Proteomics highlights common and distinct pathophysiological processes associated to ileal and colonic ulcers in Crohn's disease

Nicolas Pierre[1,*], Catherine Salée[1], Charlotte Massot[1], Noëlla Blétard[2], Gabriel Mazzucchelli[3], Nicolas Smargiasso[3], Denis Morsa[4], Dominique Baiwir[4], Edwin De Pauw[3], Catherine Reenaers[5], Catherine Van Kemseke[5], Jean-Philippe Loly[5], Philippe Delvenne[2], Marie-Alice Meuwis[1,5,#] & Edouard Louis[1,5,#]

[1]Laboratory of Translational Gastroenterology, GIGA-institute, University of Liège, 4000 Liège, Belgium; [2]Department of Anatomy and Pathology, GIGA-institute, Liège University Hospital CHU, 4000 Liège, Belgium; [3]Laboratory of Mass Spectrometry, Chemistry department, University of Liège, 4000 Liège, Belgium; [4]GIGA Proteomics Facility, University of Liège, 4000 Liège, Belgium; [5]Hepato-Gastroenterology and Digestive Oncology Department, Liège University Hospital CHU, 4000 Liège, Belgium. [#]*Equally contributed to this work.*

Short title: Distinct proteomic profiles in ileal and colonic CD ulcers

***Corresponding author:**

Nicolas Pierre

Address: Translational Gastroenterology, GIGA institute Bât. B34 Quartier Hôpital, avenue de l'Hôpital 11, 4000 Liège 1, Belgique

Tel: +32 4 3662538

Fax: +32 4 3667889

Email: nicolas.pierre@uliege.be

Abstract

Background and Aims: Based on genetics and natural history, Crohn's disease can be separated in two entities, an ileal and a colonic disease. Protein based-approaches are needed to elucidate whether such subphenotypes are related to distinct pathophysiological processes.

Methods: The proteome of ulcer edge was compared to the one of paired control tissue (n=32 biopsies) by differential proteomics in the ileum and the colon of Crohn's disease patients (n=16). The results were analysed though a hypothesis-driven (based on literature) and a hypothesis-free approach (pathway enrichment analysis). To confirm one of the key pathway highlighted by proteomics, two proteins were also studied by immunochemistry in tissue biopsies.

Results: In the ileum and the colon, 4428 and 5204 proteins, respectively, were identified and quantified. Ileal and colonic ulcer edge differed by a distinct distribution of proteins of epithelial–mesenchymal transition, neutrophil degranulation and ribosome. Ileal and colonic ulcer edge were similarly characterised by an increase of the proteins implicated in the pathway of *protein processing in endoplasmic reticulum* and a decrease of mitochondrial proteins. Immunochemistry confirmed the presence of endoplasmic reticulum stress in the mucosa of ileal and colonic ulcer edge.

Conclusion: This study provides protein-based evidences showing partly distinct pathophysiological processes associated to ileal and colonic ulcer edge in Crohn's disease patients. This could constitute a first step toward the development of gut segment-specific diagnostic markers and therapeutics.

Keyword: Crohn's disease; ulcers; proteomics

1. Introduction

Crohn's disease (CD) is characterised by a chronic inflammation that can affect the entire gastrointestinal tract. This situation can lead to transmural lesions, strictures and fistulae. The pathophysiology of CD remains poorly understood; it is currently recognised as an inappropriate immune response against microbiota in genetically predisposed individuals[1]. Current treatments of CD patients are mainly based on corticosteroids, anti-metabolites and biologics directed against the tumour necrosis factor α (TNFα), the interleukins 12/23 and the α4β7 integrin[2]. Although these treatments improve the patients' quality of life by reducing flares, they remain only partly satisfactory. Indeed, the most used biologics in CD, the anti-TNFα, show a rate of primary non-response and a loss of response ranging between 10-30% and 23-46%, respectively[3]. The complexity and the heterogeneous presentation of CD encourage the development of personalised treatment[4]. In this context, accumulating evidences support the distinction of CD subphenotypes based on clinical characteristics, including disease location. Patients presenting an ileal or a colonic disease, the segments mainly affected in CD, show distinct biological and clinical characteristics. First, ileal CD is associated with a higher risk and an earlier need for surgery than colonic CD, mainly reflecting a greater inclination to develop strictures[5,6]. Second, the fecal proteins correlated with the presence of CD endoscopic lesions, calprotectin and lactotransferrin, show a higher performance to detect active disease in colonic than ileal CD[7]. Third, based on genetic risk factors, colonic CD is as far from ileal CD as from ulcerative colitis (UC), the other inflammatory bowel disease (IBD) only affecting the colon and the rectum[5]. Such a subphenotype classification opens new perspectives for the development of personalised diagnostic markers and treatments based on disease location. To this end, protein based-approaches are required to determine the relation between disease location and pathological processes.

The present study aimed to compare the proteomic profiles of ileal/colonic CD lesions with paired control tissues. Among the endoscopic lesions found in Crohn's patients (erythema, aphtoid lesion, ulcer, stricture and fistula), we choose to study ulcer since it represents the most frequent lesion seen in CD. Given that ulcer itself includes a loss of epithelial and subepithelial tissue, we rather studied the edge of ulcer as a relevant tissue to investigate the pathophysiology of this lesion. To unravel the distinction between ileal and colonic ulcer edge, we first analysed our data with a hypothesis-driven approach. On the one hand, the fecal level of calprotectin and lactotransferrin (contained in neutrophil granules) is higher in colonic than ileal active CD[7], suggesting a differential neutrophil involvement between the two segments. On the other hand, the epithelial cells transdifferentiation namely epithelial-mesenchymal transition (EMT) plays a key role in the development of CD gut strictures[8,9], a pathological process preferentially found in ileal CD[5]. Based on these observations, we hypothesized that neutrophil degranulation and EMT could be differentially affected in ileal and colonic ulcer edge. Then, we analysed our data with a hypothesis-free approach. In this context, we performed pathway enrichment analyses to determine common and segment-specific cellular processes associated with ileal and colonic CD ulcer edge. We confirmed by immunohistochemistry (IHC) the increase of two proteins involved in one of the highlighted key pathways.

2. Methods

2.1. Study population and sampling

This study received approval by the Ethic reviewing board of the University Hospital of Liège, Belgium (March 26 2013) (Belgian reference: 707201317029). Adult CD patients (n=156) undergoing endoscopy as part of their disease management were prospectively recruited between 2012 and 2017 according to classical diagnostic criteria. For the present study, 16 cases were selected based on the following criteria: 1) biopsy taken at the edge of an ulcer (U); 2) paired control (C) biopsy, i.e., tissue taken close to U in a macroscopically normal mucosa.

Ulcer was defined as a zone presenting a macroscopic erosion of the epithelium. Lesions described macroscopically as aphtoid were not included in the present study. Biopsies were collected in the endoscopy room according to a standardised procedure. Then, tissues were flash-frozen and stored in liquid nitrogen.

2.2. Label-free proteomic

Biopsies (~5 mg) were combined with 60 mg of beads (Diagenode, Belgium) and lysed by sonication on a Bioruptor® (Diagenode, 5 C°, high power, 15 cycles of 30s/30s ON/OFF) in 300 µl of a pH 7.4 RIPA buffer [4% (w/v) SDS, 1 mM DTT, 25 mM HEPES, 150 mM NaCl, 1 mM NaF, 1 mM Na_3VO_4, 25 mM sodium β-glycerophosphate and 4% (v/v) protease inhibitor cocktail EDTA free 25× (Roche Applied Science, Germany)]. Homogenates were centrifuged at 15,000 *g* for 10 min at room temperature. The supernatants were immediately stored at -80 C° until further analysis. The protein concentration was determined using the RCDC Protein Assay Kit (BioRad, USA) according to manufacturer's instructions. Twenty micrograms of total protein was precipitated using the 2D-clean up assay (GE Healthcare, USA). Then, proteins were digested with the Trypsin/Lys-C Mix Mass Spec Grade (Promega, USA) according to manufacturer's instructions. Finally, 3.5 µg of the resulting peptide mixtures were purified on ZipTip C18 (Thermo Fisher Scientific, USA), dried in a vacuum centrifuge and stored at -20 C°. Just before analysis, dried samples (3.5 µg) were solubilized in 15.75 µl of a pH 10 solution containing 100 mM ammonium formate and MPDSmix (MassPREP™ Digestion Standard Mixture, Waters, USA). This standard contains four non-human proteins digested: bovine serum albumin (BSA, P02769), yeast enolase 1 (ENO1, P00924), rabbit glycogen phosphorylase b (GPB, P00489) and yeast alcohol dehydrogenase 1 (ADH1, P00330). To control the quality of the instrumental set-up, two different concentrations of MPDSmix (MPDSmix 1 and MPDSmix 2) were spiked in U and C samples. The following molar ratios

