The proteome of ulcerative colitis

Functional analyses of the active disease and the remission state in comparison with healthy controls

—

Armin Schniers

A dissertation for the degree of Philosophiae Doctor – April 2019

For Mary Ann, Emma and Carrie

Acknowledgements

First of all, I want to thank my main supervisor Terkel Hansen for the guidance and the professional and personal support throughout the journey. Thank you for the trust you had in me and my ideas, and thank you for the opportunity to be part of the project. I also want to express my gratitude to my co-supervisor Jon Florholmen, who coordinates the Advanced Study of Inflammatory Bowel disease (ASIB), of which this work is a part. Jon and his clinical expertise were indispensable in all stages of this project. I am grateful to Rasmus Goll from UNN for his contributions he made to this project with his clinical knowledge. Thank you to Renate Meyer, the study nurse of the ASIB project.

I thank all my colleagues at IFA and especially the Natural Products and Medicinal Chemistry Research Group. Thank you Martina and Trude, our engineers in the group who keep things going and contribute a lot to the great work environment in our group. I am grateful to all former and current colleagues in the group and the department.

Thanks to my co-authors Yvonne, Guro, Sveinung, Ruth, Chris, and Endre. Special thanks to Yvonne for extensive discussions on proteomics sample preparation and data handling.

The project would not have been possible without TUPP, the proteomics platform at UiT. I thank Jack, Ilona and Toril for keeping the instruments up and running. I also thank them for advice and discussions on sample preparation, operating the mass spectrometer, and data handling. I furthermore want to thank the Norwegian PhD School of Pharmacy and NORBIS for travel grants to courses and conferences.

Finally yet importantly, I want to thank my family. Thanks to my parents Christine and Hermann as well as my sister Kerstin for the continuous support. I am deeply grateful for the love and support from Emma, Carrie, and my wife Mary Ann.

Tromsø, 11.04.2019

Armin Schniers

Table of Contents

List of Papers

Paper I

The Proteome of Ulcerative Colitis in Colon Biopsies from Adults - Optimized Sample Preparation and Comparison with Healthy Controls [1]

Armin Schniers, Endre Anderssen, Christopher Graham Fenton, Rasmus Goll, Yvonne Pasing, Ruth Hracky Paulssen, Jon Florholmen and Terkel Hansen

PROTEOMICS – Clinical Applications 2017, 11(11-12):1700053.

Paper II

Ulcerative colitis: functional analysis of the in-depth proteome [2]

Armin Schniers, Rasmus Goll, Yvonne Pasing, Sveinung Wergeland Sørbye, Jon Florholmen and Terkel Hansen

Clin Proteomics 2019, 16:4.

Paper III

The proteome of ulcerative colitis in remission – functional differences in comparison with healthy controls and the active disease

Armin Schniers, Rasmus Goll, Sveinung Wergeland Sørbye, Jon Florholmen and Terkel Hansen

Manuscript

Paper IV

Prediction of the 1-year outcome for Ulcerative Colitis from the proteomic profile of treatment naïve patients

Armin Schniers, Rasmus Goll, Sveinung Wergeland Sørbye, Hans Stenlund, Jon Florholmen, and Terkel Hansen

Manuscript

Summary

This work elucidates the properties of human colon mucosa in ulcerative colitis (UC) on the protein level. We developed an optimized sample preparation method of colon mucosa biopsies for bottom-up proteomics. This method applies a lysis buffer of 8M urea, 5% sodium deoxycholate, and 100 mM TEAB. It showed superior reproducibility and identification numbers. With this method, we acquired the to-date deepest proteome data set from colon mucosa biopsies with more than 8,000 quantified proteins. The investigated states comprise active UC, remission from UC, and healthy controls. We performed enrichment analyses of GO annotations among differently abundant proteins.

We observed increased abundances among proteins related to the immune system and to protein processing in the endoplasmic reticulum (ER) in active UC compared to healthy controls. Lower abundant are metallothioneins, fibrillary collagens, bile acid transport proteins, carbonic anhydrases, and proteins related to nutrient, energy, and xenobiotic metabolism.

