

Elevated Colonic Mucin Expression Correlates with Extended Time to Surgery for Ulcerative Colitis Patients

Melanie A. Eshelman^{1,2}, N. Arjun Jeganathan², Kathleen M. Schieffer³, Bryan P. Kline², Megan Mendenhall^{1,2}, Sue Deiling², Leonard Harris², Walter A. Koltun², Gregory S. Yochum^{1,2}

1) Department of Biochemistry & Molecular Biology, Pennsylvania State University College of Medicine, Hershey, PA
2) Department of Surgery, Division of Colon and Rectal Surgery, Pennsylvania State University College of Medicine, Hershey, PA
3) The Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio, USA.

Address for correspondence:

Gregory S. Yochum
The Pennsylvania State University College of Medicine
Department of Biochemistry & Molecular Biology & Surgery
500 University Dr.
Hershey, PA, USA. 17033
gsy3@psu.edu

ABSTRACT

Background & Aims: Both genetic and environmental factors contribute to the development and persistence of ulcerative colitis (UC). As supported by differential responses to therapy, multiple subclasses of disease likely comprise UC. We reasoned that profiling the colonic transcriptomes may offer one approach to molecular subtype UC.

Methods: We conducted RNA-sequencing (RNA-seq) on full-thickness colonic tissues from 26 UC patients undergoing colectomy. Hierarchical clustering from transcriptomic data identified disease subsets. Subsets were characterized using differential gene expression analysis, cell type deconvolution, and network analysis.

Results: We identified two UC subsets that were distinguished by 957 differentially expressed genes. Cluster 1 was enriched in genes associated with intestinal epithelial cell (IEC) differentiation, while cluster 2 was enriched in genes associated with epithelial-to-mesenchymal transition (EMT) and inflammatory responses. Cluster 1 was associated with an extended time from diagnosis to colectomy [hazard ratio = 0.45 (95% CI: 0.14-0.88); $p=0.03$]. Of cluster 1 genes, elevated *MUC5B*, *MUC4*, and *MUC2* expression displayed the strongest correlation with increased time to surgery [hazard ratio = 0.37 (95% CI: 0.11-0.61); $p=0.0044$].

Conclusions: Our transcriptome analysis indicates that UC can be sub-classified into at least two molecular signatures. We found that elevated mucin gene expression correlated with prolonged time to colectomy following diagnosis. This work identified *MUC5B*, *MUC4*, and *MUC2* as potential prognostic indicators of disease severity, as reflected in time to surgery after diagnosis.

Key words: ulcerative colitis – RNA-seq – transcriptome – mucin.

Abbreviations: CD: Crohn's disease; CI: confidence interval; cDC: conventional dendritic cell; EMT: epithelial-to-mesenchymal transition; GSEA: gene set enrichment analysis; GWAS: genome-wide association studies; IBD: inflammatory bowel disease; IEC: intestinal epithelial cell; ME: module eigengene; MUC: mucin; PCA: principal component analysis; RNA-seq: RNA-sequencing; RT-qPCR: reverse transcription quantitative PCR; UC: ulcerative colitis; WGCNA: weighted gene coexpression network analysis.

INTRODUCTION

Ulcerative colitis (UC) is an inflammatory bowel disease (IBD) that results from chronic and unresolved inflammation of the colonic mucosa [1]. It is generally accepted that UC develops from an environmental trigger in genetically susceptible individuals [2]. Genome-wide association studies (GWAS) have identified hundreds of genetic variants associated with disease predisposition, and this work has

been instrumental in implicating defective immune signaling and autophagic processes in UC pathogenesis [2]. While the precise environmental insults remain unclear, those that cause microbial dysbiosis have gained recent interest [3]. As UC pathophysiology involves genetics and the environment, transcriptomics, or profiling and cataloging gene expression signatures [4] offer one approach that essentially considers both factors simultaneously. Indeed, this approach has been successful in defining molecular subsets of IBD [5-7].

While the development of biologic therapies has improved the prognosis of UC patients, between 10 and 40% will still require surgery during the course of their disease [8, 9]. Despite the success of total proctocolectomy to cure UC and remove the risk of colorectal adenocarcinoma, surgery carries inherent risks of morbidity and mortality [8-10]. Urgent (or

Received: 02.07.2019

Accepted: 09.11.2019

emergent) surgery carries significantly higher mortality rates than elective surgery [8, 9, 11]. Thus, prognostic indicators that predict when UC patients may benefit from earlier surgical intervention would improve patient outcomes. While there are clinical correlates that aid in surgical decision making, the molecular pathways that distinguish individuals who have a greater or lesser likelihood of requiring surgery are incompletely defined [12, 13]. Such biological determinants would have real clinical value.

A single layer of intestinal epithelial cells (IECs) lines the colon and forms a barrier to separate the underlying submucosal tissues from the luminal contents of the gut [14]. A mucus bilayer secreted by goblet cells protects the colonic epithelium and also serves as an environmental niche for the commensal microbiota [15, 16]. The mucin family of glycoproteins is the primary structural component of the mucus bilayer and over 20 MUC genes have been identified in humans. The predominant mucin in the colon is the gel-forming MUC2; however, membrane-bound MUC1, MUC3, MUC4 and the gel-forming MUC5B are also expressed [15, 17]. Disruption of the mucosal barrier precedes colitis in animal models and UC patients exhibit a loss of goblet cells and a thinning of the mucus layer [15, 16]. Therefore, maintenance of the protective mucus bilayer is essential to prevent UC.

We conducted RNA-sequencing (RNA-seq) analysis on surgically resected, colonic tissue from UC patients. We identified two molecular subsets of UC. In this report we focus on the subset that was characterized by differential expression of genes involved in IEC differentiation.

METHODS

Specimen collection

The Pennsylvania State University College of Medicine Institutional Review Board approved this study (IRB Protocol No. HY98-057EP-A). Prior to colectomy, patients diagnosed with UC gave informed consent to have surgically resected tissue collected and banked into the Penn State Hershey Colorectal Disease Biobank. Diagnoses were confirmed using colonoscopy and biopsy. Surgical specimens were immediately transferred from the operating room to the surgical pathology lab, where the macroscopic disease severity of the tissue, in terms of mucosal ulcerations and immune cell infiltrates, was determined by a pathologist and surgeon. Diseased tissues from the sigmoid colons were stored in RNAlater (Invitrogen) at -80°C until they were assayed. Specimens were collected from May 2010, to August 2017.

