# ORIGINAL RESEARCH—BASIC

# Proteomic Differences in Colonic Epithelial Cells in Ulcerative Colitis Have an Epigenetic Basis

Scott Jelinsky,[1] Isac Lee,[1] Mara Monetti,[2,†] Susanne Breitkopf,[2,‡] Flora Martz,[3,§] Ramya Kongala,[3] Jeffrey Culver,[2] Vanessa Vo,[2,¶] Liang Xue,[4] Richard Gieseck III,[1] Caitlyn Dickinson,[1] Marion Kasaian,[1] and James D. Lord[3]

[1]Department of Inflammation and Immunology, Pfizer, Cambridge, Massachusetts; [2]Internal Medicine Research Unit, Pfizer, Cambridge, Massachusetts; [3]Translational Research Program, Benaroya Research Institute, Seattle, Washington; and [4]Machine Learning and Computational Sciences, Early Clinical Development, Pfizer, Cambridge, Massachusetts

**BACKGROUND AND AIMS:** The colonic epithelium serves as both a barrier to lumenal contents and a gatekeeper of inflammatory responses. In ulcerative colitis (UC), epithelial dysfunction is a core feature, but little is known about the cellular changes that may underlie disease pathology. We therefore evaluated how the chromatin epigenetics and proteome of epithelial cells differs between health and UC. **METHODS:** We sorted live CD326+ epithelial cells from colon biopsies of healthy control (HC) screening colonoscopy recipients and from inflamed or uninflamed colon segments of UC patients on no biologic nor immunomodulator therapy (n = 5–7 subjects per group). Cell lysates were analyzed by proteomic evaluation and nuclei were analyzed for open chromatin with assay for transposase-accessible chromatin using sequencing. **RESULTS:** Proteins most highly elevated in inflamed UC biopsies relative to HC were those encoded by the HLA-DRA ($P = 3.1 \times 10^{-33}$) and CD74 ($P = 1.6 \times 10^{-27}$), genes associated with antigen presentation, and the antimicrobial dual oxidase 2 (DUOX2) ($P = 3.2 \times 10^{-28}$) and lipocalin-2 ($P = 2.2 \times 10^{-26}$) genes. Conversely, the water channel aquaporin 8 was strikingly less common with inflammation ($P = 1.9 \times 10^{-18}$). Assay for transposase-accessible chromatin using sequencing revealed more open chromatin around the aquaporin 8 gene in HCs ($P = 2.0 \times 10^{-2}$) and more around the DUOX2/DUOXA2 locus in inflamed UC colon ($P = 5.7 \times 10^{-4}$), suggesting an epigenetic basis for differential protein expression by epithelial cells in health and disease. **CONCLUSION:** Numerous differences exist between the proteome and chromatin of colonic epithelial cells in UC patients and HCs, some of which correlate to suggest specific epigenetic mechanisms regulating the epithelial proteome.

*Keywords:* Proteomics; ATAC-Seq; DUOX2; AQP8

## Introduction

The gut epithelium is a single layer of cells separating the host from the external environment, supported by an interdigitating network of junctional proteins, and interacting with the underlying stromal layer to maintain polarity that supports secretory function. The epithelium is a dynamic structure, with stem cells localized to the base of crypts continuously undergoing maturation and functional specialization to maintain the integrity of the barrier, secretion of mucus and antimicrobial peptides, release of serotonin and other bioactive amines, and communication with immune cells. In inflammatory bowel disease (IBD), an escalating cycle of epithelial disruption and inflammation contributes to tissue damage and epithelial erosion.[1] Agents targeting the inflammatory component are standard of care for IBD, but unmet need remains, with current treatments achieving remission in < 30% of patients vs placebo, and up to 40% of subjects showing partial response or loss of response over time.[2]

The preservation and restoration of the mucosal epithelium is a promising strategy for the development of novel IBD therapeutics,[3] but mechanistic understanding of epithelial function has been challenging. Recent technical advances have allowed the first single-cell transcriptional analysis of isolated gut epithelial cells,[4] but proteomic characterization has been more limited. In IBD, proteomics can inform on disease severity, progression, and treatment response, in relation to the degree of inflammation, tissue damage, and other correlates of histopathology.[5,6] At the

tissue level, proteomics has been instrumental in characterizing pathogenesis, providing evidence for disease-induced changes in mucus production,[7] extracellular matrix,[8,9] tissue repair,[10,11] inflammation,[5,10,12] autophagy,[8] energy metabolism,[9,13] unfolded protein response,[9] and oxidative stress.[13] Proteomic analysis targeted toward the epithelium may allow further characterization of the localization of response, and definition of the interplay between tissue damage and inflammation.

To date, the studies exploring protemics of enriched epithelium in IBD have all utilized surgical resections, rather than colonoscopic biopsies,[6,8,10] which necessarily restricts ulcerative colitis (UC) subjects to patients with either malignancy or exhaustive medical treatment exposure and failure as an indication for surgery. Additionally, and in contrast to colonoscopies, bowel resections cannot be ethically performed on truly healthy human controls for comparison, necessitating the use of heterogeneous controls with other intestinal disease as an indication for surgery. Furthermore, much existing proteomic data on the colonic epithelium in IBD used chemically and mechanically dissociated epithelium, which would necessarily be contaminated with intraepithelial leukocytes and often some lamina propria cells. This is especially confounding when comparing profiles from inflamed and uninflamed regions of the gut, as intestinal inflammation is associated with immune infiltrate and epithelial erosion,[14] making lymphoid contamination of the isolated epithelium even more likely. To avoid these caveats in the current study, we examine proteomics of sort-purified epithelial cells from colonoscopic biopsies obtained from healthy screening colonoscopy recipients, compared to those of the inflamed vs uninflamed colon segments of UC patients on no biologic or immunomodulator medication for their UC. Our findings reveal a unique profile of epithelial proteomic and epigenetic changes that differentiate the colonic epithelium of UC patients from that of healthy controls (HCs).

## Materials and Methods

### Subjects and Samples

Colon biopsies were obtained and cryopreserved from HC subjects undergoing screening colonoscopy, and from inflamed or uninflamed colon segments from UC patients on no biologic nor immunomodulator therapy (Table). Age of patients in each group was not statistically different ($P = .87$, $P = .31$ and $P = .38$ for healthy vs noninflamed, healthy vs inflamed and noninflamed vs inflamed respectively). All subjects provided written consent to participate as part of an instititional review board–approved biorepository program at the Benaroya Research Institute, in accordance with the Declaration of Helsinki.

### Epithelial Cell Purification

Biopsies were thawed and homogenized by vigorous vortexing for 20–30 minutes at 37 °C in the presence of 150 U/mL collagenase (Gibco) in Roswell Park Memorial Institute medium supplemented with 1 mM calcium and 1 mM magnesium. Homogenized cells were then passed through a 100 $\mu$m filter and stained with a viability dye (LIVE/DEAD, Life Technologies) fluorophor-conjugated anti-CD45 (clone 2D1 BioLegend) and anti-CD326 (clone 9C4 BioLegend). Live epithelial cells were purified by fluorescence-activated cell sorting on a FACSAria Fusion (BD Biosciences) for CD326$^+$, CD45$^-$ cells excluding viability dye. $1–2 \times 10^5$ sorted cells were pelleted and frozen for lysis and digestion for proteomic analysis, as below, while another $1.5 \times 10^5$ cells had their nuclei isolated, as below, for epigenetic analysis.

### Lysis, Digestion, and Tandem Mass Tag Labeling for Proteomic Analyses

Epithelial cells (CD45−/CD326+) (STUDY 1; Table) from 5 human healthy colon or UC inflamed (5 donors) and uninflamed (5 donors) biopsies were lysed in 4% SDS/0.1 M Tris, pH 8.5 and precipitated in the presence of acetone at −20 °C overnight. Protein was reconstituted in 8 M urea. A second, overlapping set of epithelial cells (CD45−/CD326+) (STUDY 2; Table) was lysed in 8 M urea directly.