126 (U/C) were expected: 1.00 (ADH1), 0.38 (GBP), 1.66 (ENO1), 8.96 (BSA). The quantity of
127 ADH1 injected was 150 fmol.
128 Samples were analysed by ultra-performance liquid chromatography/electrospray ionization
129 tandem mass spectrometry (UPLC-ESI-MS/MS). This system consists of a 2D nanoAcquity
130 chromatography (Waters) coupled online with a Q Exactive™ Plus Hybrid Quadrupole-
131 Orbitrap™ mass spectrometer (Thermo Fisher Scientific), equipped with a nano-electrospray
132 source operated in positive ion mode. The spray voltage and the temperature of heated capillary
133 were set to 2.2 kV and 300 °C, respectively. Nine µL of the solubilized peptide mixture was
134 injected on the UPLC-ESI-MS/MS system. The first dimension of UPLC separation was
135 performed at pH 10 on a X-Bridge BEH C18 5 µm column (300 µm × 50 mm, Waters). Samples
136 were loaded at 2 µL/min in a pH 10 solution of 20 mM ammonium formate. Then, samples
137 were eluted by three steps of acetonitrile: 13.3%, 19% and 65%. After a 1:10 dilution to pH 3
138 with aqueous solution of 0.1% formic acid, peptides were loaded on a trap column Symmetry
139 C18 5 µm (180 µm × 20 mm, Waters). The second dimension of UPLC separation was
140 performed on HSST3 1.7 µm (75 µm × 250 mm, Waters) column. Peptides were separated by
141 using a 140 min linear gradient of solvent A (0.1% formic acid in water) and B (0.1% formic
142 acid in acetonitrile). The flow rate was constant (250 nL/min) and the following gradient (A/B)
143 was applied: 0 min, 99/1% (v/v); 5 min, 93/7% (v/v); 140 min, 65/35% (v/v). The total run time
144 was 180 min for each eluted fraction: 140 min for the linear gradient and 40 min for cleaning
145 and re-equilibration steps. The mass spectrometer method consisted of one full MS scan
146 followed by data-dependent MS/MS scans of the 12 most intense ions. The parameters for MS
147 spectrum acquisition were set as follows: mass range from 400 to 1750 *m/z* with R=70000
148 (defined at *m/z* 200), the automatic gain control (AGC) target was 1×10^6 and the maximum
149 injection time was 200 ms. The parameters for MS/MS spectrum acquisition were: 2.0 *m/z*
150 isolation window, a stepped normalised collision energy of 25.0 with R=17500 (defined at *m/z*

200), AGC target of 1×10^5 and a maximum injection time of 50 ms (underfill ratio of 1.0 %). Raw data were recorded with Xcalibur software (Thermo Fisher Scientific).

Identifications and quantifications of proteins were obtained using the MaxQuant software Version 1.5.5.1[10]. Proteins were searched in the Uniprot-human database (20237 reviewed entries, release 2017_09) enriched with the sequences of the 4 MPDSmix proteins. The following settings were used for protein identification: trypsin as digestion enzyme, a maximal number of miscleavages equal to 2, a minimal peptide length of 7 amino acids, carbamidomethylation of cysteines as fixed modification, methionine oxidation as variable modification, a minimal number of peptide *per* protein equal to 2, a minimal number of unique peptide *per* protein equal to 1, precursor mass tolerance of 4.5 ppm, fragment mass tolerance of 0.5 Da, false discovery rate (FDR) of 1% for peptide spectrum matches and proteins. Data normalisation was performed using the MaxQuant label-free quantification (LFQ) algorithm[11]. The enrichment pathway analyses were done using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway databases[12]. The KEGG pathway enrichments were determined with the DAVID 6.8 (Database for Annotation, Visualization and Integrated Discovery) bioinformatics tool[13]. All enrichment analyses were performed using the appropriate (ileum or colon) identified proteins as background.

2.3. Anatomopathological analysis

The frozen biopsies were placed in formaldehyde for fixation during 2 hours at room temperature, followed by dehydration in 70% ethanol (v/v). Tissues were further paraffin embedded according to standard procedure[14]. Serial sections of 5 µm were obtained from each biopsy. Sections were mounted on SuperFrost Plus glass slides for haematoxylin and eosin stained staining (H&E) and immunohistochemistry.

Neutrophils, lymphocytes, and plasmocytes infiltration were evaluated by a trained anatomopathologist (N.B) after examination of the H&E stained sections. A standard method was applied to grade the infiltration of inflammatory cells and the following scores were used: 0 (none), 1 (light), 2 (moderate) and 3 (severe).

2.4. Immunohistochemistry

Tissue sections were deparaffinised and rehydrated using xylene, ethanol and water washes. Antigen retrieval was performed in Target Retrieval Solution (DAKO, Agilent Technologies, USA) using a steamer for 10 min. Endogenous peroxidase activity was quenched by incubation with 3% (v/v) hydrogen peroxide for 10 min. After 3 washes in phosphate-buffered saline (PBS), nonspecific bindings were blocked by a 20 min incubation in the Protein block serum-free ready-to-use (DAKO). Then, tissue sections were incubated at room temperature for 1h with the following primary antibodies: HSPA5 (1:500, #3177, Cell Signaling, USA); HSP90B1 (1:800, #20292, Cell Signaling). After 3 washes in PBS, tissue sections were incubated 30 min with the EnVision System-HRP labelled polymer anti-Rabbit (DAKO). The chromogen used was 3,3′- Diaminobenzidine (10 min at RT) and counterstaining was done with haematoxylin as previously described[15].

Three non anatomopathologist scorers (M-A.M, C.M, C.S), performed a blinded readings of the biopsy sections for IHC score. Conflicting results were solved by consensus with the advice of the trained anatomopathologist (N.B). The following staining scores were used: 0 (none), 1 (weak), 2 (medium), 3 (strong) and 4 (very strong). The final score was the averaged value of the different fields (minimum 5/section), weighted by the percentage of the slide sharing similar score intensity. The IHC scores were determined in the surface epithelium, the crypt and the *lamina propria*. These zones were defined as described in the literature[16].

2.5. Statistical analysis

In the proteomic experiment, data transformation and statistical tests applied on the LFQ values were done using Perseus software Version 1.6.0.7[17]. Data were Log_2 transformed to reach a normal distribution. Differences between U and C protein abundances were assessed using a two tailed paired t-test (n=8 pairs for each gut segment). To correct for multiple testing, p-values were adjusted with the Benjamini-Hochberg method. According to statistical nomenclature, the corrected p-value were named q-value. The volcano plots were done using the ggplot2 R package.

In the histological experiment, normality of the distribution has been tested with the Shapiro-Wilk test. When appropriate, either two tailed paired t-test or two tailed Wilcoxon matched-pairs signed-ranks test was applied (n=6-8 pairs for each gut segment). Statistical analysis and graphical illustrations of the data have been performed using GraphPad Prism Version 7.0 (GraphPad, USA).

For all the statistical analysis, the level of significance was set at 0.05.

2.6. Proteomic data availability

All the raw data generated by the mass spectrometer, the MaxQuant files and the related samples information have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[18] with the data set identifier PXD012284.

3. Results

3.1. Patients' characteristics and histopathological analysis

One hundred and fifty-six patients were prospectively recruited. The reviewing of the patients' medical records enabled the selection of 16 cases for the present study, 8 with ileal biopsies in ulcer edge (U) and paired control tissues (C) and 8 with colonic biopsies in U and C. The

patients' clinical characteristics are presented in Table 1. To perform proteomic and histological analyses in the same zone, two paired and adjacent biopsies (U and C) were collected for each patient. A total of 64 biopsies was analysed, 32 by proteomics (2 paired biopsies for each selected case) and 32 by histology (2 paired biopsies for each selected case). No technical replicates were performed in this study.

In U but not in C, the pathologist (N.B) confirmed the classical architectural modifications of the epithelium and the infiltration of inflammatory cells compatible with CD[19]. The persistence of the epithelium and the absence of granulation tissue confirmed that the biopsies were not taken in the ulcer itself. Our quantitative histological approach supported an increase of neutrophils, lymphocytes and plasmocytes infiltration in U of ileum and colon (Figure 1).

3.2. Proteomic dataset: description and quality control

In the Table 2, the proteomic dataset is briefly described by showing information on MS/MS, MS/MS identified and unique peptides. The mass spectra of the proteomic dataset were assigned to 36632 and 46741 unique peptides corresponding to 4428 and 5204 proteins in ileum and colon samples, respectively (Table 2). Among these proteins, 2423 (ileum) and 2915 (colon) were identified and quantified with 100% occurrence in the 2 groups (U and C). By applying the most stringent statistical approach, i.e., p-value calculated for proteins with 100% occurrence in the 2 groups (U and C) and corrected by the Benjamini-Hochberg method, we found 415 (ileum) and 402 (colon) proteins differentially abundant in U vs. C (Table 2). We identified and quantified the 4 proteins of the MPDSmix standards with coherent differential abundances (U vs. C), showing the quality of our instrumental set-up (Supplementary Figure 1).