RNA-sequencing

RNA-sequencing was performed on full thickness resected sigmoid tissue from 26 UC patients. Total RNA was isolated from diseased tissue by pestle homogenization in TRIzol (Ambion). After addition of chloroform and centrifugation at 12,000×g for 15 min., the aqueous layer was purified using the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. From purified RNAs, cDNA libraries were prepared and subjected to RNA sequencing on a MiSeq2500 (Illumina) as described previously [18, 19]. FASTQ files were uploaded to the Galaxy web platform (usegalaxy.org)

[20]. FastQC (v0.11.6) was used to validate the quality of the sequenced reads. The reads were aligned to the human genome (build hg19) and quantified as described previously [18, 19]. The data can be accessed on the Gene Expression Omnibus database using reference number, GSE130038.

Hierarchical clustering and differential gene expression analysis

Regular log (rlog) normalized counts were generated using the R package DESeq2 [21]. Normalized counts for the 5000 most variable genes were used to conduct unsupervised hierarchical clustering with a centered Euclidian correlation coefficient algorithm and a complete linkage method using the R package pheatmap [22]. Patients were divided into two clusters based on the first branch point of the clustering dendrogram. Principal components analysis (PCA) using the R package DESeq2 was used to confirm the transcriptional uniqueness of the two patient subsets. Genes that were differentially expressed between the clusters were identified using DESeq2 and subsequently visualized as a volcano plot using the R package plotly [23]. Normalized counts generated by DESeq2 were used for Gene Set Enrichment Analysis (GSEA) of the clusters [24].

Cell type deconvolution

To computationally estimate proportions of cell subsets within the tissue of patients from each cluster, we performed cell-type deconvolution using xCell [25]. We considered only cell types that had significant enrichment scores (FDR corrected *p* values < 0.2 in at least 80% of the samples). We identified the following cell types: B-cells, CD4+ memory T-cells, epithelial cells, M2 macrophages, plasma cells, smooth muscle, and cDCs. We calculated the significance between the cell proportion scores of cluster 1 and cluster 2 individuals for each cell type using a two-tailed student's *t*-test with multiple corrections.

Network analysis

Network analysis was performed from DESeq2 normalized counts using the weighted gene coexpression network analysis (WGCNA) package in R [26]. Differentially expressed genes (*padj* < 0.01) were subjected to WGCNA. Following the standard WGCNA protocol, with the use of a soft thresholding power of 8, clustering was first performed using topological overlap matrix (TOM)-based dissimilarity followed by a clustering of consensus module eigengenes (MEs). This analysis resulted in four MEs comprised of genes that are highly coexpressed. The grey ME consisted of genes that were unassigned within an ME and was not further evaluated. Pearson correlation was performed to identify the relationship between MEs and clinical traits.

RT-QPCR

Total RNA isolated for RNA-seq was transcribed into cDNA and qPCR was performed as described previously [27]. Primers used in PCR included: *PPIA*, 5'-TGC CTA TCC TAG AGG TGG CGG ATT-3' and 5'-GGT GGT GGT AAT TCA CGC AGA AGG-3'; *GAPDH*, 5'-CCA GCA AGA GCA CAA GAG GAA GAG-3' and 5'-CAA GGG GTC TAC ATG GCA ACT

GTG-3'; *TUBB*, 5'-CCG GAA CTC ACC CAG CAG GTC T-3' and 5'-GAC ATC CGA CCA CGG AAG ACA G-3'; *MUC5B*, 5'-CAT CCT GCA CAC CTA CAC CCA CGT-3' and 5'-TGG GGA TGG AGG CAG AAC GTT C-3'; *MUC4*, 5'-AGC ATG AAA CTC GAC GCG TTC TTC-3' and 5'-AGA AGC GCA GGA CCA CGA ACG T-3'; and *MUC2*, 5'-GAG AAA ACC AGC CAG CGT GAG G-3' and 5'-CAG CTC TCG ATG TGG GTG TAG GTG-3'. Expression levels were normalized to the average of *PPIA*, *GAPDH*, and *TUBB*.

Statistical analysis

Survival curves were drawn using GraphPad Prism and significance was assessed using the Mantel-Cox test. Hazard ratios were calculated using the logrank test. Spearman's correlation was used to assess the association between time to surgery and gene expression.

RESULTS

The demographic and clinical characteristics of the patients are shown in Table I.

Transcriptome analysis identifies two distinct UC subsets

Patients demographics and clinical characteristics are shown in Table I. We previously used transcriptomics to identify clinically relevant digestive disease subsets [18, 19, 28, 29]. Here, we conducted RNA-seq analysis on RNAs isolated from the diseased colonic sigmoid tissue of 26 UC patients that were previously consented into our Colorectal Disease Biobank. We included an equal number of males and females in our study and all had received medication to treat their disease, including 5-ASA compounds, immunomodulators, corticosteroids, and biologics (Table I). Patients predominantly presented with extensive colitis (Montreal classification E3) and the most common indication for surgery was failure of medical management with four patients requiring urgent colectomy for severe colitis (Table I) [30]. In addition, three patients had

Table I. Patient characteristics

Parameter	Ulcerative colitis (n=26)
Sex, male, n (%)	14 (53.8), 12 (46.2)
Smoking status, positive*, n (%)	16 (61.5), 10 (38.5)
Age at diagnosis, yrs, mean \pm SD	45.2 \pm 16.3
Age at surgery, yrs, mean \pm SD	52.1 \pm 14.6
Body mass index, kg/m ² , mean \pm SD	26.7 \pm 4.9
Type of surgery, elective, n (%), urgent, n (%)	22 (84.6), 4 (15.4)
Disease extent (Montreal classification)	19 (73.1), 3 (11.5), 4 (15.4)
- proctitis (E1), n (%)	0 (0.0)
- left-sided colitis (E2), n (%)	5 (19.2)
- extensive (pan-)colitis (E3), n (%)	21 (80.8)
Prior medication use	
5-ASA compounds, n (%)	21 (80.8)
immunomodulators, n (%)	10 (38.5)
corticosteroids, n (%)	18 (69.2)
biologics, n (%)	20 (76.9)