For both sets of samples (STUDY 1 and STUDY 2), protein disulfide bonds were reduced in the presence of dithiothreitol (5 mM final concentration), followed by alkylation in the presence of 2-iodoacetamide (10 mM final concentration). The urea concentration was then reduced to 1.5 M using 0.05 M ammonium bicarbonate, followed by digestion with trypsin (Pierce Biotechnology; 1:50 w/w) at 37 °C overnight.

Peptide digest was desalted by C18 Sep-Pak cartridge (Waters, Milford, MA). Desalted peptides were reconstituted in 50 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid prior to quantification by NanoDrop (Thermo Fisher Scientific, Waltham, MA). Peptides were labeled with TMTPro 16plex reagent (Thermo Fisher Scientific) according to the manufacturer's instructions.

### Off-Line Basic Reversed-Phase (High pH) High-Pressure Liquid Chromatography Fractionation and Liquid Chromatography-Mass Spectrometry/MS

TMT-labeled peptides were fractionated by offline reversed-phase high-pH fractionation using an Agilent 1200 High-Pressure Liquid Chromatography (Santa Clara, CA) and separated on an Agilent ZORBAX Extend-C18 column into a total of 12 fractions. Fractions were reconstituted in 5% Acetonitrile in 0.1% Formic Acid prior to injection onto an EasyNano 1200 LC. Peptides were separated on an EasySpray column (50 cm, ES903; Thermo Fisher Scientific). The LC was coupled to a Thermo QExactive-HF mass spectrometer. A 'Top 10' data-dependent acquisition method was employed with the instrument set in positive detection mode. For MS1, the resolution was set to 60,000, the scan range was set to 350–1800 m/z, and the automatic gain control target was set to 5e5, with a maximum IT of 50 ms. For MS2, the resolution was set to 60,000, the normalized capillary electrophoresis was set to 38, the isolation window was set to 0.7 m/z, the automatic gain control target was set to 1e5, and the maximum IT was set to 105 ms.

**Table.** Patient Characteristics and Study Samples

| Subject/ biopsy | Age at draw | Gender | Race | Years of UC | Uceis score | UC medications | Patient type | Inflammation status | Anatomic location | Proteomics STUDY 1 | Proteomics STUDY 2 | ATAC-seq |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HC1 | 46 | Female | Caucasian | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | x | x | x |
| HC2 | 52 | Male | Declined | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | x | | |
| HC3 | 71 | Female | Caucasian | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | x | | x |
| HC4 | 61 | Female | Caucasian | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | x | x | x |
| HC5 | 71 | Female | Caucasian | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | x | x | |
| HC6 | 56 | Female | Caucasian | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | | x | x |
| HC7 | 27 | Female | Caucasian | n/a | n/a | n/a | Non-IBD | Uninflamed | Ascending colon | | x | x |
| UCNC1 | 49 | Male | Caucasian | 7.4 | 4 | Mesalamine | UC | Uninflamed | Ascending colon | x | x | x |
| UCNC2 | 47 | Male | Asian | 19.5 | 5 | None | UC | Uninflamed | Ascending colon | x | x | x |
| UCNC3 | 56 | Male | Caucasian | 37.9 | 0 | None | UC | Uninflamed | Ascending colon | x | x | x |
| UCNC4 | 78 | Female | Caucasian | 3.5 | 2 | None | UC | Uninflamed | Ascending colon | x | x | x |
| UCNC5 | 51 | Male | African American | 4 | 2 | Balsalazide | UC | Uninflamed | Transverse colon | x | | x |
| UCNC6 | 22 | Female | Asian | 3.6 | 4 | Mesalamine | UC | Uninflamed | Ascending colon | | x | x |
| UCIC1 | 25 | Male | Asian, Caucasian | 0 | 2 | None | UC | Inflamed | Ascending colon | x | x | x |
| UCIC2 | 55 | Male | Caucasian | 2.4 | 1 | Mesalamine | UC | Inflamed | Ascending colon | x | x | x |
| UCIC3 | 47 | Male | Asian | 19.5 | 3 | None | UC | Inflamed | Descending colon | x | x | x |
| UCIC4 | 35 | Female | Caucasian | 13.6 | 7 | Mesalamine, budesonide | UC | Inflamed | Descending colon | x | x | x |
| UCIC5 | 61 | Male | Declined | 2.7 | 5 | Mesalamine | UC | Inflamed | Colon | x | | x |
| UCIC6 | 27 | Male | Caucasian | 0.5 | 3 | None | UC | Inflamed | Ascending colon | | x | x |

UCIC, ulcerative colitis inflamed; UCNC, ulcerative colitis uninflamed.

## Proteomic Data Analysis and Normalization

All liquid chromatography-mass spectrometry/MS runs were analyzed using the Sequest algorithm (SEQUEST HT, Thermo Fisher Scientific) within Proteome Discoverer 2.4 (Thermo Fisher Scientific) against the UniProt human database. A 10 ppm MS1 error tolerance was used. Trypsin was set as the enzyme, allowing for 2 missed cleavages. A false discovery rate (FDR) level of 0.01 was used, with a reporter ion integration tolerance of 20 ppm for the most confident centroid in MS2. The MSstatsTMT workflow was applied for peptide quantification and normalization. Reporter abundances were based on signal intensities. Peptides were quantified by summing reporter ion intensity across all matching peptide-spectral, and then used to calculate protein abundance in each sample. The log2 peptide intensities are median normalized assuming equal input loading of all channels. Peptide intensities were summarized to protein intensities using Tukey's median polish algorithm.[15]

## Nucleus Isolation for Epigenetic Analyses

Cell suspensions were centrifuged at $300 \times g$ for 5 minutes at 4 °C. Pellets were resuspended in PBS + 0.04% BSA for 1000 cells/$\mu$L cell suspension. Suspensions were centrifuged again, and supernatants were mixed with equal parts chilled Lysis Buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% IGEPAL, 0.01% Digitonin, 1% BSA in nuclease-free water). Suspensions were gently mixed and incubated on ice for 3–5 minutes. Wash Buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 1% BSA, 0.1% Tween20 in nuclease-free water) was added without mixing. After centrifugation, chilled Diluted Nuclei Buffer (10X Genomics, Pleasanton, CA) was added to the pellet without mixing. After recentrifugation, nuclei pellets were resuspended in Freezing Buffer (40% glycerol, 5% Tris-HCl, 5 mM $MgCl_2$, 0.1 mM ethylenediaminetetraacetic acid in water) and stored at −80 °C until shipment to Active Motif for assay for transposase-accessible chromatin using sequencing (ATAC-seq).

## Epigenetic Analysis

Epigenetic analysis of open chromatin in purified nuclei was performed by Active Motif (Carlsbad, CA) using a standardized commercial ATAC-Seq protocol. Sequencing libraries were generated from tagmented DNA using standard library preparation protocols. Generated DNA libraries were purified using magnetic beads and amplified using polymerase chain reaction. After final polymerase chain reaction amplification, libraries were purified using Agencourt magnetic beads and assessed for quality by TapeStation and quantified with Qubit. Libraries were quantified with KAPA Library Quantification for Illumina platforms following the manufacturer's instructions. Sequencing was performed on an Illumina NextSeq system using the following parameters: (a) paired end 42 bp reads (PE42) with 80% bases higher than Q30, (b) minimum of 30 million reads per sample.

## Statistical Analysis

Protein expression differences were calculated using a limma R packages.[16] Meta-Analysis was performed using the metafor R package (2.0) and effect sizes were calculated using a fixed effects model. ATAC-seq reads were aligned to hg38 genome using Burrows-Wheeler aligner mem with default settings.[17] Peak calling was performed using MACS2,[18] and differentially accessible peaks were found using DESeq2.[19] Peaks were annotated based on the RefSeq database (hg38): peaks localized within ±2.5 kb of the transcription start site (TSS) of a protein-coding gene transcript were defined as promoter peaks with the gene of the nearest transcript as their target genes. All analyses were implemented in R version 3.6.2 and all visualizations were created using the ggplot2 R package version 3.3.6.