3.3. Distinct pathophysiological processes associated to ileal and colonic ulcer edge in CD

First, we tested whether the magnitude of neutrophil degranulation is higher in colonic than ileal U. To this end, we selected this family of proteins in our dataset by using the Reactome database R-HSA-6798695[20]. To catch the mainly affected ones, we only kept those with the highest fold change (U/C>2) and an occurrence of minimum 50% in the two groups (U and C). Based on this strategy, we selected 32 and 16 proteins in the colon and the ileum, respectively (Supplementary Table 1, Figure 2A and 2B). The well-recognised fecal markers of CD activity[7], calprotectin (S100A8 and S100A9 subunits) and lactotransferrin (LTF), showed a higher fold change (U/C) in colonic than ileal U: 6.7 vs. 3.0 for S100A8, 7.1 vs. 3.2 for S100A9 and 16.6 vs. 4.2 for LTF (Supplementary Table 1; Figure 2A and 2B). The lactotransferrin presented a higher fold change than the calprotectin in ileal and colonic U (Supplementary Table 1, Figure 2A and 2B). Such effects are in line with those observed at the fecal level[7], thus supporting the consistency of the present dataset. The matrix metalloproteinase-9 (MMP9) and the myeloperoxidase (MPO) are proteins contained in neutrophil granules which have been identified as potential biomarkers of CD activity[21,22]. Herein, MMP9 and MPO showed a much higher fold change (U/C) in the colon than the ileum: 58.4 vs. 6.6 for MMP9 and 11.6 vs. 3.3 for MPO (Supplementary Table 1; Figure 2A and 2B). Taken together, these data clearly support that neutrophil degranulation and/or infiltration is higher in colonic than ileal U.

Second, we tested whether the EMT is more impacted in ileal than colonic U. The key proteins regulated during this cellular process have been selected through searches in the literature[23–26]. Among those proteins, 13 are acquired and 17 are attenuated during EMT (Supplementary Table 2). We illustrated their distribution in the Figure 2C and 2D. In the ileum but not the colon, 7 markers of epithelial cells were significantly attenuated (U vs. C) and 2 markers of mesenchymal cells were augmented (Figure 2C and 2D). The cytokeratin-20 (KRT20), a specific keratin of the intestinal epithelium, was attenuated in ileal but not colonic U (Figure 2C and 2D). The distribution of proteins implicated in adherent junction (CTNNB1: β-catenin;

MLLT4: afadin:) and desmosome (DSP: desmoplakin), seem also specifically altered in ileal U (Figure 2C and 2D). The distribution of proteins involved in the tight junction, occludin (OCLN) and tight junction protein ZO-1 (TJP1), were not impacted in ileal nor colonic U (Figure 2C and 2D). On the other hand, the marker of mesenchymal cells, fibronectin type-III domain-containing protein 3A (FNDC3A), was specifically increased in ileal U (Figure 2C and 2D). Thus, our results indicate that EMT seems a more pronounced feature of ileal than colonic U.

We also investigated, without hypothesis, the pathophysiological processes more specifically involved in ileal or colonic U. To this end, we selected proteins detected in the two segments and differentially abundant (U vs. C, proteins with 100% occurrence in the 2 groups, q-value<0.05) in either the ileum or the colon (Figure 3). Thus, two lists of proteins were generated: 1) proteins with abundance specifically affected in U of the ileum (n=293, Supplementary Table 3); 2) proteins with abundance specifically affected in U of the colon (n=286, Supplementary Table 4). Those lists of proteins were submitted to pathway enrichment analysis. Remarkably, ileal U presented a highly significant enrichment of ribosomal proteins which was not observed in colonic U (Supplementary Table 5). The distribution of ribosomal proteins is represented in Figure 4. By separating the cytosolic and the mitochondrial ribosomal proteins, we found two clear patterns characterising ileal or colonic U (Figure 4). While cytosolic ribosomal proteins increased in ileal U, the mitochondrial ribosomal proteins presented a clear trend to decrease in colonic U (Figure 4). It seems a general pattern since both significantly and non-significantly affected (U vs. C) ribosomal proteins showed a similar distribution (Figure 4). More precisely, 24 ribosomal proteins showed an increased abundance in ileal but not colonic U (Figure 4). The 60S acidic ribosomal protein P2 (RPLP2) was the only ribosomal protein showing an increased abundance in colonic U, this was not observed in ileal U (Figure 4). On the other hand, 5 mitochondrial ribosomal proteins show a decreased

abundance in colonic but not ileal U (Figure 4). Remarkably, one mitochondrial ribosomal protein (39S ribosomal protein L15, MRPL15) showed an increase in its abundance in ileal but not colonic U (Figure 4). In addition, colonic U exhibited an enrichment of proteins implicated in the *oxidative phosphorylation* pathway which was not observed in ileal U (Supplementary Table 5). Thus, the mitochondrial proteins of the *oxidative phosphorylation* appear particularly affected in colonic U. Altogether, our hypothesis-free approach indicates that the ribosomal proteins and the proteins of the *oxidative phosphorylation* pathway are differently affected in ileal and colonic U.

3.4. Common pathophysiological processes associated to ileal and colonic ulcer edge in CD

To investigate the pathophysiological similarities between ileal and colonic CD, we selected a list of proteins for which the abundance was similarly affected in the two segments. To this end, we merged the ileal and the colonic proteins differentially abundant between U and C (proteins with 100% occurrence in the 2 groups, q-value<0.05). We obtained 106 proteins (Figure 3B). Among these proteins, only one showed an opposite distribution (U/C) between ileum and colon: the protein kinase C and casein kinase substrate in neurons protein 2 (PACSIN2), increased in ileal U but decreased in colonic U. Thus, 105 proteins were similarly affected in ileal and colonic U (Supplementary Table 6). These proteins were enriched in pathways regulating: 1) metabolism and 2) protein processing in endoplasmic reticulum (ER) (Supplementary Table 7). The enrichment of metabolic pathways (*Valine leucine and isoleucine degradation*, *Fatty acid degradation*, *Fatty acid metabolism*, *Carbon metabolism*, *Biosynthesis of antibiotics*, *beta-Alanine metabolism*, *Metabolic pathways*, *Propanoate metabolism*, *Cardiac muscle contraction*, *Citrate cycle* and *Fatty acid elongation*, Supplementary Table 7) is driven by the presence of a high proportion of mitochondrial proteins (31/105, Supplementary Table 6). Indeed, the same analysis performed without the mitochondrial proteins results in no

enrichment of metabolic pathways (data not shown). The pathway of *protein export* (Supplementary Table 7) regroups proteins regulating the entry of nascent polypeptide inside the ER and the export of proteins outside the ER. These proteins are already included in the pathway of *protein processing in ER*. Thus, mitochondrial proteins and proteins implicated in *protein processing in ER* are similarly affected in ileal and colonic U.

The distribution of mitochondrial proteins showed a general reduction of their abundance in ileal and colonic U (Figure 5). Only one mitochondrial protein (MRPL15) presented a significant abundance increase in ileal U (Figure 5A). Among the mitochondrial proteins with a differential abundance between U and C (orange dots above the horizontal line in Figure 5), 50/51 and 74/74 were reduced in ileal and colonic U, respectively. Remarkably, the protein with the highest negative fold change (U/C<1) in the ileum (U/C=0.09) and the colon (U/C=0.20) was a mitochondrial protein: the hydroxymethylglutaryl-CoA synthase (HMGCS2) (Supplementary Table 6, Figure 5).

The proteins involved in *protein processing in ER* displayed a general increase of their abundance in ileal and colonic U. This result is indicative of ER stress, a situation where unfolded proteins accumulate inside the ER. Only one protein (Calpain-2 catalytic subunit, CAPN2) of *protein processing in ER* showed a significant decrease of its abundance in colonic U (Figure 5B). Among the proteins involved in *protein processing in ER* with a significant differential abundance between U and C (black dots above the horizontal line in Figure 5), 25/25 and 27/28 were increased in U of ileum and colon, respectively. To illustrate the pathway of *protein processing in ER*, we highlighted the distribution of 2 well-recognised markers of ER stress: HSPA5 (heat shock 70 kDa protein 5) also known as BiP (binding immunoglobulin protein) and HSP90B1 (heat shock protein 90 kDa beta member 1) also known as endoplasmin (Figure 5). In ileal and colonic U, the effect observed on the mitochondrial proteins and the proteins of *protein processing in ER* seems general since whatever the value of their statistics

(U vs. C) their patterns are similar (Figure 5). Taken together, these results highlight that ileal and colonic U appear to be characterised by a reduction of mitochondrial proteins and an increase of the proteins of *protein processing in ER*.

