnostic accuracy of SSA/P *versus* HP with 85% sensitivity and 82% specificity. Furthermore, MUC17 was found to be a significant contributor to the panel of diagnostic markers that discriminated SSA/P from TA. Due to the paucity of information regarding the association of MUC17 with BRAF mutation, CpG island methylator phenotype (CIMP), microsatellite instability (MSI) and microsatellite instability-high (MSI-H) colon carcinomas, additional studies addressing these aspects would help establish MUC17 as a marker of neo-transformation through the serrated neoplasia pathway.

Several studies have previously reported MUC5AC expression in normal colon, HP, SSA/P, TSA, TA, VA, TVA and mixed polyps [29, 30, 42-44] without emphasizing on its diagnostic utility for discriminating these polyp subtypes. Consistent with these studies, we also observed significant alterations in the expression of MUC5AC in HP, SSA/P, and TA when compared to the NC. Further, significant overexpression of MUC5AC was observed in SSA/P when compared to HP and TA. Conceptually, high positivity of MUC5AC represents an early gain of gastric differentiation in SSA/P and is significantly associated with MSI-H CRCs sporadic tumors ( $p < 0.001$ ) [45]. Increased expression of MUC5AC in SSA/P has recently been attributed to MUC5AC hypomethylation which in turn, is specifically associated with serrated lesions with BRAF mutation, CIMP-H or MSI [27]. Therefore, increased expression of MUC5AC in SSA/P may be reflective of neoplastic process. Interestingly, combination of MUC5AC and MUC17 discriminated SSA/P from HP with an area under curve (AUC) of 0.865 and SN/SP of 86%/82%. Moreover, MUC5AC was the only marker which could significantly differentiate SSA/P from HP and TA on basis of its differential localization pattern. Given these results, IHC for these markers may be helpful

to pathologists for effective discrimination of SSA/P from benign HP. Studies characterizing the role of MUC5AC during the serrated pathway of CRC progression would establish its utility as a predictive marker of neoplastic transformation.

Increased expression of MUC6 was found to be highly specific in distinguishing SSA/P from HP [16, 26]. Contrary to this, it was also reported that increased expression of MUC6 lacks specificity in distinguishing SSA/P and SSA/P with dysplasia from HP [31]. Owens *et al.* concluded that neither anatomic location nor polyp size account for the differences in MUC6 expression in serrated polyps [16, 42]. However, Bartley *et al.* concluded strong association of MUC6 expression with proximal location of serrated polyps [26]. Looking at these confounding reports, we also investigated expression of MUC6 in SSA/P, TA and HP samples. However, MUC6 immunoreactivity was absent in majority of cases (data not shown). These observation conflicts from previous studies could be attributed to difference in tissue sites, patient ethnicity, antibody used for study and/or limited sample numbers.

CA19-9 is a monosialosyl Lewis (a) blood group antigen that has been shown to be differentially expressed during adenoma-carcinoma sequence [22]. The importance of CA19-9 serum levels at the beginning of malignant alterations for monitoring colorectal tumors [46] and to discriminate premalignant lesions from cancer has previously been reported [47]. This is the first study to evaluate CA19-9 in SSA/P and assess its efficacy to discriminate colorectal polyps in combination with other mucins. The present study clearly demonstrated that CA19-9 expression is significantly increased in all colorectal polyp subtypes with highest levels in SSA/P, and lowest in TA in comparison to HP. Diagnostic utility of CA19-9 was evident from the fact that CA19-9 in combination with MUC1 and MUC17 predicted SSA/P from TA with very high sensitivity and specificity (92.3%/97.2%). Our assessment from multivariate models was further corroborated by a decision tree model where CA19-9 proved to be an effective marker to clearly discriminate TA from SSA/P and HP at the very first step.

Further, this study utilized, the 'Tree-based models' as an additional exploratory way of analyzing the data. Such models are advantageous in the sense that interactions are handled automatically as well as monotonic transformations of the x and y variables. They also easily handle multiple categories for the classification variable. However, 'Tree based models' are usually preferred when sample sizes are large. For our case, since the sample size was relatively small, we tested accuracy using both multivariate and tree-based model.