## Pathway Analysis

Hypergeometric testing was used to assess the enrichment of specific pathways in the meta-analyzed protein expression data. Input regulated proteins ($P < .05$) were tested against pathways derived from Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), REACTOME or the Molecular Signatures Database (MSigDB) hallmark gene set collection. Significant regulated pathways were visualized using the TMOD package in R.

All authors had access to the study data and had reviewed and approved the final manuscript.

# Results

## Differentially Expressed Proteins

Differentially expressed proteins in freshly isolated CD326+, CD45-epithelial cells from inflamed and uninflamed colon were identified using mass spectrometry. In Study 1, cells were lysed in SDS, and proteins precipitated overnight in acetone, then reconstituted in urea prior to trypsin digestion and mass spectrometry. A total of 2939 distinct proteins were detected with 50 proteins differentially regulated ($FDR < 0.05$) between inflamed and uninflamed UC samples (N = 5 each), 582 proteins differentially regulated between inflamed UC and HC tissue (N = 5), and 0 proteins differentially expressed between uninflamed UC and HC biopsies. To improve protein detection in Study 2, the protein isolation protocol was optimized by lysing the cell pellets directly in urea prior to trypsin digestion, and a repeat proteomic analysis was performed on a separate set of samples from a partially overlapping cohort of subjects (Table). In Study 2, a total of 3715 distinct proteins was identified. Of these, 376 proteins were differentially regulated between inflamed and uninflamed, 451 proteins were differentially regulated between inflamed and healthy tissue and 1 protein was differentially expressed between uninflamed and healthy. Principal component analysis of the proteome profiles showed that samples from inflamed regions formed distinguishable clusters, while samples from uninflamed and normal regions co-clustered (Figure 1A and B). The expression patterns from the inflamed tissue were distinct from the expression patterns of healthy tissue, while the expression patterns of the uninflamed tissue were more similar to the pattern from healthy tissue (Figure 1C and D). To increase our sensitivity and overcome the variability between the different studies, a meta-analysis was
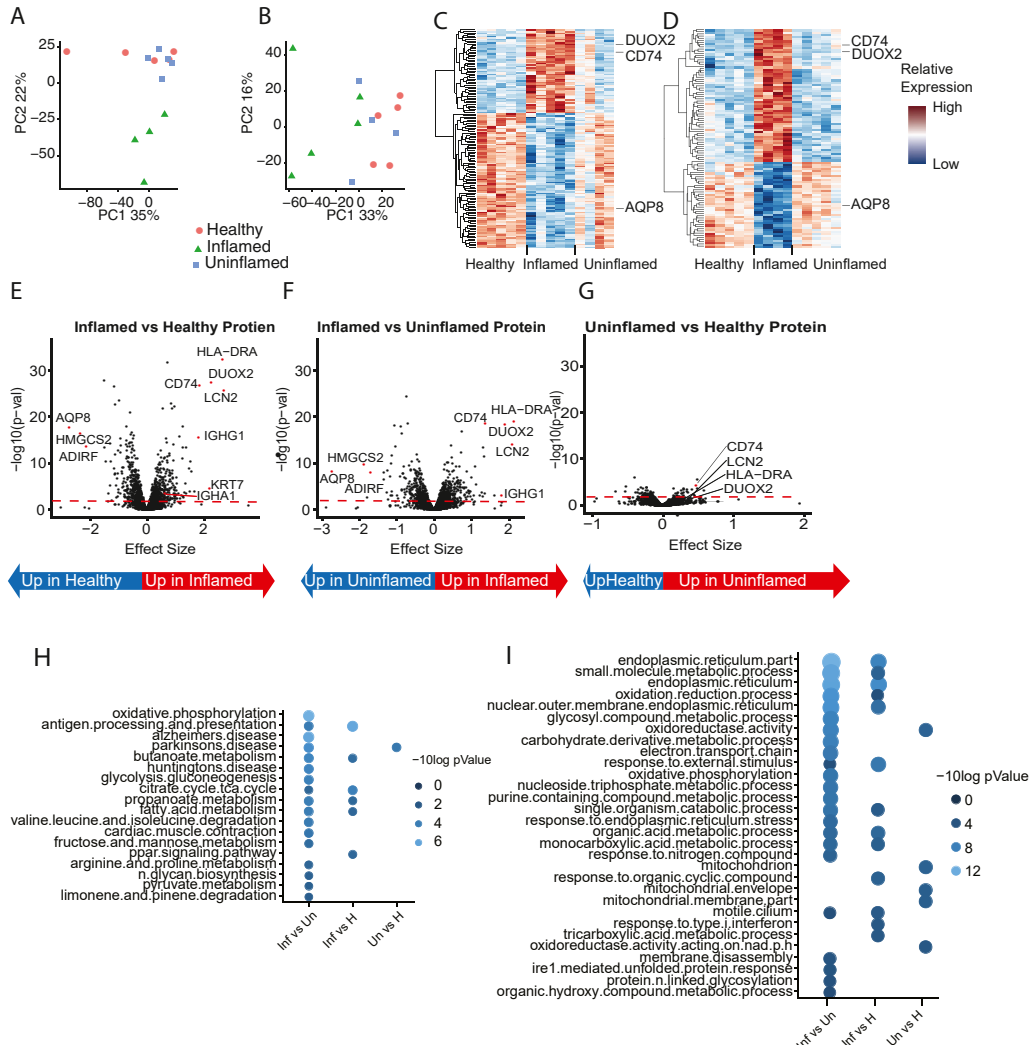
**Figure 1.** Protein expression from inflamed and uninflamed epithelial tissue from IBD patients. (A and B) Principal Component Analysis (PCA) plot of protein expression data for inflamed, uninflamed, and healthy epithelial tissue from 2 independent studies (study 1 (A) and study 2 (B)). Each point represents a sample, and the colors indicate sample groupings. The first 2 principal components are shown on the x- and y-axes, respectively. (C and D) Heatmap of protein expression levels for the top differentially expressed proteins between inflamed, uninflamed, and healthy subjects. Rows correspond to proteins, and columns correspond to individual samples. Protein expression levels are represented by a color gradient from low (blue) to high (red). Scatterplot of change in protein expression levels for all proteins based on meta-analysis between the 2 studies for (E) inflamed and healthy subjects (F), inflamed and uninflamed subjects, and (G) uninflamed and healthy subjects. Each dot represents a protein, and the x- and y-axes show estimate and −log10 P value, determined by a moderated t-statistic as implemented in limma packages,[16] respectively. Pathway analysis for protein expression. Dot plots of enrichment analysis for (H) Kegg pathways and (I) GO ontology. The color scale and dot size indicate the degree of enrichment based on hypergeometric test for proteins that are regulated (P < .05) in inflamed vs un inflamed (Inf vs Un), inflamed vs healthy (Inf vs H) and uninflamed vs healthy (Un vs H).

performed on the proteins that were detected in both studies. An important benefit of meta-analysis is its capacity to merge data from multiple studies, leading to increased statistical power and improved precision in estimating the effect size. Consequently, this enhanced statistical power enables the detection of a greater number of differentially regulated proteins compared to what could be achieved through individual studies alone. Of the 2613 distinct proteins, 656 were differentially identified between inflamed and uninflamed UC, 791 between inflamed UC, and HC, and

3 between HC and uninflamed UC biopsies (Figure 1E–G and Table A1).

Enrichment analysis utilizing KEGG pathways revealed that the proteome of inflamed UC epithelial cells exhibited significant enrichment in pathways related to oxidative phosphorylation, antigen processing and presentation, and several pathways associated with energy metabolism. Furthermore, when GO pathways were employed for enrichment analysis, oxidation/reduction processes, oxidoreductase activity, and oxidative phosphorylation were

identified as prominent terms, along with various cellular metabolic pathways. Notably, several pathways that were found to be enriched in the proteomes of inflamed vs uninflamed cells were also enriched in the proteomes of inflamed vs healthy cells, as shown in Figure 1H and I.