*Positive smoking status includes both current and former smokers at the time of surgery

surgery for suspected dysplasia. Using hierarchical clustering of transcriptional signatures, we identified two patient clusters consisting of 16 and 10 individuals, referred to as cluster 1 and cluster 2, respectively (Fig. 1A). Patients in these clusters did not differ with respect to gender, smoking status, age at diagnosis or surgery, or BMI (Table II). Additionally, the disease extent in both groups was similar, with most individuals in each group presenting with extensive colitis (Table II). Prior medication use was also comparable among both groups (Table II). In both clusters, the most common indication for surgery was failure of medical management with two patients in each group requiring urgent colectomy for severe colitis. Two patients in cluster 1 and one patient in cluster 2 had surgery for dysplasia.

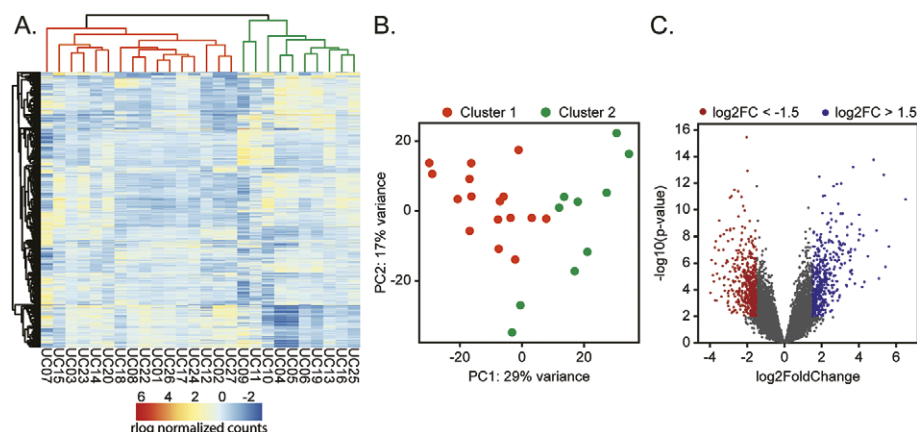


Fig. 1. Transcriptome analysis identifies two distinct UC subsets. (A) Heatmap depicting results from hierarchical clustering analysis of the transcriptional signatures of UC (n=26) patients. Cluster 1 and 2 are shown in the dendrogram as orange and green, respectively. (B) Principal components analysis plot of cluster 1 and cluster 2 patients. (C) Volcano plot of differentially expressed genes in cluster 1 (n=16), compared to cluster 2 (n=10). Blue and red points represent significantly up- (padj < 0.05 and log2FoldChange > 1.5) and down-regulated (padj < 0.05 and log2FoldChange < -1.5) genes, respectively.

Table II. Patient demographics and clinical characteristics for UC clusters

Parameter	Cluster 1	Cluster 2	p-value
Patients (n)	16	10	n.a.
Sex, male, n (%)	9 (56.2)	5 (50.0)	1.000
Smoking status, positive*, n (%)	8 (50.0)	8 (80.0)	0.218
Age at diagnosis, yrs, mean \pm SD	42.8 \pm 17.2	49.0 \pm 14.7	0.355
Age at surgery, yrs, mean \pm SD	51.4 \pm 15.5	53.3 \pm 13.6	0.758
Body mass index, kg/m ² , mean \pm SD	26.9 \pm 5.3	26.3 \pm 4.4	0.801
Disease extent (Montreal classification)			1.000
- left-sided colitis (E2), n (%)	3 (18.8)	2 (20.0)	
- extensive (pan-)colitis (E3), n (%)	13 (81.2)	8 (80.0)	
Prior medication use			
- 5-ASA compounds, n (%)	13 (81.2)	8 (80.0)	1.000
- immunomodulators, n (%)	6 (37.5)	4 (40.0)	1.000
- corticosteroids, n (%)	11 (68.8)	7 (70.0)	1.000
- biologics, n (%)	12 (75.0)	8 (80.0)	1.000

*Positive smoking status includes both current and former smokers at the time of surgery; Continuous variables analyzed by two-tailed t-test, categorical variables analyzed by two-tailed Fisher's Exact Test

The transcriptomes of cluster 1 and cluster 2 patients occupied distinct spaces in a principal components' analysis demonstrating that they represented separate subsets of UC (Fig. 1B). Differential gene expression analysis found that in comparison to cluster 2, expression of 957 genes were significantly different ($\text{padj} < 0.05$), with decreased expression of 511 genes ($\log_2 \text{FC} < -1.5$) and increased expression of 446 genes ($\log_2 \text{FC} > 1.5$) (Fig. 1C) in cluster 1.

The cell composition and molecular signatures of UC subsets are distinct

We used deconvolution of the RNA-seq data to predict which cell types were present in the patients' tissues of each cluster. We noted an enrichment of epithelial cells in cluster 1 whereas in cluster 2, we observed increases in M2 macrophages, plasma cells, and conventional dendritic cells (cDCs) (Fig. 2A). Given the prevalence of epithelial cells in cluster 1, we generated a set of established IEC differentiation genes and using gene set enrichment analysis (GSEA), we found that markers of IEC differentiation were expressed more abundantly in cluster 1 than cluster 2 (Fig. 2B and C). For cluster 2, GSEA using the hallmark gene set collection identified gene sets enriched for EMT and inflammatory responses (Fig. 2D-G).

