Utility of Novel Genomic Technologies for Biomarker Identification in Inflammatory Bowel Disease

Christina Yacoub
christina.yacoub@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/srhonors_theses

Recommended Citation
Yacoub, Christina, "Utility of Novel Genomic Technologies for Biomarker Identification in Inflammatory Bowel Disease" (2021). Honors Scholar Theses. 838.
https://opencommons.uconn.edu/srhonors_theses/838
Utility of Novel Genomic Technologies for Biomarker Identification in Inflammatory Bowel Disease

Christina Yacoub

University of Connecticut, School of Nursing

NURS 4597W: Senior Thesis in Nursing

Dr. Wendy A. Henderson

April 28, 2021
Abstract

The purpose of this study is to determine whether differences in protein expression exist between patients with inflammatory bowel disease (IBD), both Crohn’s Disease (CD) and Ulcerative Colitis (UC), irritable bowel syndrome (IBS), and healthy controls (HC). A total of fourteen colonic biopsies (n=14, 8-IBD, 4-IBS, 2-HC) underwent nucleus counts using the nCounter software of the Nanostring GeoMx Digital Spatial Profiler (NanoString Technologies, Inc., Seattle, Washington, USA). Three regions of interest were stained according to tryptase, crypt, and connective tissue and visualization markers were attached to fluoresce thirty inflammatory and oncological proteins of interest. After nucleus counts for proteins of interest were plotted in Tableau (2020.4.0), overexpression of AKT, beta-catenin, histone H3, CD44, S6, STAT3 were apparent for both IBD and IBS. The overexpressed proteins endorse mostly positive correlational relationships according to the bivariate correlation conducted in IBM SPSS Statistics (Version 27) predictive analysis software. Establishing validated levels of elevated protein expression offers the clinical opportunity to devise diagnostic biomarkers. Solidifying knowledge of the relationships between the inflammatory proteins provides potential understanding into the similarities of IBS and IBD.

Keywords: IBD, Crohn’s Disease, Ulcerative Colitis, IBS, protein expression, inflammation, oncologic activity, biomarkers

Background

The incidence of inflammatory bowel disease (IBD) has been increasing internationally with a total of 5.2 million people in North America and Europe currently diagnosed (Ananthakrishnan et al., 2020). IBD is an autoimmune illness encompassing Crohn’s Disease.
(CD) and Ulcerative Colitis (UC) which can vary from mild to severe based on symptoms and colon integrity. UC is generally superficial, continuous inflammation occurring in the submucosa of the large intestine whereas CD is transmural inflammation mixed between healthy intestines that can manifest anywhere in the gastrointestinal tract. The etiology is unknown but a genetic link has been observed and environmental factors such as smoking, microbiome diversity, oral contraceptives, antibiotics, nonsteroidal anti-inflammatory drugs (NSAIDS), and appendectomies have been studied with some correlational evidence (Ko et al., 2014).

More importantly, IBD can induce tissue damage and subsequently, through chronic inflammation, colon cancer (Beaugerie & Itzkowitz, 2015; Nebbia et al., 2020; Stidham & Higgins, 2018). Some studies have shown those with IBD are twice as likely to get colorectal carcinoma compared to healthy individuals (Beaugerie & Itzkowitz, 2015). According to the American Gastroenterology Association, endoscopy surveillance is recommended every 1-2 years for adults with IBD (Shah & Itzkowitz, 2020). Noninvasive tools such as the Crohn’s Disease Activity Index, Harvey-Bradshaw Index, and Partial Mayo Scoring Index score the severity of the disease based on symptomology subjectively reported by patients. However, these methods of scoring do not provide an accurate picture of what is occurring endoscopically or histologically making them less than ideal for tracking inflammation (Lewis et al., 2020). Additionally, serum biomarkers like C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR) lack sensitivity and do not correlate with inflammation of the bowel wall (Porter et al., 2020). Fecal calprotectin is able to identify colonic inflammation but does not show small intestine inflammation (Porter et al., 2020). This leaves colonoscopies or other forms of endoscopy to be the gold standard in diagnosing and monitoring IBD. It is important to note that for some patients, it is contraindicated to perform colonoscopies. Studies have shown patients
with IBD are at an 8-fold increased risk of colonoscopy-induced adverse events such as perforations (Mukewar et al., 2014). Colonoscopies are costly, and demanding procedures requiring a clear-liquid diet 24 hours before the procedure and a liquid prep to cleanse the colon. As a result, many patients with IBD are not compliant with surveillance colonoscopies despite being a high-risk population for developing colon cancer (Davis et al., 2013). Therefore, it is challenging for clinicians to determine when patients are in remission or in a flare noninvasively. More biomarkers are needed to adequately assess disease progression as well as response to pharmaceutical interventions. Inflammatory marker specificity between IBS and IBD is also needed to avoid unnecessary colonoscopies or misdiagnoses.