## Differentially Accessible Chromatin

To investigate a potential epigenetic cause for differential protein expression in diseased vs healthy epithelium, differences in accessible chromatin between epithelial cell sources were evaluated by ATAC-seq. Across all samples, we observed a total of 54,904 accessibility peaks, with 43,413 peaks from HCs, 39,964 from uninflamed UC, and 43,929 from inflamed UC. To infer gene activity from chromatin accessibility, we calculated the promoter chromatin accessibility for each gene, as the average signal in peaks within $\pm 2.5$ kb of the TSS. Principal component analysis was conducted on the promoter activities, and the results indicated that, similar to protein expression, the inflamed UC samples formed a distinct cluster on PC1, while HC and uninflamed UC samples clustered together (Figure 2A). In addition, the peaks that were more open in inflamed tissue were more closed in the healthy and uninflamed tissue (Figure 2B).

We identified 411 genes with differentially accessible promoters ($FDR < 0.05$) between inflamed UC and HC biopsies, 23 genes between inflamed and uninflamed UC, and 1 gene between uninflamed UC and HC samples (Figure 2C–E and Table A1).

Additionally, we observed a specific open chromatin locus at chr15:45,114,200 near the dual oxidase 2 (DUOX2) and DUOXA2 genes. This region was open in epithelial cells from all inflamed UC biopsies and 2 of 6 uninflamed UC biopsies, but not in any from HCs. This suggests that this region may contain a crucial regulatory element driving the expression of DUOX2 and DUOXA2 genes, and hence production of $H_2O_2$ reactive oxygen species, in UC.

## Correlations Between Proteomic and Epigenetic Differences

To determine whether open chromatin regions correlate with disease-associated protein expression, we integrated our ATAC-seq data with our proteomic data on epithelial
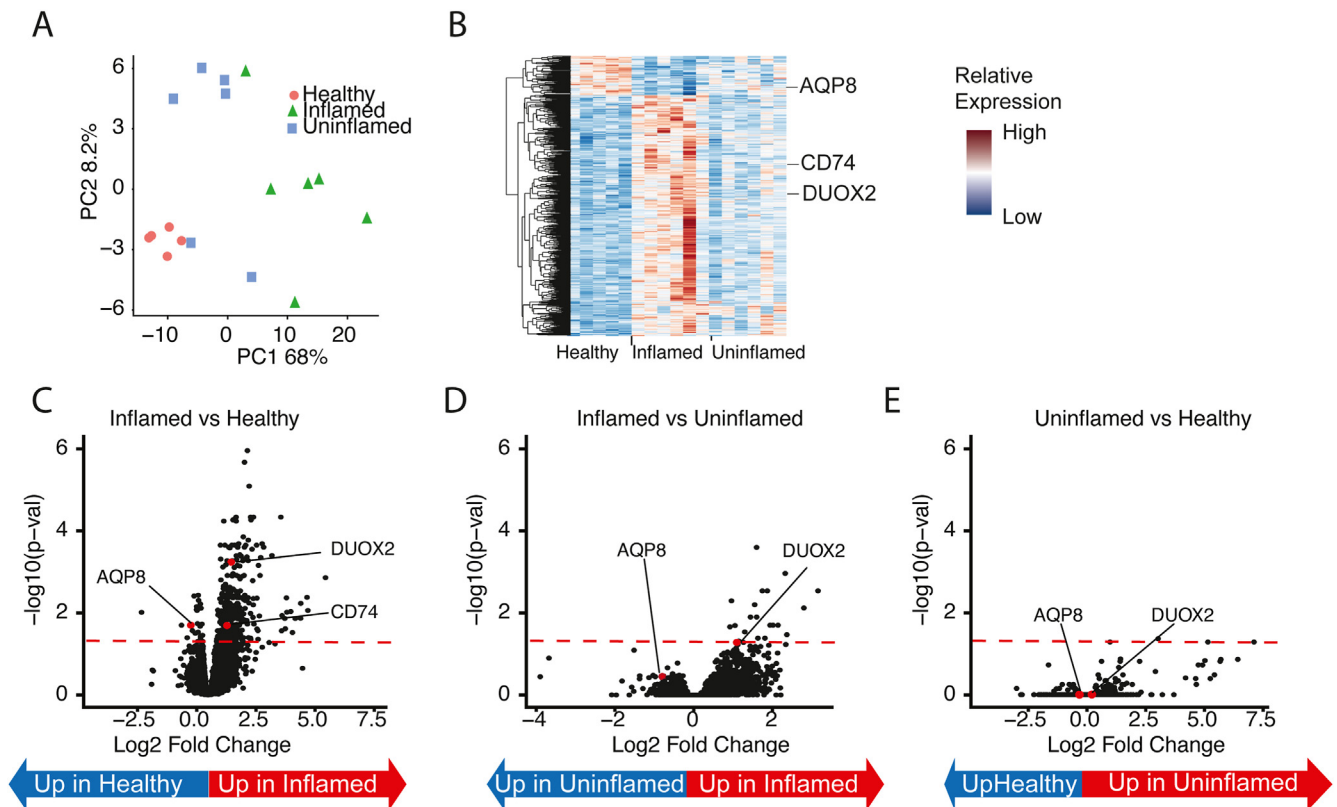


**Figure 2.** ATACseq profiles of inflamed and uninflamed epithelia from IBD patients. (A) Principal Component Analysis (PCA) plot of ATAC-seq peaks in inflamed, uninflamed, healthy epithelial tissue. Each point represents a sample, and the colors indicate sample groupings. The first 2 principal components are shown on the x- and y-axes, respectively. (B) Heatmap of ATAC-seq peaks for the top differentially accessible peaks between inflamed, uninflamed, and healthy subjects. Rows correspond to peaks, and columns correspond to individual samples. Red indicates a more accessible peak and blue indicates a less accessible peak. (CDE)Volcano plot showing change in peaks accessibility for all detected peaks between (C) inflamed and healthy subjects (D), inflamed and uninflamed subjects and (E) uninflamed and healthy subjects. Each dot represents a peak, and the x- and y-axes show the difference in peak accessibility and the negative logarithm of the P value as determined by the Wald test as implemented in the r packages DESeq2[18,19] respectively.
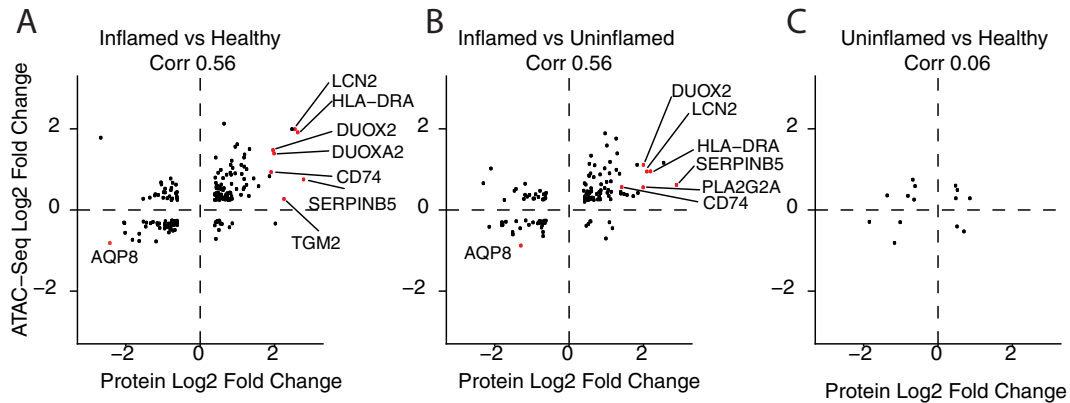
**Figure 3.** Correlation of protein expression and ATAC-seq peaks in epithelia of ID patients. Scatterplot of protein expression difference from study 2 and ATAC-seq peak accessibility differences in: (A) inflamed vs healthy, (B) inflamed vs uninflamed, and (C) uninflamed vs healthy epithelial tissue from IBD patients. Each dot represents a distinct protein. The x-axis shows the change in protein expression levels, while the y-axis shows the change in ATAC-seq peak accessibility. Correlations were determined by Pearson correlation coefficient method.

cells sorted from the same biopsies. Out of the 376 differentially expressed proteins between inflamed and healthy in Study 2, we identified 361 proteins with open chromatin peaks upstream of their protomer, 28 of which had statistically different peak sizes. For proteins that were differentially expressed and had open chromatin peaks within their promotor region (2.5 kb upstream and downstream of TSS) we calculated the correlation between differences in promoter accessibility and differences in protein expression between sample types. This correlation was strong between inflamed UC and HCs (Figure 3A, $R^2 = 0.565$ (Spearman)), and between inflamed and uninflamed UC (Figure 3B, $R^2 = 0.558$ (Spearman)), indicating that proteins with increased expression in inflamed tissue tended to have an increased amount of open chromatin in the promoter for their gene, while proteins with decreased expression in inflamed tissues tended to have a decreased amount of open chromatin. We did not observe any correlation in the uninflamed UC vs HC comparison due to a few differentially expressed proteins (Figure 3C).