In general, the remission state seems to be a blend of healthy and diseased state. We characterized the remission state based on the proteins that were significantly different abundant in remission compared to active UC and/or healthy controls. A small fraction of these proteins (associated functions: hormones, vitamins, lipoproteins, muscle) is higher abundant in remission than in both active UC and healthy controls. Most proteins (associated functions: immune system, protein processing, collagen) show similar abundances in remission as in healthy controls. About one fourth of the remission abundances (associated functions: nutrient and energy metabolism, PPAR signaling) was between those in active UC and healthy controls and significantly different from both. Approximately one eighth of the proteins was at similar levels as in active UC (associated functions: immunoglobulins, metallothioneins, prostaglandin metabolism). Protein abundances that are not at equal levels as in healthy controls may contribute to relapses and symptoms in remission.

In addition to improving our understanding of UC, our findings have some clinical implications. Several functions apart from the inflammation could be readily addressable with medication. For instance, late metabolites such as the ketone body β-hydroxybutyrate could potentially circumvent some impairments in the nutrient metabolism and thereby ameliorate the energy deficiency in colonocytes. It may furthermore lead to metallothionein induction.

Our results have implications for the use of biomarkers in UC. The abundance changes of the routinely used calprotectin proteins and lactotransferrin are representative for only a small minority of differently abundant proteins. An additional assessment of more representative proteins may be useful. We furthermore present a model for the prediction of the 1-year-outcome that could bring great benefit for clinical decision-making.

List of Figures

List of Tables

Abbreviations

2-DE	Two-dimensional gel electrophoresis
5-ASA	5-aminosalicylic acid
6TGN	6-thioguanine nucleotide
ABP1	Amiloride binding protein 1
AIEC	Adherent-Invasive *Escherichia coli*
ANOVA	Analysis of variance
AP-1	Activator protein 1
ASIB	Advanced Systems Biology Analysis of Early events and Clinical Outcomes in Inflammatory Bowel Disease
AZA	Azathioprine
BCA	Bicinchoninic acid
CD	Crohn's disease
CID	Collision-induced dissociation
CMV	Cytomegalovirus
COX	Cyclooxygenase
CRP	C-reactive protein
CTAB	Cetyltrimethylammonium bromide
DSS	Dextran sulfate sodium
DTT	Dithiothreitol
EIM	Extraintestinal manifestation
ER	Endoplasmic reticulum
ETC	Electron transport chain
Etk	Endothelial/epithelial tyrosine kinase
FA	Formic acid
FABP	Fatty acid binding protein
FASP	Filter-aided sample preparation
FDR	False discovery rate
GC	Glucocorticoid
GI tract	Gastrointestinal tract
GO	Gene ontology
GPM DB	Global Proteome Machine Database
GR	Glucocorticoid receptor
HCD	Higher-energy collisional dissociation
HFBA	Heptafluorobutyric acid
HILIC	Hydrophilic interaction chromatography
HLA	Human leukocyte antigen
IAA	Iodoacetamide
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
IUPAC	International Union of Pure and Applied Chemistry
IκBα	NF-κB inhibitor alpha
JAK	Janus kinase

LC	Liquid chromatography
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LPS	Lipopolysaccharide
LT	Lymphotoxin
MAdCAM-1	Mucosal addressin-cell adhesion molecule 1
MMP	Matrix metalloproteinase
MP	6-mercaptopurine
mRNA	Messenger RNA
MRS	Modified Riley Score
MS	Mass spectrometry
MT	Metallothionein
mTNF	Transmembrane TNF
NF-κB	Nuclear factor-κB
NO	Nitric oxide
NOD2	Nucleotide-binding oligomerization domain 2
NOS	Nitric oxide synthase
NOS2	Inducible nitric oxide synthase
PAGE	Polyacrylamide gel electrophoresis
PCA	Principal component analysis
PPAR	Peroxisome proliferator-activated receptor
PTGES	Prostaglandin E synthase
PTM	Posttranslational modification
RHI	Robarts Histopathology index
RIP1	Receptor interacting protein-1
ROS	Reactive oxygen species
RP-IAA	Restorative proctocolectomy with ileal pouch-anal anastomosis
SASP	Sulfasalazine
SCFA	Short chain fatty acid
SCX	Strong cation exchange
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulfate
SL	Sodium laurate
SODD	Silencer of death domain
sRNA	Small RNA
TCA cycle	Tricarboxylic acid cycle
TEAB	Triethylammonium bicarbonate buffer
TFA	Trifluoroacetic acid
T_H cell	T helper cell
TMT	Tandem mass tag
TNF	Tumor necrosis factor
TNFR1	Tumor necrosis factor receptor 1
TNFR2	Tumor necrosis factor receptor 2