IBS is diagnosed based on fulfilling two of the categories on the Rome Criteria. Patients with abdominal pain once or more per day or week are screened for increased or decreased pain with defecation, changes in stool frequency, and changes in stool appearance. IBS-C is a decrease in the number of bowel movements with more solid stools as designated by the Bristol stool chart. IBS-D refers to an increase in frequency of bowel movements with more loose stools as designated by six or seven on the Bristol stool chart. Evidently, the symptoms of IBS and IBD can be similar. IBS is considered the most prevalent functional gastrointestinal disorder yet there is no biomarker or gold standard of diagnosing IBS (Lacy & Patel, 2017). One study showed 10% of IBD patients were misdiagnosed with IBS for several years (Card et al., 2014). Treatment for IBS does not address the inflammatory process thus a misdiagnosis prevents patients from getting adequate care. More biomarkers in the clinical setting would mean an increase in patient safety through a decrease in invasive procedures, a decrease in healthcare costs, and greater opportunity for accurate diagnoses.
Introduction

Quantitative protein expression has been a major prospect of cancer research and clinical biomarker development (McNamara et al., 2021; Mungenast et al., 2021; Kang et al., 2021). The Nanostring GeoMx DSP has allowed researchers to explore the tumor microenvironment and generate information regarding cancer development. Previous studies conducted by Galon et al. (2011) and Bindea et al. (2013), looked at the ratio of the markers CD3 and CD45RO, CD3 and CD8, or CD8 and CD45RO as colorectal cancer diagnostic measures. The protein ratios can be further evaluated using GeoMx to confirm significance in cancer progression and thus develop more specific cancer treatments (Mungenast et al., 2021). The proteins discussed in this paper include AKT, CD44, STAT3, beta-catenin, histone H3, and S6. AKT, being a protein kinase, phosphorylates other proteins that can either promote or inhibit certain cell activity. AKT regulates metabolism, proliferation, cell survival and growth, angiogenesis and the uptake of glucose into the cell. AKT is a central protein in signal transduction pathways and has been involved in susceptibility to colon cancers (Bateman et al., 2020). CD44 is a cell-receptor protein involved in cell adhesion and migration. CD44 plays a pivotal role in the signal transduction pathway resulting in activation of T-lymphocytes and inflammation (Bateman et al., 2020). Elevated levels of CD44 have been associated with certain cancers. STAT3 or signal transducer and activator of transcription 3 mediates cellular response to interleukins and growth factors. It also regulates the inflammatory response by regulating differentiation of CD4+ T-cells into T-Helper cells. High levels of STAT3 are associated with certain cancers. Beta-Catenin regulates cell adhesion as a downstream signal of canonical Wnt pathway. Increased activity of beta-catenin is linked to several cancers, including colorectal (Bateman et al., 2020). Histone H3 is a transcription regulator involved in DNA repair and post-translational modifications. S6 is part of
the small 40S ribosomal subunit which is responsible for cell growth and proliferation (Bateman et al., 2020). Understanding the cellular pathways of the proteins involved in the inflammatory and oncogenic process continues to be a focal point of targeted therapy for IBD and cancers.

**Methods**

Data published by Henderson et al. (2020) was used to compare protein expression among IBD, IBS and healthy control. IRB approval was not required since data is public and contains no information that could potentially identify participants. Formalin-fixed paraffin-embedded (FFPE) colonic samples were masked according to three regions of interest, one being tryptase, two is crypt and the third is connective tissue. The total sample size is fourteen, four colonic biopsies were taken from patients with IBS, eight from patients with IBD, and two were from healthy patients. All patients experienced moderate abdominal pain and were untreated prior to collection of biopsies. Nucleus counts were performed using Nanostring GeoMx Digital Spatial Profiler (DSP). GeoMx is done by attaching photocleavable oligonucleotides to the 30 inflammatory/oncological proteins’ antibodies and manually staining for the regions of interest as seen in figure 1. One protein was Rabbit IgG which was used as an internal negative control because there should be no expression of rabbit antibodies in human samples. Once regions of interest are stained/fluorescent, the ultraviolet light detaches the oligonucleotides which can then be placed onto a microtiter plate for the GeoMx DSP software to count. For this analysis the nuclei counts for the three regions of interest were utilized for comparison. GeoMx is a novel instrument that allows for the quantification of proteins without damaging the sample or requiring amplification. The nucleus counts were then separated according to region of interest and a bivariate correlation between proteins was conducted. Data were then visualized using Tableau software and proteins appearing overexpressed were run through the String database.
Figure 1

*Note.* Process of collecting nucleus counts indicated in the protein pathway in the image.