The promising results from this study lay the foundation for stratification of colorectal polyp subtypes on the basis of mucins and associated O-glycans immunostaining. The proposed panel of markers

from this study, in conjunction with current histologic diagnostic criterion may aid pathologists in better discrimination of the colorectal polyp subtypes. However, this study is limited by its small sample size and lack of enough samples from patients who had multiple polyps from different sites or of different subtypes. Further, Torlakovic *et al.* histologically subclassified HPs on basis of the epithelial mucin content into goblet cell-rich, microvesicular, and mucin-poor types [48]. Therefore, future studies could focus on further investigating the mucins and associated O-glycans expression in various subtypes of HP for further improvement in discrimination from SSA/P.

Cancer-associated post-translational modifications frequently produce aberrant O-linked glycoproteins which in combination with protein backbone induce auto-antibodies [49]. One recent study reported cancer-associated autoantibodies against a set of aberrantly glycosylated glycopeptides derived from mucin MUC1 and MUC4 in sera of CRC patients [49]. Therefore, future research efforts should also focus attention on identification of CRC-specific O-glycans on differentially expressed mucins (MUC1, MUC2, MUC4, MUC5AC, and MUC17) and develop CRC-specific glycopeptide microarrays for serum profiling of patients harboring colorectal polyps with malignant potential.

Efforts should continue to further elucidate the currently unexplored roles of mucins MUC4, MUC17, and MUC5AC in the progression of CRC *via* different pathways, along with their association with various genetic and epigenetic mutations in CRC. This will firmly establish their utility as markers for molecular classification of polyps. Also, investigation of molecular mechanisms associated with the upregulation of mucins like MUC17 and MUC5AC in SSA/P may further help in devising therapies targeting these lesions.

In conclusions, this study is the first critical step addressing the “gray zone” in differentiating SSA/P from other polyp subtypes utilizing aberrant mucin expression and glycosylation. The efficient discrimination of different colorectal polyp subtypes utilizing the novel panel of molecular markers (CA19-9/MUC17/MUC5AC) from the decision tree model, in conjunction with currently available tests and colonoscopy, may aid clinicians in devising improved colon cancer screening recommendations.

## MATERIALS AND METHODS

### Tissue specimens

After institutional review board (IRB) approval (IRB#167-15-EP), endoscopically resected formalin-fixed paraffin embedded colonic polyp tissue specimens- HP (*n*

= 33), SSA/P (*n* = 39), and TA (*n* = 36) were obtained from the Department of Pathology at University of Nebraska Medical Center (UNMC). The polyp diagnoses were based on the current WHO criteria [8]. Additionally, only those cases were considered for the study where the original histological diagnosis by pathologists was in consensus to the diagnosis made by a second pathologist (Dr. Yuri Sheinin at UNMC). The colorectal polyps used for this study includes samples from the same patient having different subtypes and/or multiple polyps of one type from different sites in colon. However, their numbers were insignificant. The normal colon tissue adjacent to the polyp lesion was also determined and considered for comparative analysis by the pathologist (Dr. Yuri M. Sheinin) on the basis of histological examination of Hematoxylin & Eosin stained tissue sections.

### Immunohistochemistry (IHC)

IHC was performed on the polyps following a standard protocol [50]. After overnight baking at 56°C, tissue sections were deparaffinized with xylene (10 min, 4X), rehydrated with graded alcohol (5 minutes each wash). Endogenous peroxidase activity was quenched using methanolic 3% hydrogen peroxide (30 minutes). Antigen was retrieved in 0.05M citrate buffer (pH 6.0, 95°C, 15 minutes), and blocked with 2.5% horse serum (ImmPRESS Universal antibody kit; Vector Laboratories, Burlingame, CA) (two hours). Tissue sections were then incubated with anti-mucin and associated O-glycans specific antibodies overnight at 4°C (Supplementary Table 1). Next, the tissue sections were washed with PBS-T (10 minutes, 4X) and incubated with anti-rabbit/anti-mouse secondary antibody conjugated with Horseradish peroxidase (ImmPRESS Universal antibody kit; Vector Laboratories) (30 minutes). Next, the sections were washed with PBS-T (10 minutes, 4X), developed for colorimetric detection by 3, 3' diaminobenzidine kit (Vector Laboratories, Burlingame, CA, USA) and counterstained with hematoxylin. Tissues were dehydrated (graded alcohol and xylene), dried, mounted with Permount, and evaluated by the pathologists.