As ER stress appears as a critical event in the pathophysiology of CD[27], we decided to go further by confirming and localising this cellular perturbation. To this end, we analysed by IHC the abundances of HSPA5 and HSP90B1 in the paired biopsies (U vs. C) collected in the same zones as those used for the proteomic analysis. In the surface epithelium, the abundance of ER stress markers was not modified (Figure 6A, 6B, 6E and 6F). In the crypt, the ER stress markers increased in colonic but not ileal U (Figure 6A, 6C, 6E and 6G). In the *lamina propria* of colonic U, HSP90B1 and HSPA5 were increased (Figure 6A, 6D, 6E and 6H). In the *lamina propria* of ileal U, HSP90B1 but not HSPA5 was increased (Figure 6A, 6D, 6E and 6H). This IHC analysis confirms part of our proteomic data and characterise the mucosal localisation of ER stress in ileal and colonic U.

4. Discussion

In this study, we generated a highly relevant proteomic dataset describing the proteome of ileal and colonic ulcer edge in CD. Few studies investigated the proteome of colonic CD lesions; to the best of our knowledge, none analysed the proteome of ileal CD lesions. In CD, ileum is probably less studied than colon since its access remains more difficult by endoscopy. Another explanation could be that, until recently, no clear evidences indicated a distinct pathological process between ileal and colonic CD, thus making the analysis of ileum unnecessary. However, a genome-wide association study showed that ileal and colonic CD could be distinguished by genetic risk factors, suggesting a potential relation between disease location and pathological process[5]. Such a hypothesis needs to be investigated by proteomics since transcripts and genes are far from predicting proteins levels[28,29]. Thanks to both hypothesis-driven and hypothesis-

free approaches, we highlighted common and gut segment-specific pathophysiological processes in ileal and colonic ulcers edge of CD patients.

In agreement with our hypothesis-driven approach, the markers of neutrophil degranulation were more impacted in colonic than ileal ulcer edge. This could be linked to the higher microbial load in the colon than the ileum[30]. This result is also in line with the higher fecal level of calprotectin and lactotransferrin in colonic than ileal active CD[7]. Interestingly, neutrophil infiltration is well-known to impair the chronic wound healing in IBD patients[31,32]. In this context, the proteins from neutrophil granules, such as MMPs and MPO, have been identified as key deleterious actors[31,33]. Taken together, these observations indicate that a pharmacological approach or biomarkers targeting proteins from neutrophil granules could exhibit a differential success in colonic and ileal CD.

By measuring 30 EMT markers, our study provides a comprehensive overview of this cellular transdifferentiation. According to our hypothesis, we found a higher level of EMT in the ileal than the colonic ulcer edge. Given the role of EMT in fibrosis[8], our result highlights a potential mechanism by which the ileum is more affected by fibrotic stricture than the colon in CD. In this disease, a pharmacological agent targeting fibrosis remains an unmet clinical need[8]. This makes attractive the comprehension of such pathological process.

Based on our hypothesis-free approach, ileal and colonic ulcer edge were also distinguished by a differential distribution of ribosomal proteins. While ileal ulcer edge showed an increased abundance of the cytosolic ribosomal proteins, we found a decreased abundance of the mitochondrial ribosomal proteins in colonic ulcer edge. The meaning of these findings remains unknown.

Our hypothesis-free analysis also highlighted common pathophysiological process associated to ileal and colonic ulcer edge. In ileum and colon, ulcer edge is characterised by a striking decrease of mitochondrial proteins. Our finding supports a previous proteomic study showing

400 that 93.5 % of mitochondrial proteins decrease in colon of paediatrics CD patients compared to
401 healthy subjects[34]. In a transcriptomic study, it was reported that HMGCS2 transcript, a mRNA
402 coding for a mitochondrial protein, was the most downregulated gene in colonic erythematous
403 zones of CD patients[35]. Accordingly, we found that HMGCS2 was the protein with the most
404 decreased abundance in ileal and colonic U (Figure 5). In UC patients, a proteomic study found
405 that the majority (8/12) of proteins with a lower abundance in disease active tissue were
406 localised in the mitochondria[36]. Others revealed a decreased activity of the mitochondrial
407 respiratory chain complex II, III and IV in colon from UC compared to healthy patients[37]. At a
408 morphological level, electron microscopy analysis showed swollen mitochondria with irregular
409 cristae in inflamed ileum and colon from CD and UC patients, respectively[36,38]. Taken together,
410 these results indicate that mitochondrial dysfunction in the mucosa might be a feature of IBD
411 whatever the type of lesion, the segment and the age of the patients.
412 In addition to mitochondrial proteins, we found another common feature of ileal and colonic
413 CD ulcer edge: the involvement of the pathway of *protein processing in ER*. From translation
414 to protein export, *protein processing in ER* encompasses all the steps regulating protein
415 secretion: translocation of nascent polypeptide into the ER, protein folding, folding quality
416 control, export of correctly folded proteins to Golgi for secretion, unfolded protein response
417 (UPR), degradation of unfolded proteins through endoplasmic-reticulum-associated protein
418 degradation (ERAD) and apoptosis. In ileal and colonic U, we found an increase of the proteins
419 involved in all these cellular processes, thus highlighting anomalies in the whole ER
420 proteostasis. The HSPA5, an ER-resident chaperon, is a well-recognised marker of ER stress
421 playing a key role in protein folding and UPR activation[39]. As shown by studies on transcripts
422 and protein levels, HSPA5 increases in inflamed ileum from CD patients compared to healthy
423 control[27,40]. In agreement with this finding, we found by proteomics an increase of HSPA5
424 abundance in ileal and colonic U. We confirmed this increase by IHC in colonic but not ileal

ulcer edge. This could be due to a higher increase of HSPA5 in colonic (+41%) than ileal (+22%) ulcer edge. Such small effect may be difficult to confirm by a quantitative approach based on a discrete visual scale. In ileal and colonic U, we confirmed by IHC the increase of HSP90B1, another ER-resident chaperon increased upon ER stress. These results support previous findings showing that ER stress plays a key role in CD pathophysiology[27,41]. Furthermore, our IHC analysis brings additional information by showing that ER stress seems to affect the crypt and the *lamina propria* rather than the surface epithelium.

Alteration of mitochondria and ER homeostasis can be linked to the pathophysiological process regulating ulcer formation and healing. In intestinal epithelial cells (IECs), such stresses are known to reduce stemness[42,43]. This effect could in turn alter the capacity of re-epithelization since stem cells activity play a capital role in mucosal healing[44]. In line with this objective, alleviating ER stress with chemical chaperones reduces dextran sodium sulfate (DSS)-induced colitis in mice[45]. On the other hand, mitochondrial dysfunction has been shown to impair the maintenance of epithelial barrier probably since this process requires a considerable amount of energy[46,47]. In T84 epithelial cell and gut tissues from CD patients, perturbation of mitochondria function induced by uncoupler of oxidative phosphorylation causes epithelial barrier defect[47]. In mice, the use of mitochondria-targeted antioxidant reduces DSS-induced colitis[47]. In this context, targeting mitochondrial dysfunction and ER stress could be an attractive goal to restore the regenerative capacity of the epithelium.

The strengths of our work include the paired design of the tissue analysis. Such experimental approach reduces the biological noise induced by patient's genetic and environmental heterogeneity, a main issue when using human samples[48]. As a consequence, a paired design leads to a higher statistical power than a non-paired design all things being equal[49]. In addition, a paired design minimises the effects of cofounding factors since the compared groups are *de facto* perfectly balanced. The definition of the studied tissue is also a fundamental factor which

gives meaning to the results. In the literature, affected tissue is regularly defined as “active zone” or “involved zone”. The use of such terms is confusing when interpreting the results. In the present study, experienced endoscopists systematically biopsied the ulcer edge as the lesion tissue. By studying such tissue, our objective was clearly to investigate the pathological processes regulating ulcer formation and healing in ileum and colon.