Results

The masked FFPE colonic samples for each patient are indicated in Figures 2 through 15. The nCounter software can quantify protein expression through nucleus counts. Graphs were created to visualize the data in Figures 16 through 18. The proteins with higher nucleus counts for each region were included in bivariate correlation Tables 1 through 3.

The results of the visualization conducted on Tableau indicate increased protein expression for CD44, beta-catenin, STAT3, S6 and histone H3 as seen in Figure 16. In Figure 17 and 18, AKT, beta-catenin, histone H3, STAT3, and S6 had marked increases compared to healthy control samples. Expression of VISTA was noted only in connective tissue and crypt for patients with IBD.

The bivariate correlation run through SPSS revealed significant correlation between most of the proteins. However, when considered with the data visualization, the bivariate correlation served as a confirmation of positive association between the proteins of interest. However, for STAT3, positive association was not significant. STAT3 also did not have any relationship with beta-catenin and histone when run through the String database. Beta-catenin, has experimentally, database curated and is coexpressed with AKT, All the other proteins have experimental interactions.
**Figure 2**

*FFPE Fluorescent IBS-5 Slides Masked with Visualization Markers with ROI Circled*

Note. **A.** The red tryptase visualization marker allows for distinguishing the tryptase rich environment for selection in the image directly below. **B.** The blue nuclei visualization marker highlights the crypt microenvironment for protein analysis in the region. **C.** PanCK green visualization marker displays the connective tissue environment for protein analysis.

**Figure 3**

*FFPE Fluorescent IBS-10 Slides Masked with Visualization Markers with ROI Circled*
Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

Figure 4

*FFPE Fluorescent IBS-4 Slides Masked with Visualization Markers with ROI Circled*

Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

Figure 5

*FFPE Fluorescent IBS-1 Slides Masked with Visualization Markers with ROI Circled*
Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

**Figure 6**

*FFPE Fluorescent IBD-6 Slides Masked with Visualization Markers with ROI Circled*

Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

**Figure 7**

*FFPE Fluorescent IBD-19 Slides Masked with Visualization Markers with ROI Circled*

Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

**Figure 8**

*FFPE Fluorescent IBD-18 Slides Masked with Visualization Markers with ROI Circled*
Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

Figure 9

*FFPE Fluorescent IBD-7 Slides Masked with Visualization Markers with ROI Circled*

Note. A. Tryptase ROI-1, B. Crypt-ROI-2, C. Connective-ROI-3

Figure 10

*FFPE Fluorescent IBD-3 Slides Masked with Visualization Markers with ROI Circled*
Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

Figure 11

*FFPE Fluorescent IBD-15 Slides Masked with Visualization Markers with ROI Circled*

![Image](image1.png)

Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

Figure 12

*FFPE Fluorescent IBD-12 Slides Masked with Visualization Markers with ROI Circled*

![Image](image2.png)

Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

Figure 13

*FFPE Fluorescent IBD-13 Slides Masked with Visualization Markers with ROI Circled*

![Image](image3.png)
Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

**Figure 14**

*FFPE Fluorescent Healthy Control-1 Slides Masked with Visualization Markers with ROI Circled*

Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3

**Figure 15**

*FFPE Fluorescent Healthy Control-2 Slides Masked with Visualization Markers with ROI Circled*
Note. A. Tryptase ROI-1, B. Crypt ROI-2, C. Connective ROI-3
**Note.** Protein nucleus count data for tryptase stained colonic biopsies were plotted in Tableau software. The x-axis is scaled independently for each patient with ranges of nucleus counts from 0-200. Increased expression of beta-catenin, CD44, histone H3, S6, and STAT3 is comparatively visible.
Figure 17

**Nucleus Count for Inflammatory and Oncologic Proteins in Crypt Masked Colonic Samples**

*Note.* Protein nucleus count data for crypt stained colonic biopsies are visualized. The x-axis is scaled independently for each patient with ranges of nucleus counts from 0-40. Increased expression for AKT, beta-catenin, histone H3, S6, and STAT3 is comparatively noted.
Figure 18