The staining scores corresponding to the polyps as well as the normal colon adjacent to polyps was determined following the H-Score system published previously [51]. The intensity of staining for each of the markers was scored on a scale of 0-3 (0-negative, 1-weak, 2-intermediate, 3-strong intensity) to obtain intensity score (IS). For a given intensity, the percentage of cells positive for mucin or associated O-glycans within a given lesion was also scored on a scale of 0-1 (0: no cells staining, 1: 100% cells positively stained). The IS and corresponding score for the percentage of immunoreactive cells were then multiplied to obtain a histology Score (H-Score), representing overall expression ranging from 0-3. The respective tissue specimens were further classified by



H-score as having no reactivity (H-score = 0), focal reactivity (H-score > 0 but ≤ 0.1), mild reactivity (H-score > 0.1 but ≤ 1), moderate reactivity (H-score > 1.0 but ≤ 2.0) and strong reactivity (H-score > 2.0). Further, differential localization pattern of mucin and associated O-glycans expression within colon crypts was also considered for analysis.

### Statistical analysis

Patient characteristics for each polyp subtype were compared using chi-square or Fisher's exact test for categorical variables and ANOVA or Kruskal-Wallis test for continuous variables. H-score for each polyp subgroup were compared using Kruskal-Wallis test and Wilcoxon rank sum test for pairwise comparisons. Wilcoxon signed rank test was used to compare H-Score of polyp subtypes and the respective adjacent normal controls. Adjustments for multiple comparisons were made using Bonferroni's method. H-score was also categorized into reactivity groups, and these were compared by polyp subtype using Chi-square test. The localization pattern of secretory mucin MUC5AC in HP and SSA/P crypts was compared using Chi-square test. Univariate and multivariate logistic regression models and ROC curves were used to examine the individual markers (MUC1, MUC4, MUC17, MUC2, MUC5AC, CA19-9 and Tn/STn-MUC1), and a combinatorial panel of markers as predictors of each polyp subtype. ROC curves were used to determine optimal H-score cutpoints for markers. Firth's penalized maximum likelihood estimation was used to reduce bias in the parameter estimates because of separability in some of the multivariate models. Accuracy of multivariate models to differentiate SSA/P from TA and HP was calculated as the number of cases classified correctly divided by the total number of cases considered. SAS software version 9.3 was used for all other statistical analyses (SAS Institute Inc., Cary, NC). Further, the R package rpart was used to create a classification tree useful for biomarker-based medical decision making. Accuracy of tree based model was calculated as the number of cases classified correctly divided by the total number of cases considered.

### Abbreviations

AUC, area under curve; CI, confidence interval; CRC, colorectal cancer; H-Score, histology- score; HP, hyperplastic polyps; IS, intensity score; MUC, mucin; NC, adjacent normal colon; OR, odds ratio; ROC, receiver operating characteristics; SE, standard error; SN, sensitivity; SP, specificity; SSA/P, sessile serrated adenoma/polyps; TA, tubular adenoma; TSA, traditional serrated adenoma; TVA, tubulovillous adenoma; VA, villous adenoma

### ACKNOWLEDGMENTS

We would like to thank Dr. Sandra J. Gendler at Mayo Clinic, USA for anti-MUC1 antibody (HMFG2), Dr. Henrick Clausen of the University of Copenhagen, Sweden and Dr. Michael A. Hollingsworth from University of Nebraska Medical Center for providing with the anti-Tn/STn on MUC1 antibody (5E5). We would also like to thank Kavita Mallya (UNMC) for technical support.

### CONFLICTS OF INTEREST

Authors declare no conflict of interest.

### GRANT SUPPORT

This work was supported, in parts, by the grants from National Institutes of Health (Cancer Center Support Grant P30 CA036727, EDNRN CA111294, SPORE P50 CA127297, RO1 CA195586, RO1 CA183459 and U54 CA163120).

### REFERENCES

1. Jain P, Mondal SK, Sinha SK, Mukhopadhyay M, Chakraborty I. Diagnostic and prognostic significance of different mucin expression, preoperative CEA, and CA-125 in colorectal carcinoma: A clinicopathological study. *J Nat Sci Biol Med.* 2014; 5: 404-408.
2. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology.* 2015; 149: 1177-1190.
3. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology.* 2015; 149: 1204-1225.
4. Pancione M, Remo A, Colantuoni V. Genetic and epigenetic events generate multiple pathways in colorectal cancer progression. *Patholog Res Int.* 2012; 2012: 509348.
5. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology.* 2008; 135: 1079-1099.
6. Crockett SD, Snover DC, Ahnen DJ, Baron JA. Sessile serrated adenomas: an evidence-based guide to management. *Clin Gastroenterol Hepatol.* 2015; 13: 11-26.
7. IJspeert JE, Vermeulen L, Meijer GA, Dekker E. Serrated neoplasia-role in colorectal carcinogenesis and clinical implications. *Nat Rev Gastroenterol Hepatol.* 2015; 12: 401-409.
8. Rex DK, Ahnen DJ, Baron JA, Batts KP, Burke CA, Burt RW, Goldblum JR, Guillem JG, Kahi CJ, Kalady MF, O'Brien MJ, Odze RD, Ogino S et al. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol.* 2012; 107: 1315-1329.