Our work has some limitations. Legitimately, it could be argued that our results actually reflect the presence of a distinct cellular population between the control and the affected tissues. Such bias is inherent to the use of homogenised biopsies where cellular population are mixed together. Although we have tried to minimise this bias by avoiding biopsies in the ulcer itself and rather targeting the ulcer edge, this prevents to know whether the results obtained are due to a perturbation of the cellular homeostasis and/or a modification of the cellular population. In the biopsies that were analysed, the ulcer edge was associated with crypt hyperplasia, villous atrophy and shortening of the surface epithelium thickness. Consequently, the proportion of crypt-based cells is most probably higher in ulcer edge than non-lesional tissue. Compared to surface epithelium, crypts contain a higher level of proteins involved in *protein processing in ER* and a lower level of mitochondrial proteins[50,51]. This is what we observed in ileal and colonic U thus supporting the idea that part of our results is driven by an elevated proportion of crypt-based cells in ulcer edge. However, such hypothesis can only explain part of our results. Indeed, the proposed modification of the epithelial cellular population could not explain the fact that we found, by IHC, the presence of ER stress also in the *lamina propria*. In addition, ER stress and morphological alteration of the mitochondria have been reported in non-lesional and mildly inflamed mucosa of CD patients, respectively[27,38].Thus, alteration of mitochondria and ER homeostasis seem to precede a change of cellular population caused by the eroding process. In addition, the bias discussed in this section does not appear suitable to explain the distinct pathophysiological processes associated to ileal and colonic ulcer edge. Indeed, a change in

cellular population caused by the inflammatory process should induce a similar effect in the two segments. Finally, if this limitation is relevant for pathophysiology and new treatment perspective, it is not for the search of new biomarkers of disease activity.

Others limitations of our study could be the potential effects associated to: 1) the different colonic regions biopsied (ascending, transverse, descending and sigmoid); 2) gender difference when comparing ileum and colon. However, to the best of our knowledge, there is no published data suggesting an effect of gender and colonic regions on the disease process. Therefore, it is very unlikely that it explains the observed results.

In summary, we highlighted for the first time protein-based evidence showing different pathophysiological processes between ileal and colonic CD. These distinctions rely on the proteins of the epithelial-mesenchymal transition, neutrophil degranulation and ribosome. Such results underline the need to consider the pathophysiology of colonic and ileal CD as partly distinct. Our data also support that ileal and colonic ulcer edges are characterised by a reduction of mitochondrial proteins and an increase of the proteins belonging to the pathway of *protein processing in ER*. This study offers perspectives for the development of personalised CD diagnostic markers and treatment based on disease location.

Funding

This work was supported by the Walloon Region and the Fond européen de développement régional (FEDER) [Grant portofollio 246099-510388]. Additional support was provided by internal funding of Uliège, CHU de Liège and the MSD.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Samira Azarzar, Nancy Rosière, Lisette Trzpiot and Nanou Tanteliarisoa Haingo for their precious technical assistance. We thank the Biothèque Hospitalière Universitaire de Liège (ULiège and CHU de Liège) and the Immunohistology Facility of the GIGA-institute for the preparation of the sample specimens. We also thank the GIGA Proteomics Facility for their expertise.

Author contributions

N.P., M-A.M. and E.L. designed the experiment. E.L., M-A.M., C.M., C.R., C.V-K., J-P.L, provided the samples and the clinical information. N.P. performed the sample preparation for the proteomic analysis. D.B., D.M., N.S., G.M., and E.D-P. managed the injection of the proteomic experiments and provided advises for the data analysis. M-A.M., C.M., C.S., N.B., and P.D. performed the histological analysis. N.P., M-A.M. and E.L. analysed the data and wrote the paper.

Supplementary Data

Supplementary data are available at ECCO-JCC online.

References

1. Kohr B, Gardet A, Xavier RJ Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;**474**(7351):307–17.
2. Ha F, Khalil H Crohn's disease: a clinical update. *Therap Adv Gastroenterol* 2015;**8**(6):352–9.
3. Roda G, Jharap B, Neeraj N, Colombel JF Loss of Response to Anti-TNFs: Definition, Epidemiology, and Management. *Clin Transl Gastroenterol* 2016;**7**(1):e135-5.
4. Danese S New therapies for inflammatory bowel disease: From the bench to the bedside. *Gut* 2012;**61**(6):918–32.

5. Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: A genetic association study. *Lancet* 2016;**387**(10014):156–67.

6. Louis E, Collard A, Oger A-F, Belaiche J Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut* 2001;**49**(6):777–82.

7. Sipponen T, Savilahti E, Kolho KL, Nuutinen H, Turunen U, Färkkilä M Crohn's disease activity assessed by fecal calprotectin and lactoferrin: Correlation with Crohn's disease activity index and endoscopic findings. *Inflamm Bowel Dis* 2008;**14**(1):40–6.

8. Pariente B, Hu S, Bettenworth D, Speca S, Desreumaux P, Meuwis M-A, et al. Treatments for Crohn's Disease–Associated Bowel Damage: A Systematic Review. *Clin Gastroenterol Hepatol* 2019;**17**(5):847–56.

9. Jiang H, Shen J, Ran Z Epithelial-mesenchymal transition in Crohn's disease. *Mucosal Immunol* 2018;**11**(2):294–303.

10. Tyanova S, Temu T, Cox J The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 2016;**11**(12):2301–19.

11. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol Cell Proteomics* 2014;**13**(9):2513–26.

12. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;**44**(D1):D457–62.

13. Huang DW, Sherman BT, Lempicki RA Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;**37**(1):1–13.

14. Bancroft JD, Gamble M *Theory and practice of histological techniques*. 6th ed. Churchill Livingstone; 2008.

15. Quesada-Calvo F, Massot C, Bertrand V, Longuespée R, Blétard N, Somja J, et al. OLFM4, KNG1 and Sec24C identified by proteomics and immunohistochemistry as potential markers of early colorectal cancer stages. *Clin Proteomics* 2017;**14**(1):9.

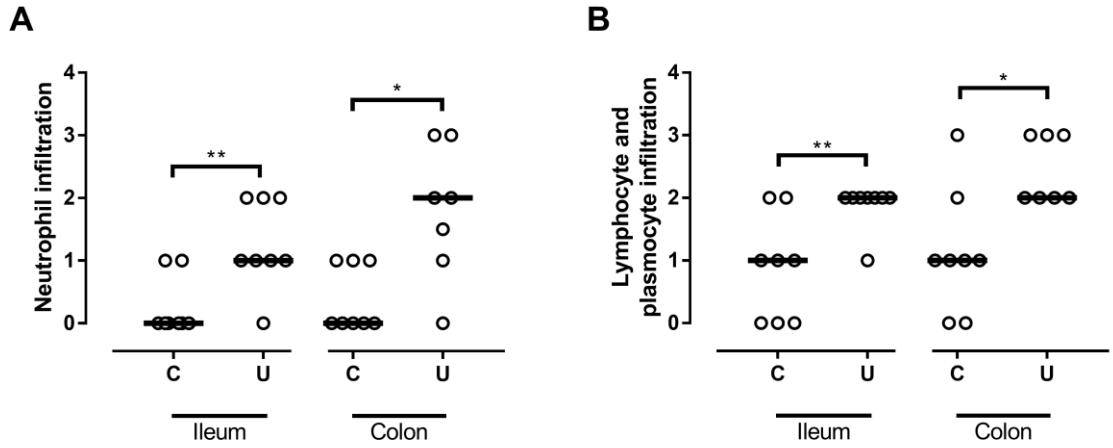
16. Geboes K Histopathology of Crohn ' s Disease and Ulcerative Colitis. *Inflamm Bowel Dis*

554 2003;**18**:255–76.
555 17. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus
556 computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*
557 2016;**13**(9):731–40.
558 18. Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the
559 PRIDE database and its related tools. *Nucleic Acids Res* 2016;**44**(D1):D447–56.
560 19. Magro F, Langner C, Driessen A, Ensari A, Geboes K, Mantzaris GJ, et al. European
561 consensus on the histopathology of inflammatory bowel disease. *J Crohn's Colitis*
562 2013;**7**(10):827–51.
563 20. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome
564 Pathway Knowledgebase. *Nucleic Acids Res* 2018;**46**(D1):D649–55.
565 21. Duvoisin G, Lopez RN, Day AS, Lemberg DA, Gearry RB, Leach ST Novel Biomarkers and
566 the Future Potential of Biomarkers in Inflammatory Bowel Disease. *Mediators Inflamm*
567 2017;**2017**(CD).
568 22. Kofla-Dlubacz A, Matusiewicz M, Krzystek-Korpacka M, Iwanczak B Correlation of MMP-3
569 and MMP-9 with crohn's disease activity in children. *Dig Dis Sci* 2012;**57**(3):706–12.
570 23. Kalluri R, Neilson EG Epithelial-mesenchymal transition and its implications for fibrosis. *J*
571 *Clin Invest* 2003:1776–84.
572 24. Zeisberg M, Neilson EG Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest*
573 2009:1429–37.
574 25. Sleeman JP, Thiery JP SnapShot: The epithelial-mesenchymal transition. *Cell*
575 2011;**145**(1):162–162.e1.
576 26. Francart M, Lambert J, Vanwynsberghe AM, Thompson EW, Bourcy M, Polette M, et al.
577 Epithelial – Mesenchymal Plasticity and Circulating Tumor Cells : Travel Companions to
578 Metastases. *Dev Dyn* 2018;**247**:432–50.
579 27. Kaser A, Lee AH, Franke A, Glickman JN, Zeissig S, Tilg H, et al. XBP1 Links ER Stress to
580 Intestinal Inflammation and Confers Genetic Risk for Human Inflammatory Bowel Disease.
581 *Cell* 2008;**134**(5):743–56.