Among the proteins differentially expressed between inflamed and uninflamed tissue, DUOX2 and aquaporin 8 (AQP8) also showed some of the highest epigenetic differences and had strong peak signals (Figure 4). A significant increase in DUOX2 protein and decrease in AQP8 protein in inflamed tissue relative to uninflamed correlated with increased accessibility in the DUOX2 promoter and decreased accessibility in the AQP8 promoter in inflamed tissue by ATACseq (Figure 4).

Additional proteins that were significantly elevated in inflamed UC as compared to uninflamed UC or biopsies from healthy subjects included several involved in inflammation or antigen presentation. Lipocalin 2, an antimicrobial protein of the innate immune system encoded by the lipocalin-2 (LCN2) gene, was increased in epithelial cells from inflamed biopsies ($P = 2.2 \times 10^{-26}$ vs HC, $P = 8.7 \times 10^{-15}$ vs uninflamed UC; Table A1) which showed correspondingly more open chromatin near the LCN2 promoter ($P = 6.6 \times 10^{-7}$) (Figure 5A).

Similarly, for the antigen-presenting major histocompatibility complex (MHC) class II protein HLA-DRA (Figure 5B) and its chaperone, CD74 (Figure 5C), protein levels were significantly higher in inflamed than uninflamed UC biopsies ($P = 8.2 \times 10^{-20}$ for HLA-DRA, and $P = 2.3 \times 10^{-19}$ for CD74) or HC biopsies ($P = 3.1 \times 10^{-33}$ for HLA-DRA, and $P = 1.6 \times 10^{-27}$ for CD74; Table A1). In both cases, promoter accessibility for these genes was increased in accordance with the higher protein expression in the inflamed epithelium ($P = 1.4 \times 10^{-4}$ for HLA-DRA, and $P = 2.8 \times 10^{-4}$ for CD74) (Figure 5B and C). Levels of other proteins involved in antigen processing and presentation, such as the proteasome protein encoded by the proteasome subunit beta type-9 (PSMB9) gene and the MHC class I peptide transporters transporter associated with antigen processing (TAP-1 and TAP-2), were also significantly higher in inflamed UC than HC epithelium ($P = 6.2 \times 10^{-23}$ for PSMB9, $P = 2.0 \times 10^{-6}$ for TAP1, and $P = 5.6 \times 10^{-4}$ for TAP2; Figure 5D–F and Table A1). However, in contrast to LCN2, HLA-DRA, and CD74, there was no correlation between inflammation and the accessibility of chromatin near the PSMB9 or TAP1 and TAP2 genes (Figure 5D–F).

In the absence of inflammation, differences in protein expression between uninflamed UC and HC epithelial cells were much more modest. However, IgA was the most differentially associated with uninflamed UC biopsies relative to HCs (IGHA1; $P = 5.2 \times 10^{-4}$; Table A1). Likewise, IgG1 was among the proteins most strongly associated with inflamed UC biopsies compared to HCs (IGHG1; $P = 2.0 \times 10^{-16}$; Table A1). Unsurprisingly, neither IGHA1 nor IGHG1 was detected in the chromatin analysis by ATAC-seq, as antibodies are exclusively made by B cells, not epithelial cells.

## Determination of Cell Type-Specific Protein Expression

We aimed to determine the cell type-selective expression of the proteins we identified as regulated. To accomplish this,
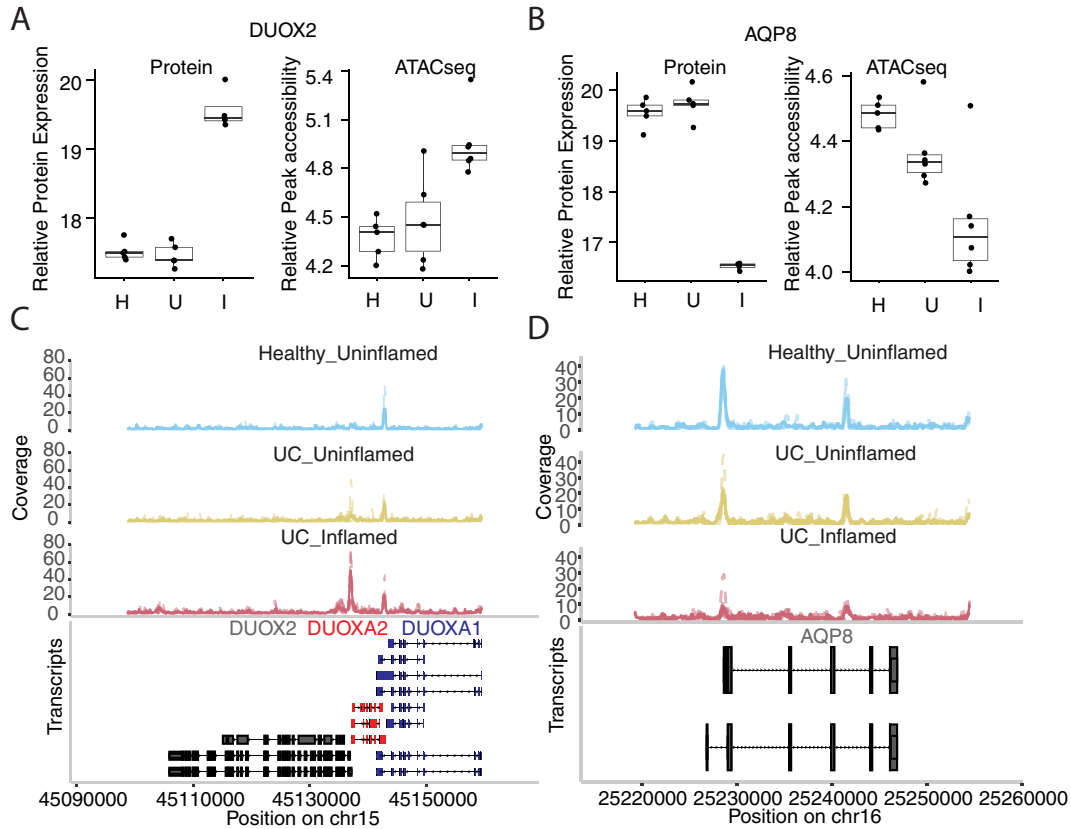
**Figure 4.** Chromatin accessibility and protein expression for DUOX2 and AQP8. (A and B) Protein expression and ATAC-seq peak accessibility for DUOX2 and AQP8 in inflamed and uninflamed epithelial from IBD subjects. (C and D) Genome browser view of DUOX2 and AQP8 showing open chromatin regions align with promoter start sites for DUOX2 and AQP8. Genome browser views consist of a combination of multiple individual samples.

we utilized a publicly available single-cell RNA sequencing dataset.[4] This data confirmed that the expression of DUOX2 mRNA is restricted to epithelial populations from inflamed sites (Figure 6) with significantly reduced expression in noninflamed sites and very low expression in healthy sites. In contrast, and consistent with the proteomics data, AQP8 demonstrated high mRNA expression in healthy and non-inflamed sites (Figure 6). For both DUOX2 and AQP8, mRNA expression was primarily seen in mature enterocytes, suggesting that these are the predominant epithelial cell type we purified, although AQP8 mRNA is also expressed fairly well by immature enterocyte populations, raising the possibility that a greater fraction of these immature cells could exist in our uninflamed biopsies to contribute to them having more AQP8 and less DUOX2 protein in our data. There was little to no expression of either protein's mRNA in immune cells or stromal cells (data not shown) to suggest such cells could be contaminating our data.