Cluster 1 patients are associated with a delayed time to surgical intervention after diagnosis

Given the IEC enrichment of cluster 1, which may be indicative of a restored or intact colonic epithelium, we sought to determine whether patients comprising this cluster had more favorable clinical outcomes than those of cluster 2. Using the differentially expressed genes between the clusters and weighted gene co-expression network analysis (WGCNA), we established module eigengenes (MEs) for each cluster (Fig. 3A). Cluster 1 was associated with the turquoise and yellow MEs whereas cluster 2 was associated with the blue and brown MEs (Fig. 3A). We found no correlation between any of the four MEs and gender, smoking status, age at diagnosis, age at surgery, or

BMI. However, the cluster 1 turquoise ME correlated positively with the duration in time from UC diagnosis until surgical intervention (Fig. 3A). Genes that shared expression patterns with the turquoise ME were also most highly correlated with time to surgery after diagnosis (Fig. 3B). Further analysis found that cluster 1 patients required surgery an average of 67.6 months following initial diagnosis, whereas cluster 2 patients required surgery after only 27.3 months [hazard ratio = 0.45 (95% CI: 0.14-0.88); $p=0.03$] (Fig. 3C).

Elevated MUC levels correlated with a delayed requirement for surgical intervention

We examined the relationship between each of the top 25 differentially expressed IEC differentiation genes and time to surgery by calculating the Spearman's correlation between these two variables. Several of the genes most associated with duration to surgery after diagnosis were members of the MUC family of glycoproteins (Fig. 4A). Higher expression of *MUC5B*, *MUC4*, and *MUC2* was significantly associated with increased time to surgical intervention (Fig. 4A, B). Importantly, the association between *MUC5B*, *MUC4*, and *MUC2* and time to surgery was confirmed using RT-qPCR analysis of colonic tissue RNAs isolated from the same 26 UC patients in the cohort (Suppl. Fig. 1). We ranked all UC patients in our study, indeterminate of cluster association, by their average *MUC5B/MUC4/MUC2* gene expression levels and divided them at the median into high and low *MUC* expression groups. We found that individuals who had higher *MUC* levels persisted with their disease for an average period of 7.2 years, longer after diagnosis than those with lower *MUC* levels [hazard ratio = 0.37 (95% CI: 0.11-0.61); $p=0.0044$] (Fig. 4C).

DISCUSSION

Around 15% of UC patients require colectomy during their lifetime [1]. For those who require surgery, individuals that undergo elective intervention have substantially better

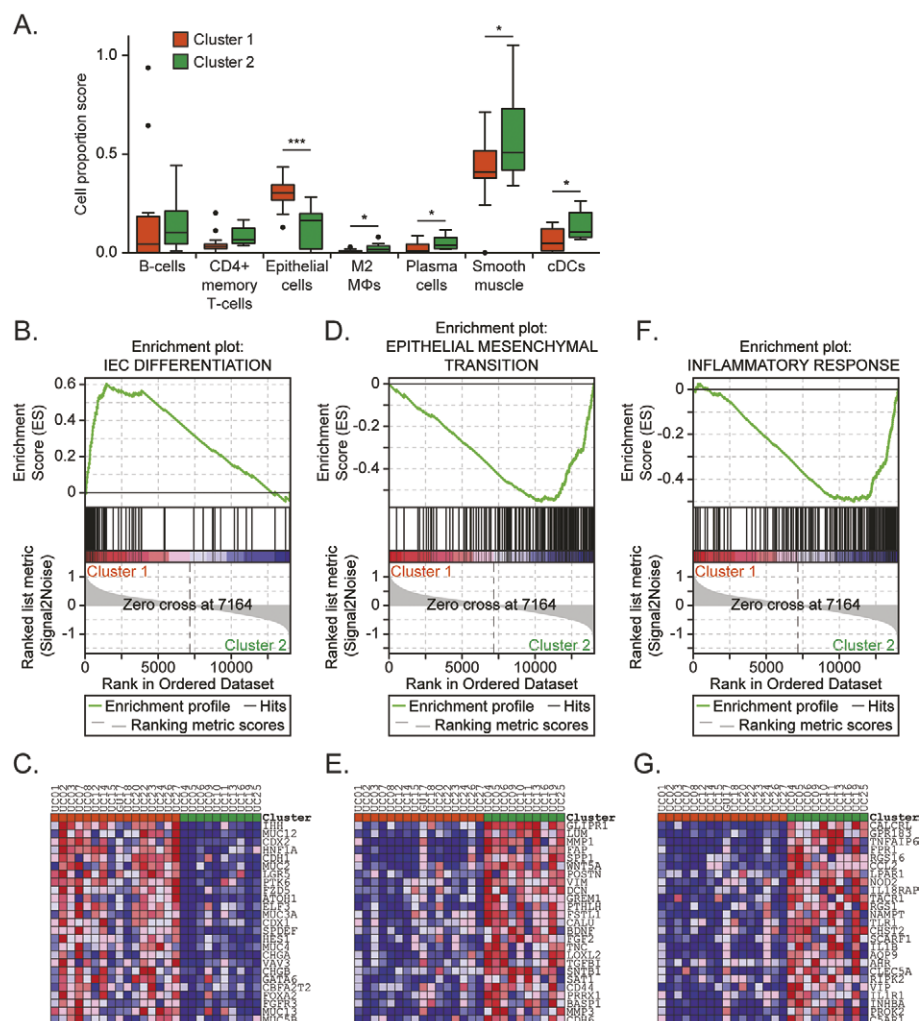


Fig. 2. The cell composition and molecular signatures of UC subsets are distinct. (A) RNA-seq deconvolution cell proportion scores of the indicated cell types in cluster 1 (n=16) and cluster 2 (n=10) individuals. (B) Enrichment of intestinal epithelial cell (IEC) differentiation genes by GSEA. (C) Heatmap of the top 25 core enrichment genes in the IEC differentiation gene set. (D) Enrichment of Hallmark epithelial-to-mesenchymal (EMT) genes by GSEA. (E) Heatmap of the top 25 core enrichment genes in the Hallmark EMT gene set. (F) Enrichment of Hallmark Immune Response genes by GSEA. (G) Heatmap of the top 25 core enrichment genes in the Hallmark Immune Response gene set (* $p < 0.05$, *** $p < 0.001$).

outcomes than those that undergo urgent or emergent intervention [11]. Clinical and genetic factors that predict who will undergo colectomy have been described [12]. However, our study is the first to implicate elevated expression of IEC differentiation genes, and in particular the *MUC* gene family, and a better prognosis in terms of prolonged time to surgery following initial diagnosis. Moreover, elevated expression of *MUC5B*, *MUC4*, and *MUC2* distinguished individuals that did not require early surgical intervention. Our findings suggest a model of UC pathogenesis, in which a subset of individuals is capable of sustaining or upregulating mucin gene expression. We propose that their enhanced mucosal barrier quells mucosal inflammation by prohibiting translocation of luminal antigens into the underlying tissue. Thus, these individuals present with less severe disease and have a prolonged period prior to requiring surgical intervention.