TRADD	TNFR-associated death domain protein
TRAF2	TNFR-associated factor 2
T$_{reg}$ cell	Regulatory T cell
tRNA	Transfer RNA
TYK	Tyrosine kinase
UC	Ulcerative colitis
UCDAI	Ulcerative Colitis Disease Activity Index
UCEIS	Ulcerative Colitis Endoscopic Index of Severity
UPR	Unfolded protein response
VEGFR2	Vascular endothelial growth factor receptor 2

1 Introduction

1.1 Inflammatory Bowel Disease

Ulcerative colitis (UC) and Crohn's disease (CD) share clinical manifestations. Hence, they are often summarized as inflammatory bowel disease (IBD), even though they are distinct pathophysiological entities.[3] Both diseases are chronic inflammatory intestinal disorders, but they differ in the location and nature of the inflammatory changes.[4] UC presents with continuous, non-transmural (i.e. limited to epithelial lining) inflammation of the mucosa that is limited to the large intestine. In 95% of the cases UC affects the rectum, from where it can extend to more proximal parts of the large intestine in a continuous and circumferential manner.[5] CD on the other hand presents with transmural (i.e. it affects the entire bowel wall) and discontinuous inflammation that can affect all segments of the gastrointestinal tract from mouth to anus.[6] UC is less prone to complications than CD and its course is in many patients comparatively mild.[3]

1.1.1 Epidemiology and role of environmental factors

At the turn of the 21st century IBD has become a global disease.[7] The worldwide burden of IBD is increasing over time.[8] After the prevalence in North America and Western Europe increased since the mid of the 20th century [3, 9], the incidence in these regions has by now stabilized.[5, 7] The incidence in newly industrialized countries that are becoming more westernized is still increasing.[7] For instance in China, IBD has changed from being rare to now being responsible for one fourth of the occupied gastroenterological and colorectal surgical hospital beds.[10] When IBD is newly diagnosed in a population, UC precedes CD and appears with a higher incidence.[3]

The highest annual incidence for UC is 24.3 per 100 000 person-years in Europe, 19.2 per 100,000 person-years in North America, and 6.3 per 100 000 person-years in Asia (see Figure 1).[8] The highest prevalence values were reported for Europe (505 per 100 000 in Norway) and North America (286 per 100,000 in the USA).[11] The peak-age for disease onset is 30-40 years [11, 12] with a slight gender predominance that rests with men.[12, 13] The highest annual incidence for CD is 12.7 per 100,000 person-years in Europe, 20.2 per 100,000 person-years in North America, and 5.0 per 100,000 person-years in Asia and the Middle East.[8] Like for UC, the highest reported prevalence values for CD are in Europe (322 per 100,000 in Germany) and North America (319 per 100,000 in Canada).[8]