582 28. Geiger T, Cox J, Mann M Proteomic changes resulting from gene copy number variations in
583 cancer cells. *PLoS Genet* 2010;**6**(9).
584 29. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of
585 human colon and rectal cancer. *Nature* 2014;**513**(7518):382–7.
586 30. Bowcutt R, Forman R, Glymenaki M, Carding SR, Else KJ, Cruickshank SM Heterogeneity
587 across the murine small and large intestine. *World J Gastroenterol* 2014;**20**(41):15216–32.
588 31. Rieder F, Karrasch T, Ben-Horin S, Schirbel A, Ehehalt R, Wehkamp J, et al. Results of the
589 2nd Scientific Workshop of the ECCO (III): Basic mechanisms of intestinal healing. *J Crohn's*
590 *Colitis* 2012;**6**(3):373–85.
591 32. Leoni G, Neumann PA, Sumagin R, Denning TL, Nusrat A Wound repair: Role of immune-
592 epithelial interactions. *Mucosal Immunol* 2015;**8**(5):959–68.
593 33. Slater TW, Finkielsztein A, Mascarenhas LA, Mehl LC, Butin-Israeli V, Sumagin R Neutrophil
594 Microparticles Deliver Active Myeloperoxidase to Injured Mucosa To Inhibit Epithelial
595 Wound Healing. *J Immunol* 2017;**198**(7):2886 LP-2897.
596 34. Mottawea W, Chiang CK, Mühlbauer M, Starr AE, Butcher J, Abujamel T, et al. Altered
597 intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat Commun*
598 2016;**7**.
599 35. Hong SN, Joung JG, Bae JS, Lee CS, Koo JS, Park SJ, et al. RNA-seq Reveals Transcriptomic
600 Differences in Inflamed and Noninflamed Intestinal Mucosa of Crohn's Disease Patients
601 Compared with Normal Mucosa of Healthy Controls. *Inflamm Bowel Dis* 2017;**23**(7):1098–
602 108.
603 36. Hsieh S-Y, Shih T-C, Yeh C-Y, Lin C-J, Chou Y-Y, Lee Y-S Comparative proteomic studies
604 on the pathogenesis of human ulcerative colitis. *Proteomics* 2006;**6**(19):5322–31.
605 37. Sifroni KG, Damiani CR, Stoffel C, Cardoso MR, Ferreira GK, Jeremias IC, et al.
606 Mitochondrial respiratory chain in the colonic mucosal of patients with ulcerative colitis. *Mol*
607 *Cell Biochem* 2010;**342**(1–2):111–5.
608 38. Nazli A, Yang PC, Jury J, Howe K, Watson JL, Söderholm JD, et al. Epithelia under Metabolic
609 Stress Perceive Commensal Bacteria as a Threat. *Am J Pathol* 2004;**164**(3):947–57.

610 39. Zhang K, Kaufman RJ From endoplasmic-reticulum stress to the inflammatory response.
611 *Nature* 2008;**454**(7203):455–62.
612 40. Deuring JJ, de Haar C, Koelewijn CL, Kuipers EJ, Peppelenbosch MP, van der Woude CJ
613 Absence of ABCG2-mediated mucosal detoxification in patients with active inflammatory
614 bowel disease is due to impeded protein folding. *Biochem J* 2012;**441**(1):87–93.
615 41. Cao SS Endoplasmic reticulum stress and unfolded protein response in inflammatory bowel
616 disease. *Inflamm Bowel Dis* 2015;**21**(3):636–44.
617 42. Heijmans J, Van Lidth de Jeude JF, Koo BK, Rosekrans SL, Wielenga MCB, Van De Wetering
618 M, et al. ER Stress Causes Rapid Loss of Intestinal Epithelial Stemness through Activation of
619 the Unfolded Protein Response. *Cell Rep* 2013;**3**(4):1128–39.
620 43. Berger E, Rath E, Yuan D, Waldschmitt N, Khaloian S, Allgäuer M, et al. Mitochondrial
621 function controls intestinal epithelial stemness and proliferation. *Nat Commun* 2016;**7**.
622 44. Neal MD, Richardson WM, Sodhi CP, Russo A, Hackam DJ Intestinal stem cells and their
623 roles during mucosal injury and repair. *J Surg Res* 2011;**167**(1):1–8.
624 45. Cao SS, Zimmermann EM, Chuang B-M, Song B, Nwokoye A, Wilkinson JE, et al. The
625 unfolded protein response and chemical chaperones reduce protein misfolding and colitis in
626 mice. *Gastroenterology* 2013;**144**(5):989–1000.e6.
627 46. Novak EA, Mollen KP Mitochondrial dysfunction in inflammatory bowel disease. *Front Cell*
628 *Dev Biol* 2015;**3**(October):1–18.
629 47. Wang A, Keita Å V., Phan V, McKay CM, Schoultz I, Lee J, et al. Targeting mitochondria-
630 derived reactive oxygen species to reduce epithelial barrier dysfunction and colitis. *Am J*
631 *Pathol* 2014;**184**(9):2516–27.
632 48. Rifai N, Gillette MA, Carr SA Protein biomarker discovery and validation: The long and
633 uncertain path to clinical utility. *Nat Biotechnol* 2006;**24**(8):971–83.
634 49. Stevens JR, Herrick JS, Wolff RK, Slattery ML Power in pairs: assessing the statistical value of
635 paired samples in tests for differential expression. *BMC Genomics* 2018;**19**(1):953.
636 50. Chang J, Chance MR, Nicholas C, Ahmed N, Guilmeau S, Flandez M, et al. Proteomic changes
637 during intestinal cell maturation in vivo. *J Proteomics* 2008;**71**(5):530–46.

638 51. Rath E, Moschetta A, Haller D Mitochondrial function — gatekeeper of intestinal epithelial
639 cell homeostasis. *Nat Rev Gastroenterol Hepatol* 2018;**15**(8):497–516.
640



641

642 **Figure 1.** Ileal and colonic ulcer edges of Crohn's disease patients are infiltrated by
643 inflammatory cells. [A] Neutrophils, [B] lymphocytes and plasmocytes infiltration evaluated
644 by histopathological score in ulcer edge (U) and control tissue (C) of ileum and colon. The
645 following scores were used: 0 (none), 1 (light), 2 (moderate) and 3 (severe). Significance was
646 tested using the paired t-test or the Wilcoxon matched-pairs signed-ranks test as appropriate.
647 *$p<0.05$, **$p<0.05$.