A compromised mucus barrier in severe UC is thought to be predominantly through loss of MUC2 protein expression

and differential post-translational modification of the MUC2 protein [31, 32]. These results, and work in genetic mouse models [33], demonstrate that MUC2 is protective against UC. Our results regarding elevated MUC2 and a prolonged tolerance to therapy prior to surgical intervention are in line with a protective role for this mucin. Likewise, decreased protein levels of MUC4 has also been reported in severe UC [31]. However, Muc4-deficient mice are less susceptible to experimental models of colitis [34]. As we found that higher MUC4 levels also correlate with prolonged time to surgery, we favor the role of MUC4 as protecting the colonic epithelium although additional work is needed to resolve disparate results from mouse and human tissues.

In our study, expression of *MUC5B* was the most strongly correlated with an extended time to surgery and little is known about the function of MUC5B in the colon. Elevated MUC5B expression has been reported in active, but not inactive, UC mucosal biopsies compared to control mucosal biopsies [35].

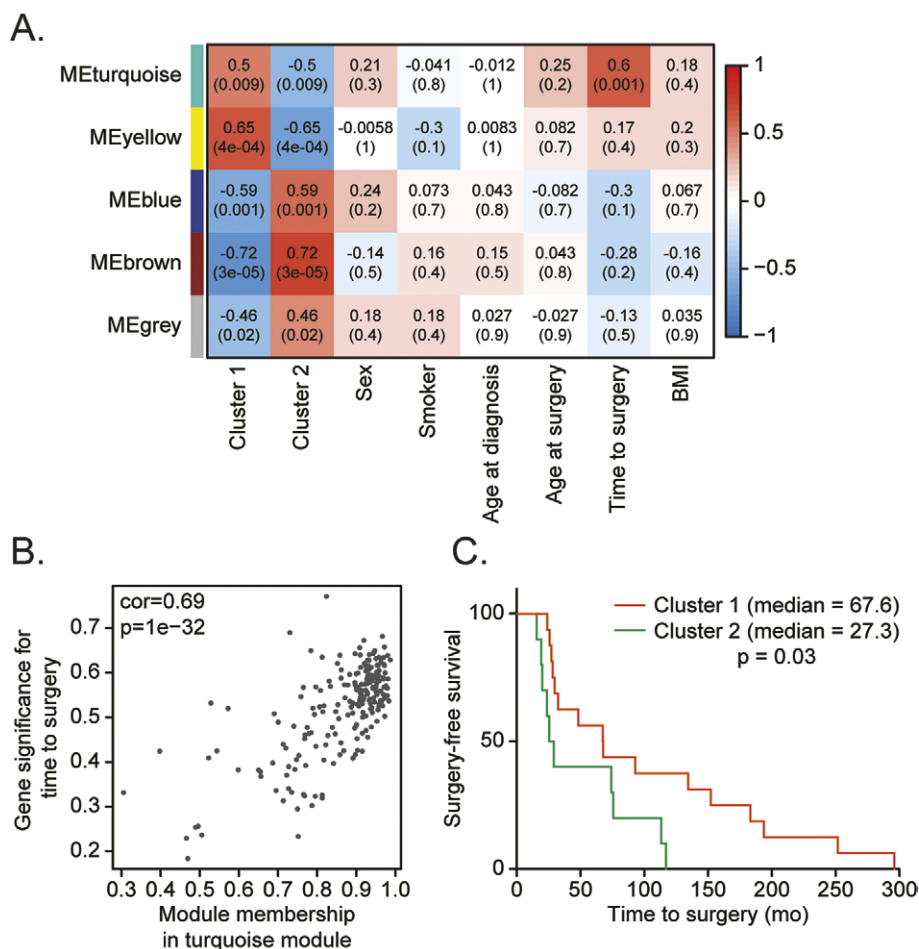


Fig. 3. Cluster 1 is associated with delayed surgical intervention. (A) Correlation between weighted gene co-expression network analysis (WGCNA) module eigengene (ME) expression and clinical traits. Each box includes the correlation coefficient, with the p-value in parentheses. (B) Association of each gene in the turquoise ME with time to surgery. (C) Kaplan-Meier curves for individuals in clusters 1 ($n=16$) and 2 ($n=10$). The log rank hazard ratio is 0.45 (95% CI: 0.14-0.88; $p=0.03$).

However, whether this increased expression is pathogenic or a protective mucosal response to inflammation was not addressed. Two studies showed that *MUC5B* was elevated in mucosal biopsies of treatment naïve, male UC patients and it was predominantly expressed towards the lower portions of the colonic crypts which harbor stem cell populations [6, 36]. Lahdaoui et al. [37] demonstrated that *MUC5B* promotes Wnt/ β -catenin signaling, epithelial cell proliferation and migration. Therefore, *MUC5B* could be targeted to protect epithelial stem cells and to improve restitution of the colonic epithelium, an overarching goal for next-generation IBD therapies [14, 32].

It will be important to determine how *MUC5B*, *MUC4*, and *MUC2* transcript levels are controlled during UC pathogenesis as their expression is induced by inflammatory cytokines [38-40]. We propose that inflammation induces sustained expression of mucins in some UC patients, leading to mucosal healing, a better responsiveness to therapeutic management, and delayed surgical intervention. Additionally, it is possible that medications, such as glucocorticoids and biologics could alter mucin expression. The similarities in medical management between cluster 1 and cluster 2 in our study suggest that treatment does not affect mucin expression.