Relatives of IBD patients and Jewish populations show a higher rate of developing CD and UC.[14] Genome analyses identified 200 risk loci, of which most contribute to both ulcerative colitis and Crohn's disease.[11] However, around 10 % of IBD can be explained by genetic variances (13.5 % for CD and 7.5 % for UC), suggesting a more complex pathogenesis.[15, 16]
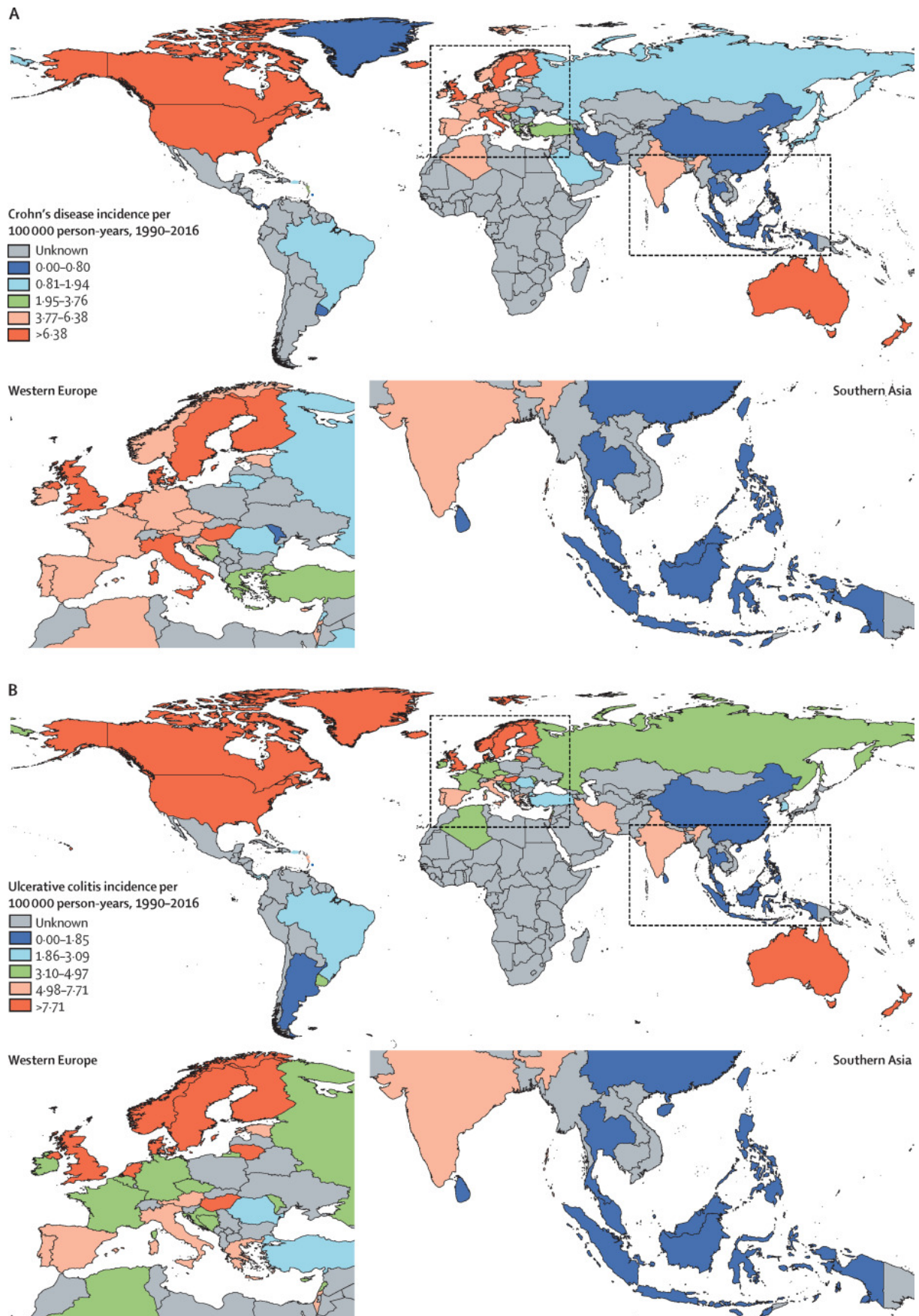
Figure 1 - Worldwide incidences of the IBD forms UC and CD. Used with permission from [7]

Both diseases are disorders of modern societies. The incidence of IBD increased strongly since the mid of the 20th century[3, 9], for which environmental factors are primarily responsible.[9] A possible explanation is that the body is not properly adapted to the rapid environmental changes. In an attempt to adjust to those changes with altered neuronal and endocrine responses a complex disruption of homeostasis occurs.[9, 17] Associations with IBD have been shown for several, seemingly unrelated, environmental factors. Most supported by evidence are associations of IBD with smoking and the enteric bacterial flora.[9]