9. Aust DE, Baretton GB. Serrated polyps of the colon and rectum (hyperplastic polyps, sessile serrated adenomas, traditional serrated adenomas, and mixed polyps)-proposal for diagnostic criteria. *Virchows Arch.* 2010; 457: 291-297.
10. Lazarus R, Junttila OE, Karttunen TJ, Makinen MJ. The risk of metachronous neoplasia in patients with serrated adenoma. *Am J Clin Pathol.* 2005; 123: 349-359.
11. Li D, Jin C, McCulloch C, Kakar S, Berger BM, Imperiale TF, Terdiman JP. Association of large serrated polyps with synchronous advanced colorectal neoplasia. *Am J Gastroenterol.* 2009; 104: 695-702.
12. Lash RH, Genta RM, Schuler CM. Sessile serrated adenomas: prevalence of dysplasia and carcinoma in 2139 patients. *J Clin Pathol.* 2010; 63: 681-686.
13. Ban S, Mitomi H, Horiguchi H, Sato H, Shimizu M. Adenocarcinoma arising in small sessile serrated adenoma/polyp (SSA/P) of the colon: clinicopathological study of eight lesions. *Pathol Int.* 2014; 64: 123-132.
14. Goldstein NS. Small colonic microsatellite unstable adenocarcinomas and high-grade epithelial dysplasias in sessile serrated adenoma polypectomy specimens: a study of eight cases. *Am J Clin Pathol.* 2006; 125: 132-145.
15. Oono Y, Fu K, Nakamura H, Iriguchi Y, Yamamura A, Tomino Y, Oda J, Mizutani M, Takayanagi S, Kishi D, Shinohara T, Yamada K, Matumoto J et al. Progression of a sessile serrated adenoma to an early invasive cancer within 8 months. *Dig Dis Sci.* 2009; 54: 906-909.
16. Owens SR, Chiosea SI, Kuan SF. Selective expression of gastric mucin MUC6 in colonic sessile serrated adenoma but not in hyperplastic polyp aids in morphological diagnosis of serrated polyps. *Mod Pathol.* 2008; 21: 660-669.
17. Farris AB, Misdraji J, Srivastava A, Muzikansky A, Deshpande V, Lauwers GY, Mino-Kenudson M. Sessile serrated adenoma: challenging discrimination from other serrated colonic polyps. *Am J Surg Pathol.* 2008; 32: 30-35.
18. Glatz K, Pritt B, Glatz D, Hartmann A, O'Brien MJ, Blaszyk H. A multinational, internet-based assessment of observer variability in the diagnosis of serrated colorectal polyps. *Am J Clin Pathol.* 2007; 127: 938-945.
19. Sandmeier D, Seelentag W, Bouzourene H. Serrated polyps of the colorectum: is sessile serrated adenoma distinguishable from hyperplastic polyp in a daily practice?. *Virchows Arch.* 2007; 450: 613-618.
20. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology.* 2012; 143: 844-857.
21. Johansson ME, Sjoval H, Hansson GC. The gastrointestinal mucus system in health and disease. *Nat Rev Gastroenterol Hepatol.* 2013; 10: 352-361.
22. Krishn SR, Kaur S, Smith LM, Johansson SL, Jain M, Patel A, Gautam SK, Hollingsworth MA, Mandel U, Clausen H, Lo WC, Fan WL, Manne U et al. Mucins and associated glycan signatures in colon adenoma-carcinoma sequence: Prospective pathological implication(s) for early diagnosis of colon cancer. *Cancer Lett.* 2016; 374: 304-314.