*Note.* Protein nucleus count data for connective tissue stained colonic biopsies were inputted into Tableau for visualization. The x-axis is scaled independently for each patient with ranges of nucleus counts from 0-40. Increased expression of AKT, beta-catenin, Histone H3, S6, and STAT3 is comparatively indicated.
<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Beta-Catenin</td>
<td>--</td>
<td>0.91**</td>
<td>0.91**</td>
<td>0.93**</td>
<td>0.65**</td>
</tr>
<tr>
<td>2. CD44</td>
<td>0.91**</td>
<td>--</td>
<td>0.80**</td>
<td>0.75**</td>
<td>0.58</td>
</tr>
<tr>
<td>3. Histone H3</td>
<td>0.91**</td>
<td>0.80**</td>
<td>--</td>
<td>0.93**</td>
<td>0.77**</td>
</tr>
<tr>
<td>4. S6</td>
<td>0.88**</td>
<td>0.75**</td>
<td>0.93**</td>
<td>--</td>
<td>0.81**</td>
</tr>
<tr>
<td>5. STAT3</td>
<td>0.65**</td>
<td>0.58</td>
<td>0.76**</td>
<td>0.81**</td>
<td>--</td>
</tr>
</tbody>
</table>

**Note.** The 31 proteins (n=31) were inputted into the bivariate correlation through IBM SPSS. The proteins with elevated expressions noted in Figure 16 from Tableau are listed in the correlational table. Significant correlation is observed between all the proteins except STAT3 and CD44, with an r value of 0.58 and p value=0.005.

**p<0.001**
Table 2

_Crypt Stained Colonic Samples’ Bivariate Correlation Between Proteins_

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AKT</td>
<td>--</td>
<td>0.815**</td>
<td>0.93**</td>
<td>0.921**</td>
<td>0.604</td>
</tr>
<tr>
<td>2. Beta-Catenin</td>
<td>0.815**</td>
<td>--</td>
<td>0.854**</td>
<td>0.872**</td>
<td>0.596</td>
</tr>
<tr>
<td>3. Histone H3</td>
<td>0.93**</td>
<td>0.854**</td>
<td>--</td>
<td>0.923**</td>
<td>0.669**</td>
</tr>
<tr>
<td>4. S6</td>
<td>0.921**</td>
<td>0.872**</td>
<td>0.923**</td>
<td>--</td>
<td>0.712**</td>
</tr>
<tr>
<td>5. STAT3</td>
<td>0.604</td>
<td>0.596</td>
<td>0.669**</td>
<td>0.712**</td>
<td>--</td>
</tr>
</tbody>
</table>

_Note._ The 31 proteins (n=31) were inputted into the bivariate correlation through SPSS. The proteins with elevated expressions noted in Figure 17 from Tableau are listed in the correlational table. STAT3 fails to have significant correlation between AKT and Beta-Catenin with p values of 0.022 and 0.025, respectively.

**p<0.001
Table 3

*Connective Tissue Masked Colonic Samples’ Bivariate Correlation Between Proteins*

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AKT</td>
<td>--</td>
<td>0.624</td>
<td>0.853**</td>
<td>0.747**</td>
<td>0.174</td>
</tr>
<tr>
<td>2. Beta-Catenin</td>
<td>0.624</td>
<td>--</td>
<td>0.645*</td>
<td>0.562</td>
<td>0.078</td>
</tr>
<tr>
<td>3. Histone H3</td>
<td>0.853**</td>
<td>0.645*</td>
<td>--</td>
<td>0.876**</td>
<td>0.228</td>
</tr>
<tr>
<td>4. S6</td>
<td>0.747**</td>
<td>0.562</td>
<td>0.876**</td>
<td>--</td>
<td>0.325</td>
</tr>
<tr>
<td>5. STAT3</td>
<td>0.174</td>
<td>0.078</td>
<td>0.228</td>
<td>0.325</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note.* The 31 proteins (n=31) were inputted into the bivariate correlation through SPSS. The proteins with elevated expressions noted in Figure 18 from Tableau are listed in the correlational table. A lack of correlation between STAT3 and AKT (p=0.589), beta-catenin (p=0.809), Histone H3 (p=0.476), and S6 (p=0.303) is evident through elevated p values. Beta-catenin shows insignificant correlation with AKT and S6 with p value equal to 0.3 and 0.57, respectively.