The role of smoking in IBD is ambiguous, as it has protective effects in UC, but increases the risk of developing CD and worsens its clinical course.[9] Former smoking has been identified as a strong risk factor for UC development, while active smoking decreases the risk to develop the disease and the severity of the disease course.[11] The mechanisms for the differential effects are not clear. However, many components of cigarette smoke possibly have immunomodulatory effect, and nicotine inhibits both innate and adaptive immune responses.[18]

Evidence suggests that both forms of IBD result from an inappropriate immune reaction to intestinal microbes[19], probably to those that form the normal enteric flora.[9] The role of the microbiota is supported by the observation that most animal models of IBD do not develop inflammation, if the animals are kept in a germ-free environment.[20] It is further supported by beneficial effects of antibiotics on CD and to a lesser degree also on UC.[9] Increasing hygiene in developed countries and the resulting decreased microbial burden are believed to contribute to the rising prevalence of IBD.[21]

Further factors associated with IBD are stress, diet, drugs (primarily oral contraceptives and nonsteroidal anti-inflammatory drugs), latitude (IBD more common in North than in the South of Europe and North America), intestinal permeability, appendectomy (negative association with UC), occupation, as well as social, economic and educational status.[9]

1.1.2 Pathogenesis of IBD

UC has a complex and not fully understood pathogenesis that involves genetic factors, environmental influences, epithelial barrier defects, and a dysregulated immune response.[9, 11, 22]

The gastrointestinal (GI) tract has a mean surface of 32 m^2, of which 2 m^2 belong to the large intestines.[23] The highest bacterial populations in humans occur in the large intestines and reach 10^{12} bacteria/cm^3. The bacterial flora of the large intestines comprises more than 1,000 species. The compositions vary from person to person.[24] As a result of the GI tract's large contact area with bacteria and their high number and variety, the GI tract is a central site to the immune system. In the healthy GI tract, the innate and the adaptive immune systems are balanced in complex interactions with the bacterial flora. Environmental factors trigger a

disruption of this balance in genetically susceptible individuals, which results in the intestinal inflammation observed in IBD.[22] These environmental factors transiently break the mucosal barrier, trigger an immune reaction, or alter the balance between pathogenic and beneficial microbes.[25]

Findings from human genetics, animal models, basic science and clinical trials indicate that UC and CD are heterogeneous diseases, in which various genetic abnormalities cause T cells to react overly aggressive to commensal bacterial flora. It is widely believed that the manifestation of IBD results when four different components intersect: An (1) overly aggressive T cell response to a (2) subset of commensal bacteria and (3) triggering environmental factors cause the onset or reactivation of the disease in (4) genetically susceptible individuals (see Figure 2).[25]
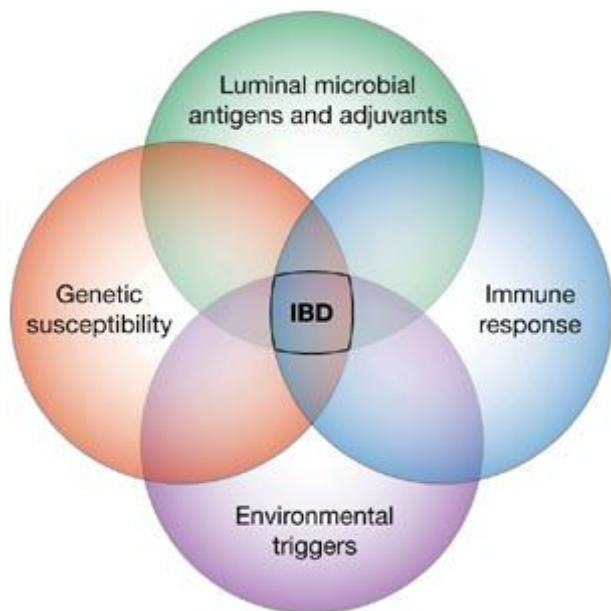
Figure 2 - Four components lead to IBD. Used with permission from [25]