However, our data does not preclude a relationship between specific therapeutics and *MUC5B*, *MUC4*, and *MUC2* levels. A longitudinal study on the effect of individual treatments on mucin expression could directly address this question.

Future prospective studies with a larger patient population are required to assess the role of *MUC5B*, *MUC4*, and *MUC2* as predictive biomarkers of disease severity and the role of these in UC pathogenesis. Additionally, prospective studies could provide evidence that expression levels have prognostic value for medical vs. surgical treatment. Importantly, for these types of studies, gene expression would have to be profiled from mucosal biopsies. Here, we focused on full-thickness surgical specimens, which is a limitation of our work. However, it is possible that the gene signature we found would be evident in superficial biopsies. Overall, this study highlights the applicability of transcriptomics to define clinically relevant, molecular subsets of IBD.

CONCLUSION

Our study identified two transcriptionally distinct clusters of UC patients that are distinguished by an IEC differentiation

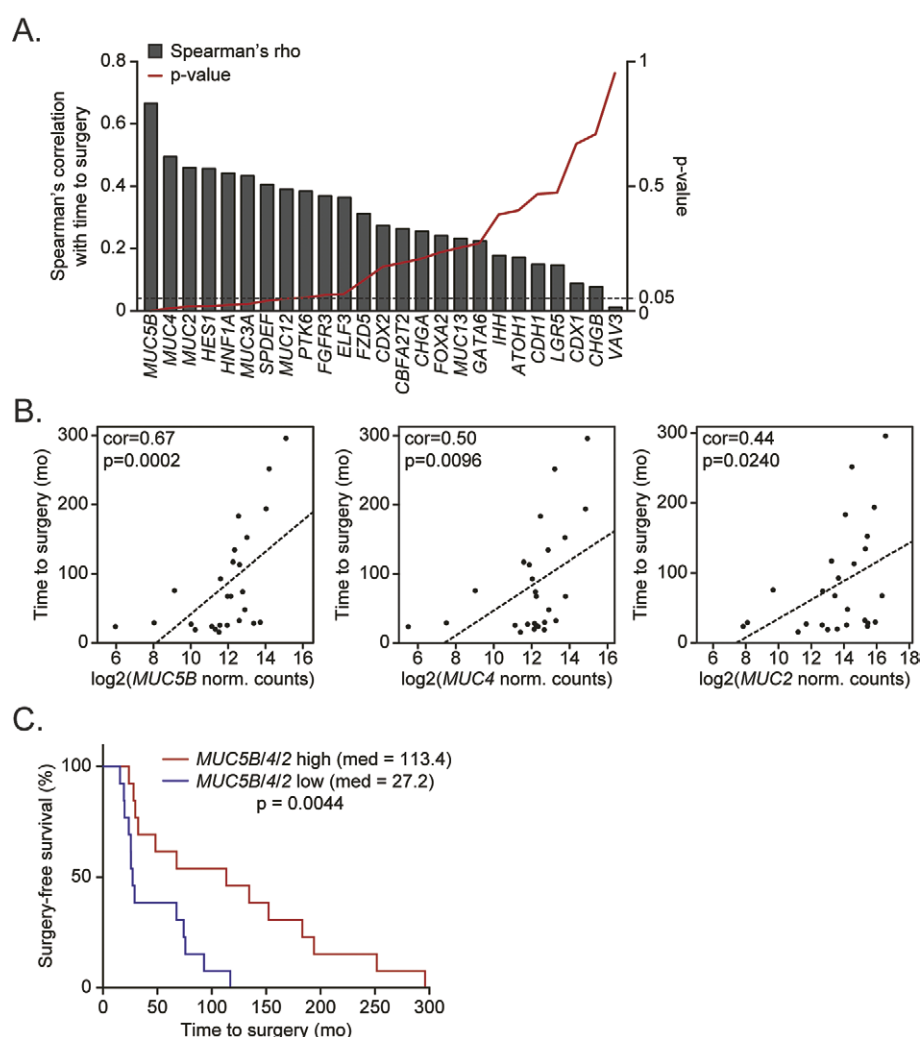


Fig. 4. Elevated MUC levels correlated with a delayed requirement for surgical intervention. (A) Spearman's correlation between gene expression and time to surgery for each of the top 25 differentially expressed genes from the IEC differentiation gene set. The p-value of the correlation is plotted on the right axis. (B) Correlation between MUC5B expression and time to surgery for each UC patient. (C) UC patients were divided into MUC low ($n=13$) and MUC high ($n=13$) expression groups, based on the median relative MUC expression for all UC patients. Surgery-free survival curves for individuals in these two groups. The log rank hazard ratio is 0.37 (95% CI: 0.11-0.61; $p=0.0044$).

gene signature. The subset with high expression of IEC genes had an extended time from diagnosis to colectomy. In particular, individuals that expressed abundant levels of mucins, *MUC5B*, *MUC4*, and *MUC2*, were protected from surgical intervention for a longer time than individuals with low mucin expression.

Conflicts of interest: No conflicts to declare.

Authors' contributions: M.A.E., K.M.S., W.A.K., G.S.Y.: study concept and design; M.A.E., N.A.J., K.M.S., B.P.K., M.M., S.D., L.H.: acquisition of data; M.A.E., N.A.J., K.M.S., B.P.K., M.M., W.A.K., G.S.Y.: analysis and interpretation of data; M.A.E., N.A.J., G.S.Y.: drafting of the manuscript; M.A.E., N.A.J., W.A.K., G.S.Y.: critical revision of the manuscript for important intellectual content; M.A.E., K.M.S.: statistical analysis; W.A.K., G.S.Y.: administrative, technical, or material support; W.A.K., G.S.Y.: study supervision. All authors approved the final version of the manuscript for publication.