## Discussion

Our unique epithelial-specific dataset addressed changes in protein expression that distinguish healthy donor colonic epithelium and the uninflamed and inflamed colonic epithelia of UC patients. These proteomic changes may underlie pathogenic changes in UC, encompassing epithelial erosion and depletion of the mucus barrier, leading to bacterial translocation, inflammation, and tissue damage. Along with this, epithelial-based repair mechanisms including antimicrobial responses and antigen presentation may distinguish inflamed and uninflamed sites.

Our analysis showed that the proteomic profile of UC inflamed biopsies segregated from those of all uninflamed biopsies, while the uninflamed profiles from UC patients clustered with proteomics of biopsies from healthy subjects. Transcriptomic analysis of IBD colonic epithelium isolated from biopsies by laser capture microdissection has shown a similar clustering pattern, with uninflamed profiles over-lapping with the healthy transcriptome.[20] In contrast, we found several hundred proteins were differentially expressed in inflamed UC biopsies as compared to biopsies from HCs. Those most increased with inflammation tended to be involved with oxidative stress, antimicrobial responses, and antigen presentation. We explored the epigenetic basis of key proteomic changes and found that open chromatin near the promoters of some but not all of the genes encoding these proteins correlates with their expression.

The water channel protein AQP8 was significantly higher in uninflamed than inflamed colonic epithelial cells. In the small and large intestines, AQP8 is localized to the apical
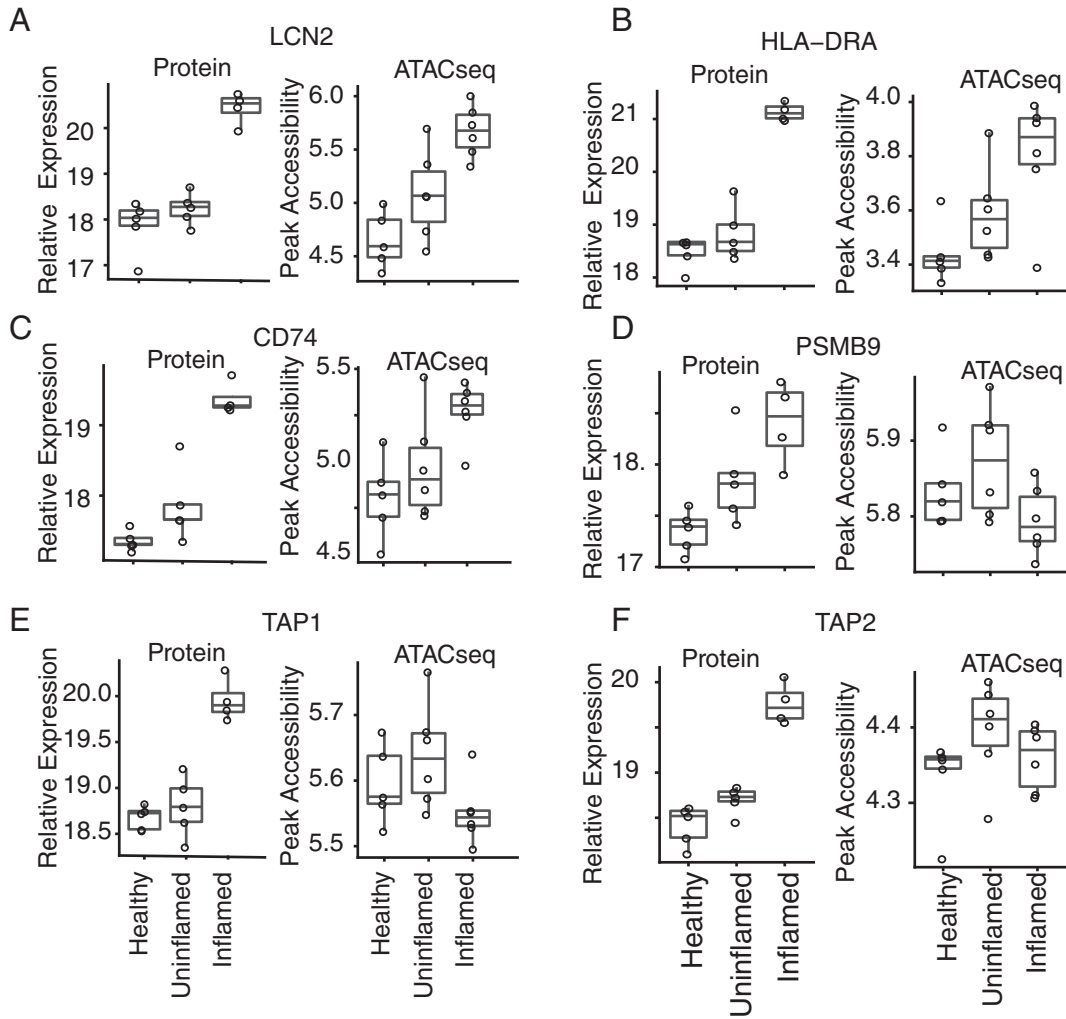
**Figure 5.** Comparison of protein expression and ATAC-peak accessibility. Expression of protein and ATAC-seq peak accessibility in inflamed, uninflamed, and healthy epithelia tissue from the gut for (A) LCN2, (B) HLA_DRA, (C) CD74, (D) PSMB9, (E) TAP1 and (F) TAP2.
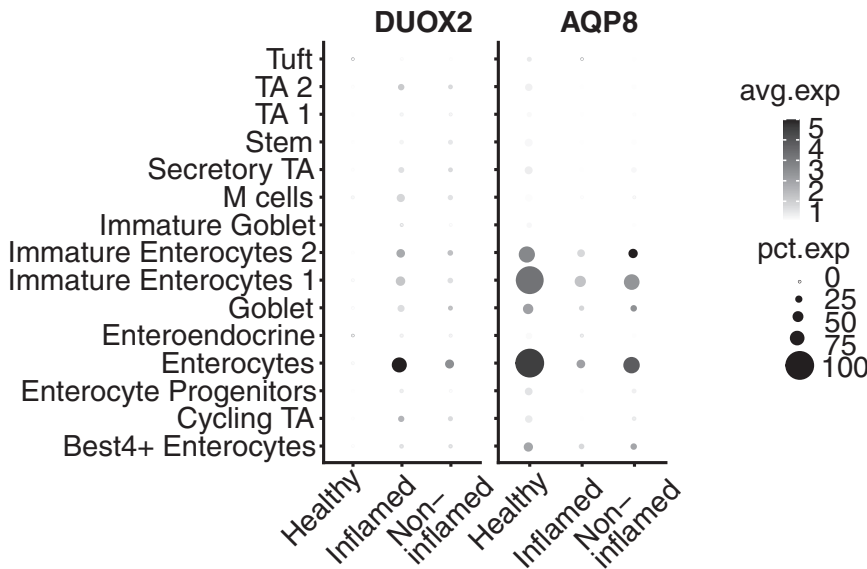


**Figure 6.** Single-cell expression of DUOX2 from IBD subjects. Single-cell expression from epithelial cell compartment derived from Smillie et al.[4] Dot intensity represents the average expression level, and the size of dots represents the percentage of cells expressing the gene of interest.

surface of mucosal epithelial cells,[21] and allows bidirectional transport of water and $H_2O_2$.[22] Reduced AQP8 expression has been linked with collagenous colitis[23] and bile acid diarrhea,[24] and has been noted in active UC,[21,25–27] Crohn's disease (CD),[26,27] and mouse models of colitis.[26,27] Under conditions of hypoxia or ER stress, the activity of AQP8 to transport $H_2O_2$ and water is impaired,[28] and downregulation of AQP8 expression has been suggested as a defense mechanism against severe oxidative stress.[27]