The UC pathogenesis has traditionally been identified as a T_H2-like disease, while CD has been linked to T_H1 cells. More recently, a role of an imbalance between T_H17 cells and regulatory T (T_{reg}) cells in IBD has been suggested.[22, 26] T_H2, T_H1 and T_H17 cells are effector T helper (T_H) cells and subgroups of CD4+ T cells. They secrete characteristic cytokines. The cytokines secreted by T_H2 cells comprise interleukin-4, 5, and 13, whereas T_H1 cells secrete interferon-γ and TNF-α and T_H17 cells secrete interleukin-17. As effector T helper cells, T_H2, T_H1 and T_H17 cells are crucial in protecting the host from pathogens and from excessive entry of luminal bacteria, but their proper regulation is a prerequisite for maintaining intestinal immune homeostasis.[19]

Specific hypotheses suggest roles in the IBD pathogenesis for persistent infections with specific pathogens, dysbiosis, mucosal barrier dysfunctions, and defective microbial clearance.[25]

A role of persistent infections with specific pathogens was suggested for various bacteria. *Mycobacterium avium paratuberculosis* was the first bacterium implicated as an IBD pathogen possibly causing CD.[27, 28] Subsequently further potential IBD pathogens were suggested, including Adherent-Invasive *Escherichia coli* (AIEC) and various Helicobacter species. AIEC is particularly of interest, because it can strongly stimulate the secretion of TNF-α, which is a key cytokine in IBD.[28] The risk for the development of both UC and CD is increased in patients who experienced episodes of Salmonella or Campylobacter gastroenteritis.[29] The "cold chain hypothesis" suggests that domestic refrigeration is a risk factor for CD due to psychrotrophic bacteria, which can grow slowly at low temperatures, e.g. *Yersinia enterocolitica*, *Listeria monocytogenes*, and *Clostridium botulinum*.[30] A role in the IBD pathogenesis has furthermore been suggested for fungi (*Saccharomyces cerevisiae*, *Candida albicans*) and viruses (especially cytomegalovirus (CMV), but also parvovirus B19, norovirus, and Epstein–Barr virus). On the contrary, in line with the "IBD hygiene hypothesis" a lack of exposure to helminthic parasites may predispose to CD development.[31, 32] Further research is necessary to attribute a causative role of specific pathogens to the IBD pathogenesis, because it is often unknown whether a pathogen associated with IBD causes the disease, or whether the disease causes conditions that are suitable for proliferation of the respective pathogen.[28]

Dysbiosis describes a state of imbalance between aggressive and beneficial microbes, which could lead to intestinal conditions that trigger chronic inflammation in susceptible hosts. An over-representation of commensal bacteria such as *Escherichia coli*, *Bacteroides*, *Enterococcus* and *Klebsiella* species seems to favor inflammation, whereas *Lactobacillus* and *Bifidobacterium* species probably have protective effects.[25, 33] Dietary components can alter the intestinal bacterial composition. For instance, non-absorbable carbohydrates like inulin and fructose oligosaccharides support the growth of the beneficial *Bifidobacterium* and *Lactobacillus* species. Iron, which is a common food additive in Western diets, enhances the growth and virulence of intracellular bacteria.[28] A role of iron in IBD is further supported by the finding that supplementation with iron enhances inflammation in rats with dextran sulfate sodium (DSS)-induced colitis.[34]

A dysfunction of the mucosal barrier could increase the uptake of luminal antigens and adjuvants, which could in turn overwhelm the suppressive capacities of the mucosal immune system.[25] The mucus layer in healthy individuals consists of two substructures, one tightly adherent inner layer that is mostly sterile, and a loosely adherent outer layer that is good for microbial growth. In IBD, particularly CD, an increase of bacteria associated with the adherent mucus layer occurs, e.g. mucosa-associated Escherichia coli.[35] An alternative mechanism is that an impairment of repair capabilities in the epithelium could potentiate damage from infections and agents like NSAIDs. One hypothesis is that IBD results from such damage and a subsequent failure to down-regulate the resulting immune reaction.[25]