Acknowledgments: We are indebted to patients who consented to have samples cataloged in the Colorectal Diseases Biobank at the Pennsylvania State University Hershey Medical Center. This research was supported by the Peter and Marsha Carlini Fund for Inflammatory Bowel Disease Research and by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR002014. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

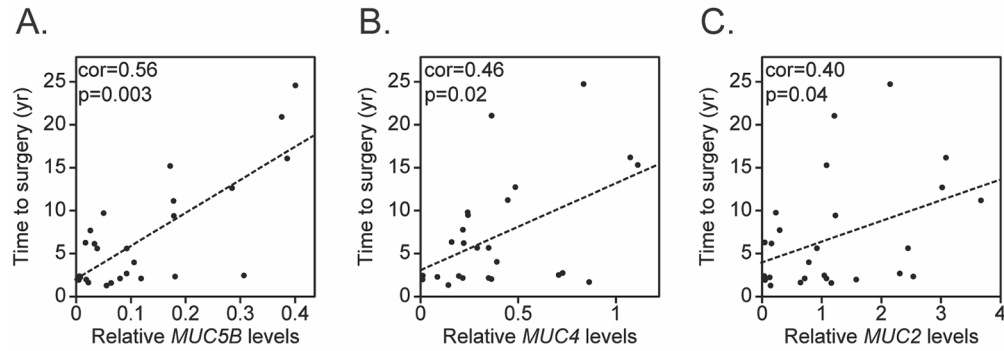
Supplementary material: To access the supplementary material visit the online version of the *J Gastrointest Liver Dis* at <http://dx.doi.org/10.15403/jgld-250>

REFERENCES

1. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *Lancet* 2017;389:1756-1770. doi:10.1016/S0140-6736(16)32126-2

2. McGovern DP, Kugathasan S, Cho JH. Genetics of Inflammatory Bowel Diseases. *Gastroenterology* 2015;149:1163-1176 e2. doi:[10.1053/j.gastro.2015.08.001](https://doi.org/10.1053/j.gastro.2015.08.001)
3. Yang Y, Jobin C. Novel insights into microbiome in colitis and colorectal cancer. *Curr Opin Gastroenterol* 2017;33:422-427. doi:[10.1097/MOG.0000000000000399](https://doi.org/10.1097/MOG.0000000000000399)
4. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol* 2017;13:e1005457. doi:[10.1371/journal.pcbi.1005457](https://doi.org/10.1371/journal.pcbi.1005457)
5. Dobre M, Milanesi E, Manuc TE, et al. M. Differential Intestinal Mucosa Transcriptomic Biomarkers for Crohn's Disease and Ulcerative Colitis. *J Immunol Res* 2018;2018:9208274. doi:[10.1155/2018/9208274](https://doi.org/10.1155/2018/9208274)
6. Taman H, Fenton CG, Hensel IV, Anderssen E, Florholmen J, Paulssen RH. Transcriptomic Landscape of Treatment-Naive Ulcerative Colitis. *J Crohns Colitis* 2018;12:327-336. doi:[10.1093/ecco-jcc/jjx139](https://doi.org/10.1093/ecco-jcc/jjx139)
7. Yang L, Tang S, Baker SS, et al. Difference in Pathomechanism Between Crohn's Disease and Ulcerative Colitis Revealed by Colon Transcriptome. *Inflamm Bowel Dis* 2019;25:722-731. doi:[10.1093/ibd/izy359](https://doi.org/10.1093/ibd/izy359)
8. Ferrari L, Krane MK, Fichera A. Inflammatory bowel disease surgery in the biologic era. *World J Gastrointest Surg* 2016;8:363-370. doi:[10.4240/wjgs.v8.i5.363](https://doi.org/10.4240/wjgs.v8.i5.363)
9. Windsor A, Michetti P, Bemelman W, Ghosh S. The positioning of colectomy in the treatment of ulcerative colitis in the era of biologic therapy. *Inflamm Bowel Dis*. 2013;19:2695-2703. doi:[10.1097/MIB.0b013e318292fae6](https://doi.org/10.1097/MIB.0b013e318292fae6)
10. Peyrin-Biroulet L, Germain A, Patel AS, Lindsay JO. Systematic review: outcomes and post-operative complications following colectomy for ulcerative colitis. *Aliment Pharmacol Ther* 2016;44:807-816. doi:[10.1111/apt.13763](https://doi.org/10.1111/apt.13763)
11. Roberts SE, Williams JG, Yeates D, Goldacre MJ. Mortality in patients with and without colectomy admitted to hospital for ulcerative colitis and Crohn's disease: record linkage studies. *BMJ* 2007;335:1033. doi:[10.1136/bmj.39345.714039.55](https://doi.org/10.1136/bmj.39345.714039.55)
12. Dias CC, Rodrigues PP, da Costa-Pereira A, Magro F. Clinical predictors of colectomy in patients with ulcerative colitis: systematic review and meta-analysis of cohort studies. *J Crohns Colitis* 2015;9:156-163. doi:[10.1093/ecco-jcc/jju016](https://doi.org/10.1093/ecco-jcc/jju016)
13. Solberg IC, Hoivik ML, Cvancarova M, Moum B; IBSEN Study Group. Risk matrix model for prediction of colectomy in a population-based study of ulcerative colitis patients (the IBSEN study). *Scand J Gastroenterol* 2015;50:1456-1462. doi:[10.3109/00365521.2015.1064991](https://doi.org/10.3109/00365521.2015.1064991)
14. Vancamelbeke M, Vermeire S. The intestinal barrier: a fundamental role in health and disease. *Expert Rev Gastroenterol Hepatol* 2017;11:821-834. doi:[10.1080/17474124.2017.1343143](https://doi.org/10.1080/17474124.2017.1343143)
15. Johansson ME, Hansson GC. Immunological aspects of intestinal mucus and mucins. *Nat Rev Immunol* 2016;16:639-649. doi:[10.1038/nri.2016.88](https://doi.org/10.1038/nri.2016.88)
16. Johansson ME, Larsson JM, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1:4659-4665. doi:[10.1073/pnas.1006451107](https://doi.org/10.1073/pnas.1006451107)
17. Shirazi T, Longman RJ, Corfield AP, Probert CS. Mucins and inflammatory bowel disease. *Postgrad Med J* 2000;76:473-478. doi:[10.1136/pmj.76.898.473](https://doi.org/10.1136/pmj.76.898.473)
18. Schieffer KM, Choi CS, Emrich S, et al. RNA-seq implicates deregulation of the immune system in the pathogenesis of diverticulitis. *Am J Physiol Gastrointest Liver Physiol* 2017;313:G277-G284. doi:[10.1152/ajpgi.00136.2017](https://doi.org/10.1152/ajpgi.00136.2017)
19. Schieffer KM, Kline BP, Harris LR, Deiling S, Koltun WA, Yochum GS. A differential host response to viral infection defines a subset of earlier-onset diverticulitis patients. *J Gastrointest Liver Dis* 2018;27:249-255. doi:[10.15403/jgld.2014.1121.273.sch](https://doi.org/10.15403/jgld.2014.1121.273.sch)
20. Afgan E, Baker D, van den Beek M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* 2016;44:W3-W10. doi:[10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343)
21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. doi:[10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8)
22. Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.10. 2018. Available at:<https://cran.r-project.org/package=pheatmap>
23. Sievert C. plotly for R. 2018. Available at:<https://plotly-book.cpsievert.me/>
24. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-15550. doi:[10.1073/pnas.0506580102](https://doi.org/10.1073/pnas.0506580102)
25. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 2017;18:220. doi:[10.1186/s13059-017-1349-1](https://doi.org/10.1186/s13059-017-1349-1)
26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559. doi:[10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559)
27. Rennoll SA, Eshelman MA, Raup-Konsavage WM, Kawasaki YI, Yochum GS. The MYC 3' Wnt-Responsive Element Drives Oncogenic MYC Expression in Human Colorectal Cancer Cells. *Cancers (Basel)* 2016;8:52. doi:[10.3390/cancers8050052](https://doi.org/10.3390/cancers8050052)
28. Kline BP, Schieffer KM, Choi CS, et al. Multifocal Versus Conventional Unifocal Diverticulitis: A Comparison of Clinical and Transcriptomic Characteristics. *Dig Dis Sci* 2019;64:3143-3151. doi:[10.1007/s10620-018-5403-y](https://doi.org/10.1007/s10620-018-5403-y)
29. Schieffer KM, Sabey K, Wright JR, et al. The Microbial Ecosystem Distinguishes Chronically Diseased Tissue from Adjacent Tissue in the Sigmoid Colon of Chronic, Recurrent Diverticulitis Patients. *Sci Rep* 2017;7:8467. doi:[10.1038/s41598-017-06787-8](https://doi.org/10.1038/s41598-017-06787-8)
30. Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006;55:749-753. doi:[10.1136/gut.2005.082909](https://doi.org/10.1136/gut.2005.082909)
31. Dorofeyev AE, Vasilenko IV, Rassokhina OA, Kondratiuk RB. Mucosal barrier in ulcerative colitis and Crohn's disease. *Gastroenterol Res Pract* 2013;2013:431231. doi:[10.1155/2013/431231](https://doi.org/10.1155/2013/431231)
32. Sun J, Shen X, Li Y, et al. Therapeutic Potential to Modify the Mucus Barrier in Inflammatory Bowel Disease. *Nutrients* 2016;8:44. doi:[10.3390/nu8010044](https://doi.org/10.3390/nu8010044)
33. Van der Sluis M, De Koning BA, De Bruijn AC, et al. Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* 2006;131:117-129. doi:[10.1053/j.gastro.2006.04.020](https://doi.org/10.1053/j.gastro.2006.04.020)
34. Das S, Rachagani S, Sheinin Y, et al. Mice deficient in Muc4 are resistant to experimental colitis and colitis-associated colorectal cancer. *Oncogene* 2016;35:2645-2654. doi:[10.1038/ncr.2015.327](https://doi.org/10.1038/ncr.2015.327)
35. Vancamelbeke M, Vanuytsel T, Farre R, et al. Genetic and Transcriptomic Bases of Intestinal Epithelial Barrier Dysfunction in Inflammatory Bowel Disease. *Inflamm Bowel Dis* 2017;23:1718-1729. doi:[10.1097/MIB.0000000000001246](https://doi.org/10.1097/MIB.0000000000001246)