Consistent with the oxidative stress response, DUOX proteins were among the most strongly and significantly elevated proteins in inflamed colonic epithelial cells of UC. DUOX2 is a hydrogen peroxide-generating enzyme involved in antimicrobial defense at the mucosal surface. The $Ca^{2+}$-dependent nicotinamide adenine dinucleotide phosphate, reduced form oxidase encoded by DUOX2, along with its maturation partner encoded by DUOXA2, mediates the production of epithelial $H_2O_2$.[29] We found the products of both DUOX2 and DUOXA2 genes to be more common in inflamed colon epithelial cells, consistent with single cell transcriptomic studies revealing enhanced expression of DUOX2 and DUOXA2 in inflamed epithelial cells of UC.[4,30] DUOX2/DUOXA2 is localized to epithelial crypts on the apical surface,[31] and is known to be upregulated in UC,[32] CD,[30] very early onset IBD,[33] and mouse models of IBD.[34] Under homeostatic conditions, epithelial $H_2O_2$ plays a protective role, to drive innate immune responses and maintain antimicrobial immunity.[29,33] Under conditions of disease, excessive generation of reactive oxygen species from DUOX2 activation can promote tissue inflammation.[31,33,35]

The major strength of our study is a parallel whole-genome survey of open chromatin performed by ATAC-seq, in conjunction with a proteomic analysis of epithelial cells purified from inflamed vs uninflamed colonic biopsies from UC patients vs HCs. This analysis revealed an epigenetic basis for some of their most significant differences in protein expression, including AQP8 and dual oxidase 2 (DUOX2 and DUOXA2). The hypoxia-inducible transcription factor, HIF-1$\alpha$, is a modulator of AQP8 in hepatocytes,[36] and was one of several transcription factor binding motifs identified in epithelial cells (data not shown). Gut epithelial expression of DUOX2 is induced by gut microbiota, downstream of inflammatory signals including NF-kB, MyD88, and p38.[37]

Gut microbiota also stimulates release of lipocalin 2, a validated fecal biomarker of intestinal inflammation that plays a key antimicrobial function in the gut.[38] Lipocalin 2 was highly increased in the proteome of inflamed biopsies, consistent with the prominent induction of LCN2 transcript in UC,[25] and elevated expression of LCN2 in proteomic analysis of unfractionated UC mucosal biopsies.[9] LCN2 can be produced by intestinal epithelial cells in addition to neutrophils.[39] In accordance with this, we detected LCN2 in our sample of enriched epithelial cells, in the absence of neutrophil markers such as calprotectin. Although LCN is considered to be primarily neutrophil-derived with epithelial cells and monocytes contributing lesser amounts,[39] it was one of the most highly induced proteins in the inflamed

UC epithelial dataset, and ATAC-seq showed enhanced chromatin accessibility. These observations suggest that epithelial cells may be a more robust source of lipocalin than previously appreciated. Interestingly, epithelial expression of LCN2 can be induced by oxidative stress,[40] suggesting a common mechanism with observed changes in DUOX2/DUOXA2.

Other notable differences in proteins quantitated between inflamed UC biopsies and HCs included HLA-DRA and CD74, involved in antigen presentation. Elevated HLA-DRA is a hallmark of the inflamed epithelium in IBD,[41] and can be up-regulated by epithelial cells upon exposure to interferon gamma.[42] CD74 functions as the invariant chain and chaperone for MHC class II, and are the receptor for the cytokine migration inhibitory factor. Functionally, it has been linked to mucosal healing in IBD, promoting intestinal epithelial regeneration and repair.[43] Current findings point to a role for chromatin remodeling in promoting the expression of HLA-DRA and CD74.

In accordance with this, several proteins associated with antigen processing and presentation were induced in UC. Proteasome subunit beta type 9 (PSMB9; LMP2), is a component of the 20S proteasome that degrades endogenous proteins. Peptides generated by the 20S proteasome are transported by the ER-resident TAP1/TAP2 complex, for loading onto MHC class I molecules. TAP1, TAP2, and PSMB9 proteins were all increased in UC inflamed tissues compared to tissues from HCs, and PSMB9 was also elevated in UC uninflamed tissues compared to healthy. Interestingly, PSMB9 (LMP2) and TAP1 share a bidirectional promoter.[44] Although PSMB9, TAP1, and TAP2 proteins were all increased in inflamed tissues, ATAC-seq findings do not indicate increased accessibility of the promoter region.

Surprisingly, antibodies were observed to be among the proteins most differentially present on epithelial cells from UC patients relative to HCs. IgA (IGHA1) was the most so on epithelial cells from uninflamed UC biopsies and IgG1 (IGHG1) was among the most enriched proteins of inflamed UC biopsies relative to HC samples (Table A1). While IgA-producing B cells are more common than IgG-producing ones in normal colon, this ratio shifts to favor IgG in UC, particularly as it gets more inflamed.[45] Because the current study was an analysis of isolated epithelial cells, lacking any B cells that can produce immunoglobins, and because open chromatin corresponding to IGHA1 or IGHG1 was not detected, the source of this protein is presumably antibodies bound to epithelial cells. Open chromatin near the FCGR1A and FCGR1B genes encoding Fc receptors for IgG were identified by ATAC-seq, but were not modulated by inflammation, and no protein for IgG receptors was identified in the epithelial cell proteome, regardless of inflammation. If UC epithelial cells are not using Fc receptors to preferentially bind these antibodies, they may be the targets of autoantibodies, suggesting a role for the humoral immune system in the pathogenesis of UC. Indeed, high serum titers of both IgA and IgG autoantibodies specific for integrin $\alpha$V$\beta$6, which is expressed on the surface of epithelial cells,

have recently been very strongly correlated with the diagnosis of UC,[46] even predating its onset by as much as a decade.[47]

Taken together, our findings are consistent with oxidative stress at the colonic epithelium in UC driving antimicrobial responses, induction of antigen processing and presentation, and tissue damage. Our analysis indicates that changes in chromatin accessibility modulate the expression of AQP8, DUOX2/DUOXA2, LCN2, HLA-DRA, CD74, and other key proteins at sites of colonic inflammation, but do not account for all of the protein modulation in the colon. Additional factors, which may include protein stabilization, hydrolyzing enzymes, or post-transcriptional modifications, could contribute to shifts in the proteome associated with active UC. By defining epithelial-associated changes, the current proteomics and ATAC-seq databases may help to define mechanisms underlying damage and repair in UC.

## Supplementary Materials

Material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.gastha.2024.04.014.