23. Kufe DW. Mucins in cancer: function, prognosis and therapy. *Nat Rev Cancer.* 2009; 9: 874-885.
24. Sheng YH, Hasnain SZ, Florin TH, McGuckin MA. Mucins in inflammatory bowel diseases and colorectal cancer. *J Gastroenterol Hepatol.* 2012; 27: 28-38.
25. Delker DA, McGettigan BM, Kanth P, Pop S, Neklason DW, Bronner MP, Burt RW, Hagedorn CH. RNA sequencing of sessile serrated colon polyps identifies differentially expressed genes and immunohistochemical markers. *PLoS One.* 2014; 9: e88367.
26. Bartley AN, Thompson PA, Buckmeier JA, Kepler CY, Hsu CH, Snyder MS, Lance P, Bhattacharyya A, Hamilton SR. Expression of gastric pyloric mucin, MUC6, in colorectal serrated polyps. *Mod Pathol.* 2010; 23: 169-176.
27. Renaud F, Mariette C, Vincent A, Wacrenier A, Maunoury V, Leclerc J, Coppin L, Crepin M, Van S, I, Leteurtre E, Buisine MP. The serrated neoplasia pathway of colorectal tumors: Identification of MUC5AC hypomethylation as an early marker of polyps with malignant potential. *Int J Cancer.* 2015; 138: 1472-1481.
28. Adler J, Robertson DJ. Interval Colorectal Cancer After Colonoscopy: Exploring Explanations and Solutions. *Am J Gastroenterol.* 2015; 110: 1657-1664.
29. Biemer-Huttmann AE, Walsh MD, McGuckin MA, Ajioka Y, Watanabe H, Leggett BA, Jass JR. Immunohistochemical staining patterns of MUC1, MUC2, MUC4, and MUC5AC mucins in hyperplastic polyps, serrated adenomas, and traditional adenomas of the colorectum. *J Histochem Cytochem.* 1999; 47: 1039-1048.
30. Percinel S, Savas B, Ensari A, Kuzu I, Kuzu MA, Bektas M, Cetinkaya H, Kursun N. Mucins in the colorectal neoplastic spectrum with reference to conventional and serrated adenomas. *Turk J Gastroenterol.* 2007; 18: 230-238.
31. Gibson JA, Hahn HP, Shahsafaei A, Odze RD. MUC expression in hyperplastic and serrated colonic polyps: lack of specificity of MUC6. *Am J Surg Pathol.* 2011; 35: 742-749.
32. Hanby AM, Poulosom R, Singh S, Jankowski J, Hopwood D, Elia G, Rogers L, Patel K, Wright NA. Hyperplastic polyps: a cell lineage which both synthesizes and secretes trefoil-peptides and has phenotypic similarity with the ulcer-associated cell lineage. *Am J Pathol.* 1993; 142: 663-668.
33. Molaei M, Mansoori BK, Mashayekhi R, Vahedi M, Pourhoseingholi MA, Fatemi SR, Zali MR. Mucins in neoplastic spectrum of colorectal polyps: can they provide predictions?. *BMC Cancer.* 2010; 10: 537.
34. Biemer-Huttmann AE, Walsh MD, McGuckin MA, Simms LA, Young J, Leggett BA, Jass JR. Mucin core protein expression in colorectal cancers with high levels of microsatellite instability indicates a novel pathway of