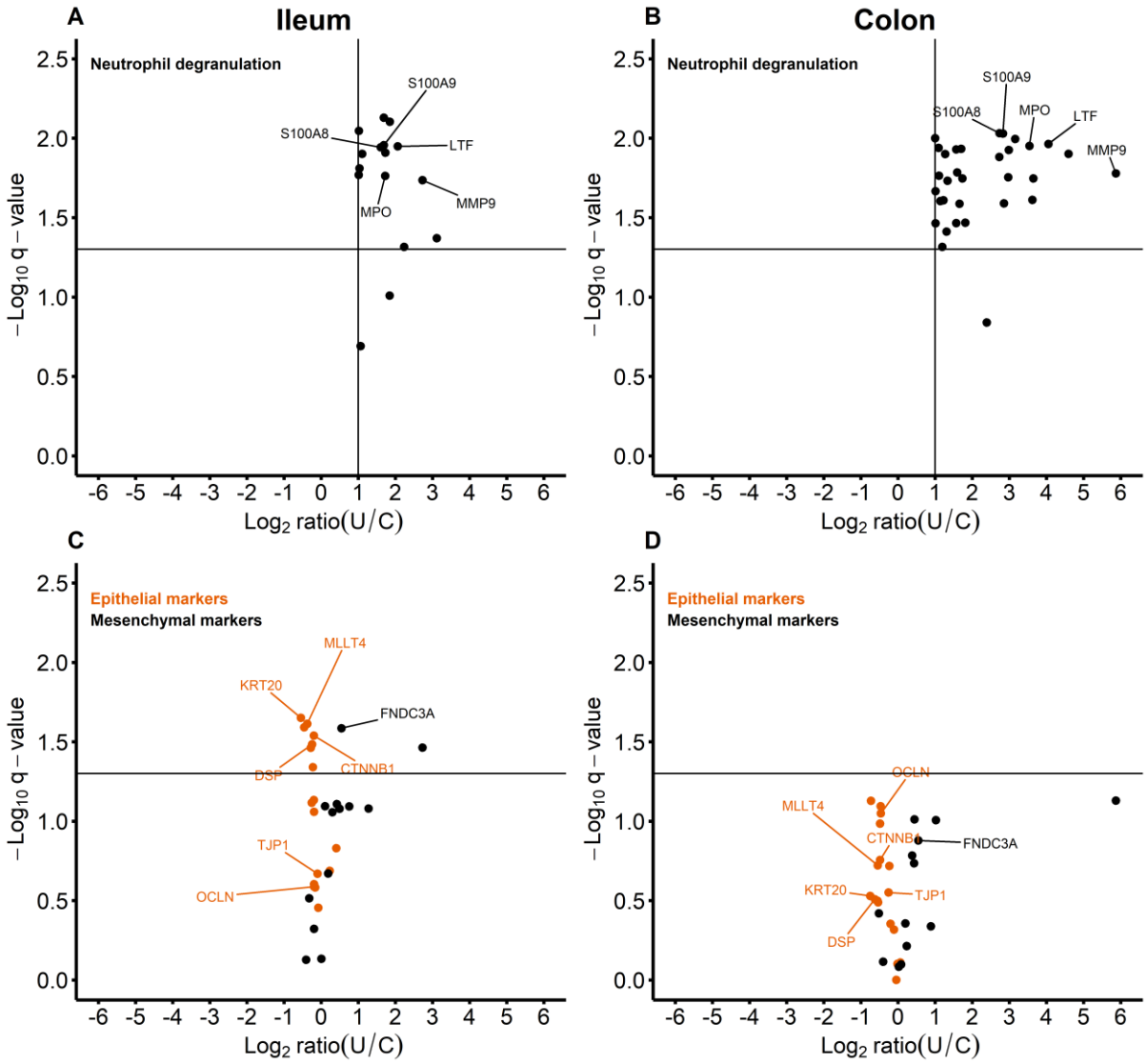


Figure 2. Markers of neutrophil degranulation and epithelial–mesenchymal transition are distinctly affected in ileal and colonic ulcer edge of Crohn's disease patients. Protein differential abundances between ulcer edge (U) and control tissue (C) were represented by plotting the -Log_{10} q-value (U vs. C) against the relative abundance expressed as the Log_2 of U/C. [A] Ileal and [B] colonic distribution of the proteins involved in neutrophil degranulation. The selection of these proteins has been performed by using the Reactome database R-HSA-6798695. Among these proteins, only those with the highest fold change (U/C>2) and an occurrence of minimum 50% in the two groups (U and C) have been selected and represented in this graph. [C] Ileal and [D] colonic distribution of the epithelial–mesenchymal transition markers. The selection of these proteins has been performed by searches in the literature (see results). The horizontal line

represents the significance threshold (q-value=0.05). The vertical lines represent the fold change (U/C) equal to 2. CTNNB1: Catenin beta-1; DSP: Desmoplakin; FNDC3A: Fibronectin type-III domain-containing protein 3A; KRT20: Keratin, type I cytoskeletal 20; LTF: lactotransferrin; MLLT4: Afadin; MMP9: matrix metalloproteinase-9; MPO: Myeloperoxidase; OCLN: occludin; S100A8 and S100A9: calprotectin subunits; TJP1: Tight junction protein ZO-1.

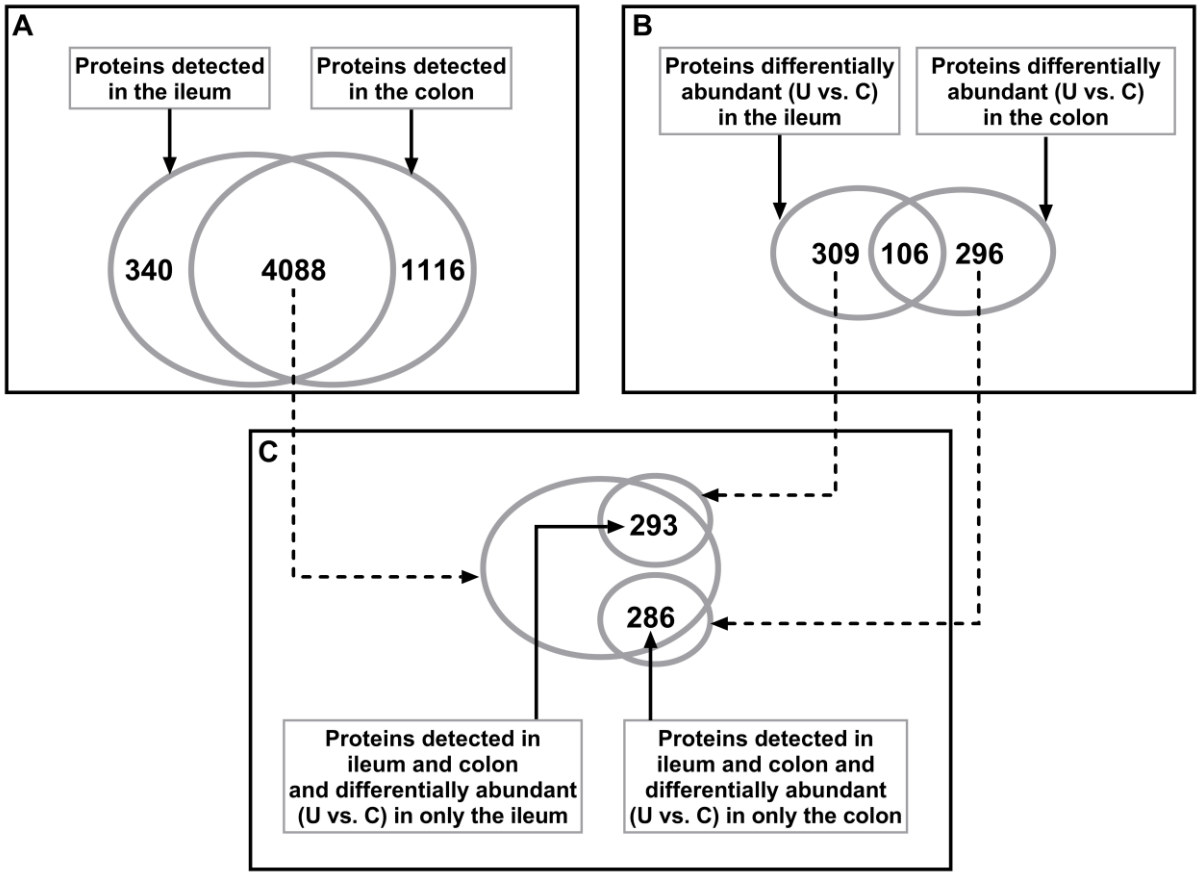
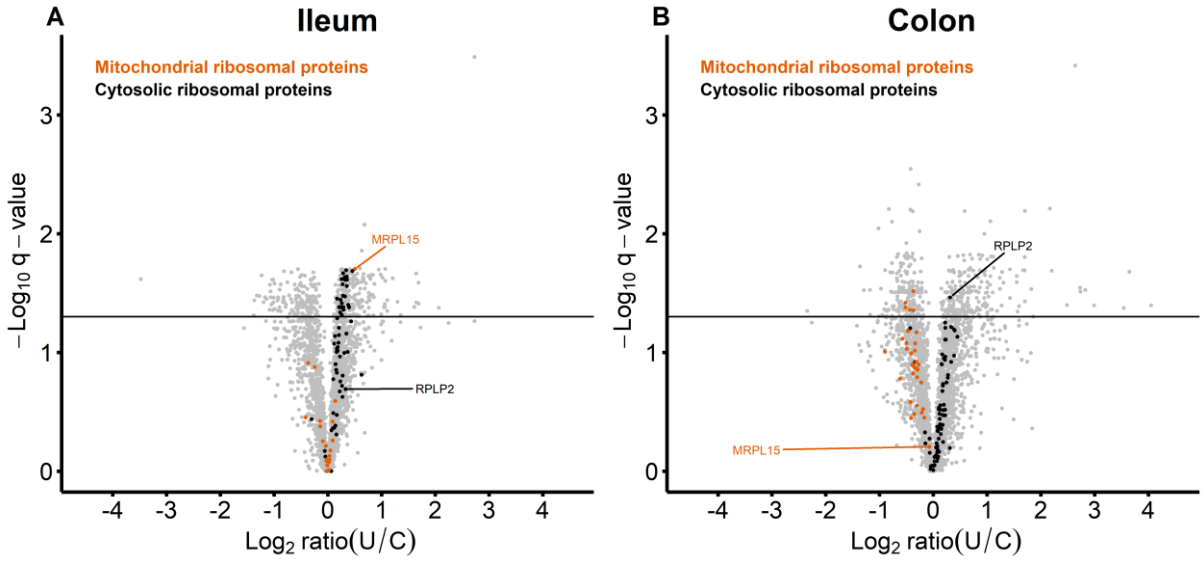