## References

1. Sommer K, Wiendl M, Muller TM, et al. Intestinal mucosal wound healing and barrier integrity in IBD-crosstalk and trafficking of cellular players. Front Med (Lausanne) 2021;8:643973.
2. Herrlinger KR, Stange EF. Twenty-five years of biologicals in IBD: what's all the hype about? J Intern Med 2021;290(4):806–825.
3. Liu CY, Cham CM, Chang EB. Epithelial wound healing in inflammatory bowel diseases: the next therapeutic frontier. Transl Res 2021;236:35–51.
4. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. Cell 2019;178(3):714–730.e22.
5. Gruver AM, Westfall MD, Ackermann BL, et al. Proteomic characterisations of ulcerative colitis endoscopic biopsies associate with clinically relevant histological measurements of disease severity. J Clin Pathol 2022; 75(9):636–642.
6. Baldan-Martin M, Chaparro M, Gisbert JP. Tissue proteomic approaches to understand the pathogenesis of inflammatory bowel disease. Inflamm Bowel Dis 2021; 27(8):1184–1200.
7. Schniers A, Anderssen E, Fenton CG, et al. The proteome of ulcerative colitis in colon biopsies from adults - optimized sample preparation and comparison with healthy controls. Proteomics Clin Appl 2017;11(11–12):1700053.
8. Moriggi M, Pastorelli L, Torretta E, et al. Contribution of extracellular matrix and signal mechanotransduction to epithelial cell damage in inflammatory bowel disease patients: a proteomic study. Proteomics 2017; 17(23–24):1700164.
9. Schniers A, Goll R, Pasing Y, et al. Ulcerative colitis: functional analysis of the in-depth proteome. Clin Proteomics 2019;16:4.
10. Fogt F, Jian B, Krieg RC, et al. Proteomic analysis of mucosal preparations from patients with ulcerative colitis. Mol Med Rep 2008;1(1):51–54.
11. Janker L, Schuster D, Bortel P, et al. Multi-omics empowered deep phenotyping of ulcerative colitis identifies biomarker signatures reporting functional remission states. J Crohns Colitis 2023;17(9):1514–1527.
12. Jin L, Li L, Hu C, et al. Integrative analysis of transcriptomic and proteomic profiling in inflammatory bowel disease colon biopsies. Inflamm Bowel Dis 2019; 25(12):1906–1918.
13. Poulsen NA, Andersen V, Moller JC, et al. Comparative analysis of inflamed and non-inflamed colon biopsies reveals strong proteomic inflammation profile in patients with ulcerative colitis. BMC Gastroenterol 2012;12:76.
14. Bressenot A, Salleron J, Bastien C, et al. Comparing histological activity indexes in UC. Gut 2015; 64(9):1412–1418.
15. Huang T, Choi M, Tzouros M, et al. MSstatsTMT: statistical detection of differentially abundant proteins in experiments with isobaric labeling and multiple mixtures. Mol Cell Proteomics 2020. 19;10:1706–1723.
16. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43(7):e47.
17. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013. http://doi.org/10.48550/arXiv.1303.3997.
18. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-seq (MACS). Genome Biol 2008;9(9):R137.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15(12):550.
20. Saeterstad S, Ostvik AE, Royset ES, et al. Profound gene expression changes in the epithelial monolayer of active ulcerative colitis and Crohn's disease. PLoS One 2022; 17(3):e0265189.
21. Ricanek P, Lunde LK, Frye SA, et al. Reduced expression of aquaporins in human intestinal mucosa in early stage inflammatory bowel disease. Clin Exp Gastroenterol 2015;8:49–67.
22. Bienert GP, Moller AL, Kristiansen KA, et al. Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. J Biol Chem 2007;282(2): 1183–1192.
23. Escudero-Hernandez C, Munch A, Ostvik AE, et al. The water channel aquaporin 8 is a critical regulator of intestinal fluid homeostasis in collagenous colitis. J Crohns Colitis 2020;14(7):962–973.
24. Yde J, Keely S, Wu Q, et al. Characterization of AQPs in mouse, rat, and human colon and their selective regulation by bile acids. Front Nutr 2016;3:46.
25. Linggi B, Jairath V, Zou G, et al. Meta-analysis of gene expression disease signatures in colonic biopsy tissue from patients with ulcerative colitis. Sci Rep 2021;11(1): 18243.
26. Hardin JA, Wallace LE, Wong JF, et al. Aquaporin expression is downregulated in a murine model of colitis and in patients with ulcerative colitis, Crohn's disease and infectious colitis. Cell Tissue Res 2004; 318(2):313–323.

27. Te Velde AA, Pronk I, de Kort F, et al. Glutathione peroxidase 2 and aquaporin 8 as new markers for colonic inflammation in experimental colitis and inflammatory bowel diseases: an important role for H2O2? Eur J Gastroenterol Hepatol 2008;20(6):555–560.

28. Medrano-Fernandez I, Bestetti S, Bertolotti M, et al. Stress regulates aquaporin-8 permeability to impact cell growth and survival. Antioxid Redox Signal 2016; 24(18):1031–1044.

29. Conner GE. Regulation of dual oxidase hydrogen peroxide synthesis results in an epithelial respiratory burst. Redox Biol 2021;41:101931.

30. Luther J, Gala M, Patel SJ, et al. Loss of response to anti-tumor necrosis factor alpha therapy in Crohn's disease is not associated with emergence of novel inflammatory pathways. Dig Dis Sci 2018;63(3):738–745.

31. Rigoni A, Poulsom R, Jeffery R, et al. Separation of dual oxidase 2 and lactoperoxidase expression in intestinal crypts and species differences may limit hydrogen peroxide scavenging during mucosal healing in mice and humans. Inflamm Bowel Dis 2017;24(1):136–148.

32. MacFie TS, Poulsom R, Parker A, et al. DUOX2 and DUOXA2 form the predominant enzyme system capable of producing the reactive oxygen species H2O2 in active ulcerative colitis and are modulated by 5-aminosalicylic acid. Inflamm Bowel Dis 2014;20(3):514–524.

33. Dang PM, Rolas L, El-Benna J. The dual role of reactive oxygen species-generating nicotinamide adenine dinucleotide phosphate oxidases in gastrointestinal inflammation and therapeutic perspectives. Antioxid Redox Signal 2020;33(5):354–373.

34. Grasberger H, Gao J, Nagao-Kitamoto H, et al. Increased expression of DUOX2 is an epithelial response to mucosal dysbiosis required for immune homeostasis in mouse intestine. Gastroenterology 2015;149(7): 1849–1859.

35. Grasberger H, Magis AT, Sheng E, et al. DUOX2 variants associate with preclinical disturbances in microbiota-immune homeostasis and increased inflammatory bowel disease risk. J Clin Invest 2021;131(9):e141676.

36. Asai Y, Yamada T, Tsukita S, et al. Activation of the hypoxia inducible factor 1alpha subunit pathway in steatotic liver contributes to formation of cholesterol gallstones. Gastroenterology 2017;152(6):1521–1535.e8.

37. Sommer F, Backhed F. The gut microbiota engages different signaling pathways to induce Duox2 expression in the ileum and colon epithelium. Mucosal Immunol 2015;8(2):372–379.

38. Kluber P, Meurer SK, Lambertz J, et al. Depletion of lipocalin 2 (LCN2) in mice leads to dysbiosis and persistent colonization with segmented filamentous bacteria. Int J Mol Sci 2021;22(23):13156.

39. Zollner A, Schmiderer A, Reider SJ, et al. Faecal biomarkers in inflammatory bowel diseases: calprotectin versus lipocalin-2-a comparative study. J Crohns Colitis 2021;15(1):43–54.

40. Makhezer N, Ben Khemis M, Liu D, et al. NOX1-derived ROS drive the expression of lipocalin-2 in colonic epithelial cells in inflammatory conditions. Mucosal Immunol 2019;12(1):117–131.

41. McDonald GB, Jewell DP. Class II antigen (HLA-DR) expression by intestinal epithelial cells in inflammatory diseases of colon. J Clin Pathol 1987;40(3):312–317.

42. Hershberg RM, Framson PE, Cho DH, et al. Intestinal epithelial cells use two distinct pathways for HLA class II antigen processing. J Clin Invest 1997;100(1):204–215.

43. Farr L, Ghosh S, Jiang N, et al. CD74 signaling links inflammation to intestinal epithelial cell regeneration and promotes mucosal healing. Cell Mol Gastroenterol Hepatol 2020;10(1):101–112.

44. Wright KL, White LC, Kelly A, et al. Coordinate regulation of the human TAP1 and LMP2 genes from a shared bidirectional promoter. J Exp Med 1995; 181(4):1459–1471.

45. Wu KC, Mahida YR, Priddle JD, et al. Immunoglobulin production by isolated intestinal mononuclear cells from patients with ulcerative colitis and Crohn's disease. Clin Exp Immunol 1989;78(1):37–41.

46. Kuwada T, Shiokawa M, Kodama Y, et al. Identification of an anti-integrin alphavbeta6 autoantibody in patients with ulcerative colitis. Gastroenterology 2021; 160(7):2383–2394.e21.

47. Livanos AE, Dunn A, Fischer J, et al. Anti-integrin alphavbeta6 autoantibodies are a novel biomarker that antedate ulcerative colitis. Gastroenterology 2023; 164(4):619–629.