36. Walsh MD, Clendenning M, Williamson E, et al. Expression of MUC2, MUC5AC, MUC5B, and MUC6 mucins in colorectal cancers and their association with the CpG island methylator phenotype. *Mod Pathol* 2013;26:1642-1656. doi:[10.1038/modpathol.2013.101](https://doi.org/10.1038/modpathol.2013.101)
37. Lahdaoui F, Messager M, Vincent A, et al. Depletion of MUC5B mucin in gastrointestinal cancer cells alters their tumorigenic properties: implication of the Wnt/beta-catenin pathway. *Biochem J* 2017;474:3733-3746. doi:[10.1042/BCJ20170348](https://doi.org/10.1042/BCJ20170348)
38. Enss ML, Cornberg M, Wagner S, et al. Proinflammatory cytokines trigger MUC gene expression and mucin release in the intestinal cancer cell line LS180. *Inflamm Res* 2000;49:162-169. doi:[10.1007/s000110050576](https://doi.org/10.1007/s000110050576)
39. Fujisawa T, Chang MM, Velichko S, et al. NF-kappaB mediates IL-1beta- and IL-17A-induced MUC5B expression in airway epithelial cells. *Am J Respir Cell Mol Biol* 2011;45:246-252. doi:[10.1165/rcmb.2009-0313OC](https://doi.org/10.1165/rcmb.2009-0313OC)
40. Mejias-Luque R, Peiro S, Vincent A, Van Seuningen I, de Bolos C. IL-6 induces MUC4 expression through gp130/STAT3 pathway in gastric cancer cell lines. *Biochim Biophys Acta* 2008;1783:1728-1736. doi:[10.1016/j.bbamcr.2008.05.020](https://doi.org/10.1016/j.bbamcr.2008.05.020)



SUPPLEMENTAL FIGURE 1 RT-qPCR confirmation of *MUC5B/4/2* levels in UC patients. (A) Correlation between *MUC5B* expression and time to surgery for each UC patient. (B) Correlation between *MUC4* expression and time to surgery for each UC patient. (C) Correlation between *MUC2* expression and time to surgery for each UC patient. Expression is normalized to the average of *GAPDH*, *PPIA*, and *TUBB*.