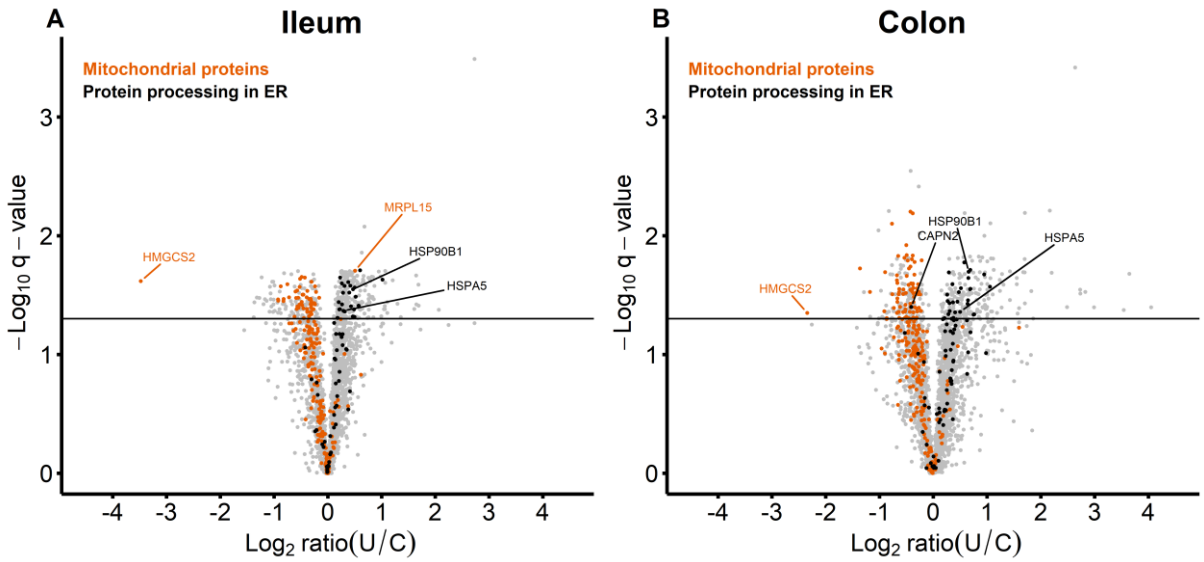
Figure 3. Selection method of the proteins similarly or specifically affected in the ileal and/or the colonic ulcer edge of Crohn's disease patients. [A] Venn diagram showing the number of proteins detected in the ileum and the colon (n=4088). [B] Venn diagram showing the number of proteins: 1) differentially abundant (U vs. C) in the ileum and the colon (n=106); 2) differentially abundant (U vs. C) in only the ileum (n=309) or only the colon (n=296). [C] Venn diagram showing the number of proteins detected in the ileum and the colon and specifically affected in ileal (n=293) or colonic (n=286) U. These lists of proteins were obtained by merging

673 the list of proteins generated in [A] and [B]. C: control tissue; U: ulcer edge tissue.



674

675 **Figure 4.** Ribosomal proteins are distinctly affected in ileal and colonic ulcer edge of Crohn's
676 disease patients. [A] Ileum and [B] colon protein differential abundances between ulcer edge
677 (U) and control tissue (C) were represented by plotting the -Log_{10} q-value (U vs. C) against the
678 relative abundance expressed as the Log_2 of U/C. Only proteins with 100% occurrence in U and
679 C have been compared and represented in this graph. Ribosomal proteins have been selected
680 through the protein name. The horizontal line represents the significance threshold (q-
681 value=0.05). MRPL15: 39S ribosomal protein L15, mitochondrial; RPLP2: 60S acidic
682 ribosomal protein P2.



683

Figure 5. Mitochondrial proteins and protein of *protein processing in ER* are similarly affected in ileal and colonic ulcer edge of Crohn's disease patients. [A] Ileum and [B] colon protein differential abundances between ulcer edge (U) and control tissue (C) were represented by plotting the -Log_{10} q-value (U vs. C) against the relative abundance expressed as the Log_2 of U/C. Only proteins with 100% occurrence in U and C have been compared and represented in this graph. Mitochondrial proteins and proteins involved in *protein processing in ER* have been selected through the protein name and the KEGG pathway database hsa04141, respectively. The horizontal line represents the significance threshold (q-value=0.05). CAPN2: calpain-2 catalytic subunit; HMGCS2: hydroxymethylglutaryl-CoA synthase; HSPA5: heat shock 70 kDa protein 5; HSP90B1: heat shock protein 90 kDa beta member 1; MRPL15: 39S ribosomal protein L15, mitochondrial.

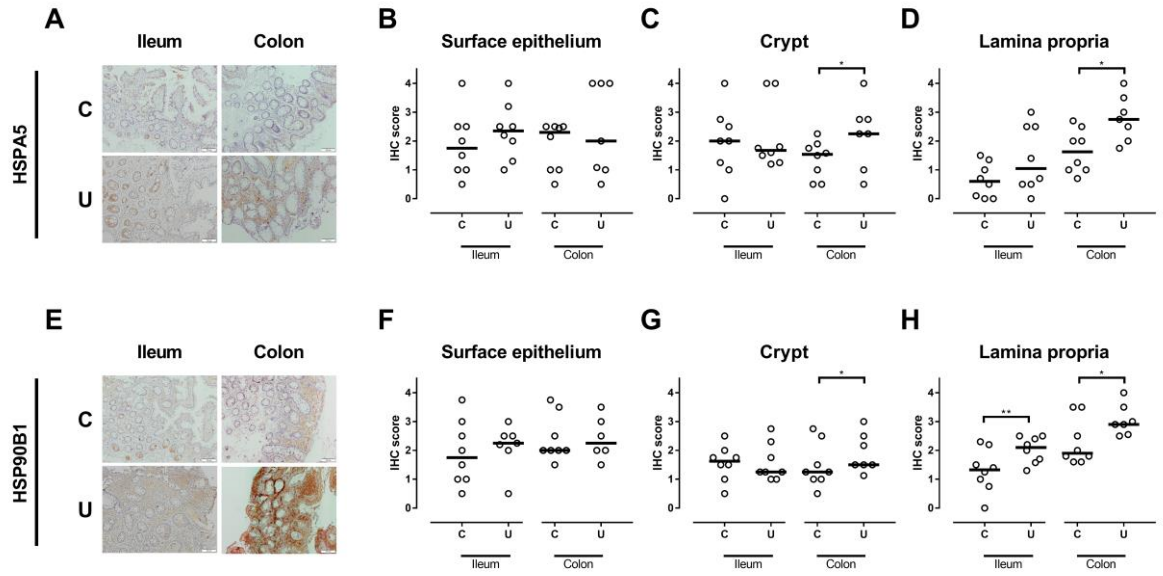


Figure 6. Activation of ER stress in ileal and colonic ulcer edge of Crohn's disease patients. [A] Representative staining of HSPA5 in ulcer edge (U) and control tissue (C) of ileum and colon. [B] Surface epithelium, [C] crypt and [D] *lamina propria* staining scores of HSPA5, horizontal lines indicate the median. [E] Representative staining of HSP90B1 in U and C of ileum and colon. [F] Surface epithelium, [G] crypt and [H] *lamina propria* staining scores of

701 HSP90B1, horizontal lines indicate the median. Significance was tested using the paired t-test
702 or the Wilcoxon matched-pairs signed-ranks test as appropriate. *p<0.05, **p<0.05.

703

Table 1. Patients' characteristics

	Patients with ileal ulcers	Patients with colonic ulcers
Patient, n (m/f)	8 (1/7)	8 (4/4)
Age, median years (min-max)	37.5 (30-68)	38.0 (30-43)
Disease duration, median years (min-max)	9 (1-41)	12.5 (0-34)
Smoking		
Yes	4	4
Former	3	2
No	1	2
Disease location at the time of the endoscopy		
Ileal	5	0
Colonic	0	3
Ileocolonic	3	5
Medication*		
Anti-TNFα	3	1
Anti-α4β7 integrin	0	2
Antimetabolites	2	1
Antibiotics	0	0
Corticoids	2	3
None	3	4

*Some patients may have several medications

704

Table 2. Description of the proteomic dataset

	Ileum	Colon
Mean (SD) MS/MS in samples	197463 (3352)	209955 (4781)
Mean (SD) MS/MS identified in samples	58735 (6112)	60528 (4209)
MS/MS identification rate (%)	29.7	28.8
Number of unique peptides	36632	46741
Mean (SD) unique peptides in samples	26661 (2046)	34455 (1597)
Number of proteins identified and quantified	4428	5204
Number of proteins differentially abundant (U vs. C)	415	402

C: control tissue; SD: standard deviation; U: ulcer edge tissue

705