*p<0.05

**p<0.001
**Figure 19**

*Visualization of Inflammatory and Oncological Protein Relationships*

**Note.** A. Image includes 23 of the 30 inflammatory or oncogenic proteins run through the String database. B. The image on the right shows only the proteins that were overexpressed within the three regions of interest. For both images, the pink line shows the experimentally confirmed interactions, the black line indicates co-expression, and the blue line represents database curated interactions. The green line has no significance besides connecting the proteins. AKT and beta-catenin (CTNNB1) have the most established relationships.

**Discussion**

The proteins found to be highly expressed across the regions of interest are generally responsible for promoting cellular proliferation. Of note is the expression of VISTA seen only in those with IBD for crypt and connective tissue masked samples thus being an ideal candidate for future research. VISTA is an immune checkpoint protein that suppresses T-cell activation and subsequent cytokines (Lines et al., 2014). VISTA could be a potential mechanism to counteract
excessive inflammation and thus was not seen with IBS. Due to the fact that it is only expressed in certain microenvironments of patients with IBD, it could be used as a biomarker for diagnosis. GeoMx allows for quantification of proteins in specific colonic regions thus improving biomarker and cancer research. The efficacy of specific treatments through targeted pathway inhibition can be explored more easily. It is expected that inflammatory proteins would be elevated in colonic samples of patients with IBD. However, the patients with IBS also exhibited increased expression of the same proteins suggesting more research is needed to understand the pathogenesis of IBS. Moreover, the oncogenic activity of beta-catenin, STAT3, and CD44 provides potential noninvasive colon cancer surveillance. As more research is conducted, these protein markers can be used to monitor disease progression. Understanding expression of these proteins can also individualize pharmacological therapies for patients with IBD and determine efficacy of response to treatment. For instance, tofacitinib (Xeljanz), a JAK/STAT inhibitor, has evidence of efficacy in patients with ulcerative colitis when anti-tumor necrosis factor blockers are ineffective. If a patient has elevated STAT3, which was seen in the colonic samples of patients with IBD, it can be predicted that tofacitinib’s inhibitory effect would lead to a reduction in inflammation for these patients. Ultimately, knowing which inflammatory proteins are highly expressed, would indicate to a clinician which inflammatory pathway(s) is/are specifically activated and therefore prescribe treatment to target it.

**Limitations**

This research included a small sample size. Only 14 colonic samples were analyzed thus preventing generalizations and adequate statistical analysis. All proteins discussed in this study are known inflammatory or oncological proteins therefore a bias is present. The data were also
missing a positive internal control so expression seen could not be compared to a known colonic protein. Finally, GeoMx DSP is a novel instrument and therefore is not validated.

Conclusion

IBD is increasing internationally with no valid noninvasive methods to diagnose and monitor the disease. The purpose of this research was to compare protein expression between IBD, IBS and healthy control using GeoMx DSP to quantify nucleus counts. By researching protein expression, insight into pathophysiology can be established and thus improve diagnostics and treatment. The results revealed beta-catenin, CD44, histone H3, S6, STAT3 were overexpressed in ROI-1 with a strong positive correlation seen with the exception of CD44 with STAT3. For ROI-2 AKT, beta-catenin, histone H3, S6, STAT3 were overexpressed with a positive correlation except for STAT3 with beta-catenin and AKT. For ROI-3, AKT, beta-catenin, histone H3, S6, STAT3 were overexpressed but STAT3 had no correlation with any of the other proteins. In general, these proteins are involved in cellular growth and proliferation making them practical to observe in the future in relation to IBD. Individualizing therapies based on the guidance of biomarkers would increase quality of treatment and decrease the trial and error of bottom-up treatment. More biomarkers are needed to address the deficiency of diagnostic options, the steep cost, and dangers of endoscopy.

Acknowledgements

The data used in this project were provided by Wendy Henderson with explanation provided by Jeffery Robinson. The statistical analysis was guided by Yiming Zhang, a Statistics Graduate Assistant at the University of Connecticut.
References:


Robinson, Jeffrey; Gong, Jingjing; Liang, Yan; Ren, Xing; Turkington, Simon; Henderson, Wendy (2020), “Spatial segmented protein expression profiles from IBD and IBS-associated colonic biopsy FFPE sections using Nanostring GeoMx/DSP”, Mendeley Data, V1, doi: 10.17632/kytd2vkms.1
[https://doi.org/10.1093/ecco-jcc/jjaa040](https://doi.org/10.1093/ecco-jcc/jjaa040)