- morphogenesis. *Clin Cancer Res.* 2000; 6: 1909-1916.
35. Shanmugam C, Jhala NC, Katkooori VR, Wan W, Meleth S, Grizzle WE, Manne U. Prognostic value of mucin 4 expression in colorectal adenocarcinomas. *Cancer.* 2010; 116: 3577-3586.
  36. Haque TR, Bradshaw PT, Crockett SD. Risk factors for serrated polyps of the colorectum. *Dig Dis Sci.* 2014; 59: 2874-2889.
  37. Yang B, Jacobs EJ, Gapstur SM, Stevens V, Campbell PT. Active smoking and mortality among colorectal cancer survivors: the Cancer Prevention Study II nutrition cohort. *J Clin Oncol.* 2015; 33: 885-893.
  38. Allais L, Kerckhof FM, Verschuere S, Bracke KR, De SR, Laukens D, Van den Abbeele P, De VM, Boon N, Brusselle GG, Cuvelier CA, Van de Wiele T. Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environ Microbiol.* 2015; 18:1352-1363.
  39. Momi N, Ponnusamy MP, Kaur S, Rachagani S, Kunigal SS, Chellappan S, Ouellette MM, Batra SK. Nicotine/cigarette smoke promotes metastasis of pancreatic cancer through alpha7nAChR-mediated MUC4 upregulation. *Oncogene.* 2013; 32: 1384-1395.
  40. Khanh dT, Mekata E, Mukaisho K, Sugihara H, Shimizu T, Shiomi H, Murata S, Naka S, Yamamoto H, Endo Y, Tani T. Transmembrane mucin MUC1 overexpression and its association with CD10(+) myeloid cells, transforming growth factor-beta1 expression, and tumor budding grade in colorectal cancer. *Cancer Sci.* 2013; 104: 958-964.
  41. Caruso M, Moore J, Goodall GJ, Thomas M, Phillis S, Tyskin A, Cheetham G, Lerda N, Takahashi H, Ruzskiewicz A. Over-expression of cathepsin E and trefoil factor 1 in sessile serrated adenomas of the colorectum identified by gene expression analysis. *Virchows Arch.* 2009; 454: 291-302.
  42. Fujita K, Hirahashi M, Yamamoto H, Matsumoto T, Gushima M, Oda Y, Kishimoto J, Yao T, Iida M, Tsuneyoshi M. Mucin core protein expression in serrated polyps of the large intestine. *Virchows Arch.* 2010; 457: 443-449.
  43. Higuchi T, Sugihara K, Jass JR. Demographic and pathological characteristics of serrated polyps of colorectum. *Histopathology.* 2005; 47: 32-40.
  44. Mochizuka A, Uehara T, Nakamura T, Kobayashi Y, Ota H. Hyperplastic polyps and sessile serrated 'adenomas' of the colon and rectum display gastric pyloric differentiation. *Histochem Cell Biol.* 2007; 128: 445-455.
  45. Kim JH, Kim KJ, Rhee YY, Bae JM, Cho NY, Lee HS, Kang GH. Gastric-type expression signature in serrated pathway-associated colorectal tumors. *Hum Pathol.* 2015; 46: 643-656.
  46. Stajic S, Novakovic R, Kostic N, Josifovski J, Simic P, Katic V, Kovijanic M. The importance of monitoring colorectal adenomas by tumor markers (ROC analysis). *Med Pregl.* 1993; 46 Suppl 1: 69-71.
  47. Goldberg EM, Simunovic LM, Drake SL, Mueller WF, Jr., Verrill HL. Comparison of serum CA 19-9 and CEA levels in a population at high risk for colorectal cancer. *Hybridoma.* 1989; 8: 569-575.
  48. Torlakovic E, Skovlund E, Snover DC, Torlakovic G, Nesland JM. Morphologic reappraisal of serrated colorectal polyps. *Am J Surg Pathol.* 2003; 27: 65-81.
  49. Pedersen JW, Blixt O, Bennett EP, Tarp MA, Dar I, Mandel U, Poulsen SS, Pedersen AE, Rasmussen S, Jess P, Clausen H, Wandall HH. Seromic profiling of colorectal cancer patients with novel glycopeptide microarray. *Int J Cancer.* 2011; 128: 1860-1871.
  50. Kaur S, Shimizu T, Baine MJ, Kumar S, Batra SK. Immunohistochemistry of pancreatic neoplasia. *Methods Mol Biol.* 2013; 980: 29-42.
  51. Kaur S, Momi N, Chakraborty S, Wagner DG, Horn AJ, Lele SM, Theodorescu D, Batra SK. Altered expression of transmembrane mucins, MUC1 and MUC4, in bladder cancer: pathological implications in diagnosis. *PLoS One.* 2014; 9: e92742.
  52. Burchell J, Gendler S, Taylor-Papadimitriou J, Girling A, Lewis A, Millis R, Lampion D. Development and characterization of breast cancer reactive monoclonal antibodies directed to the core protein of the human milk mucin. *Cancer Res.* 1987; 47: 5476-5482.
  53. Moniaux N, Varshney GC, Chauhan SC, Copin MC, Jain M, Wittel UA, Andrianifahanana M, Aubert JP, Batra SK. Generation and characterization of anti-MUC4 monoclonal antibodies reactive with normal and cancer cells in humans. *J Histochem Cytochem.* 2004; 52: 253-261.
  54. Moniaux N, Junker WM, Singh AP, Jones AM, Batra SK. Characterization of human mucin MUC17. Complete coding sequence and organization. *J Biol Chem.* 2006; 281: 23676-23685.
  55. Sorensen AL, Reis CA, Tarp MA, Mandel U, Ramachandran K, Sankaranarayanan V, Schwientek T, Graham R, Taylor-Papadimitriou J, Hollingsworth MA, Burchell J, Clausen H. Chemoenzymatically synthesized multimeric Tn/STn MUC1 glycopeptides elicit cancer-specific anti-MUC1 antibody responses and override tolerance. *Glycobiology.* 2006; 16: 96-107.