Defective microbial clearance (i.e. defective bacterial killing) is a possible factor in the pathogenesis of CD.[25] This is supported by the finding that approximately one-third of CD patients have loss-of-function mutation in the NOD2 (synonym: CARD15) gene. Nucleotide-binding oligomerization domain 2 (NOD2) is an intracellular receptor for muramyl dipeptides, which is a component of bacterial peptidoglycan. Upon activation, NOD2 leads to NF-κB-dependent pro-inflammatory cytokine expression[36], α-defensin expression[37], and possibly killing of intracellular bacteria.[25, 38]

NOD2 was the first susceptibility gene identified for CD. Further genes associated with CD are related to autophagy, which is a process for clearing unnecessary components. Among those genes are for instance those coding for Autophagy-related protein 16-1 (ATG16L1)[39] and Immunity-related GTPase family M protein (IRGM).[19, 40] Among the gene loci associated with UC are human leukocyte antigen (HLA) and several that contribute to the mucosal barrier function, e.g. the genes coding for Cadherin-1 (CDH1) and Hepatocyte nuclear factor 4-alpha (HNF4A).[11, 41] Genetic associations with both UC and CD have been shown for multiple genes of the interleukin-23–T_H17 pathway, which is crucial for the function of T_H17 cells. The interleukin-23–T_H17 pathway mediates intestinal inflammation and defense against microbes. The genes coding for Interleukin-23 receptor (IL23R), Interleukin-12 subunit beta (IL12B) and Signal transducer and activator of transcription 3 (STAT3) are associated with both UC and CD, whereas the gene coding for C-C chemokine receptor type 6 (CCR6) is associated only with CD.[19, 42]

1.1.3 Pathophysiology of Ulcerative Colitis

Infiltration of innate immune cells (neutrophils, macrophages, cytotoxic T cells, dendritic cells) and adaptive immune cells (B cells and T cells, especially pro-inflammatory T cell subgroups) into the intestinal lamina propria is the hallmark of active IBD. The increased number and an activation of these immune cells result in elevated secretion of pro-inflammatory cytokines (TNF-α, interleukin-1β, interleukin-6, interleukin-12, interleukin-23) and chemokines. The increased chemokine levels cause further infiltration of leukocytes, leading to a sustained cycle of inflammation.[19]

Tumor necrosis factor (TNF) is a dominant cytokine in ulcerative colitis.[43, 44] It is also referred to as TNF-α. However, this specification became obsolete, because the term TNF-β for the cytokine lymphotoxin (LT) was abandoned after its two forms LT-α and LT-β were identified.[45] Macrophages, monocytes and neutrophils secrete TNF in ulcerative colitis.[46-48] The TNF signaling is complex, but the known signaling pathways involve Tumor necrosis factor receptor 1 (TNFR1) or Tumor necrosis factor receptor 2 (TNFR2), see Figure 3. Unstimulated TNFR1 is associated with silencer of death domain (SODD), which is released upon TNF binding. This allows binding of another death domain containing protein, TNFR-associated death domain protein (TRADD). TRADD then recruits two further proteins, receptor interacting protein-1 (RIP1) and TNFR-associated factor 2 (TRAF2).[49] The complex is then

internalized[50] and the TRADD-RIP1-TRAF2 complex gets released from TNFR1. The TRADD-RIP1-TRAF2 interacts with further proteins in the cell, eventually resulting effects like mediation of cell survival and pro-inflammatory signals through NF-κB and Activator protein 1 (AP-1) and apoptosis through caspase 3. The signaling by TNFR2 is less well defined, but seems to involve TRAFs. TNFR2 can further activate endothelial/epithelial tyrosine kinase (Etk), which has functions in cell adhesion, migration, survival and proliferation. In endothelial cells, TNFR2, Etk, and vascular endothelial growth factor receptor 2 (VEGFR2) form a complex upon activation by TNF. In this complex, Etk and VEGFR2 phosphorylate one another, which results in phosphatidylinositol-3 kinase (PI3K) activation.[49]
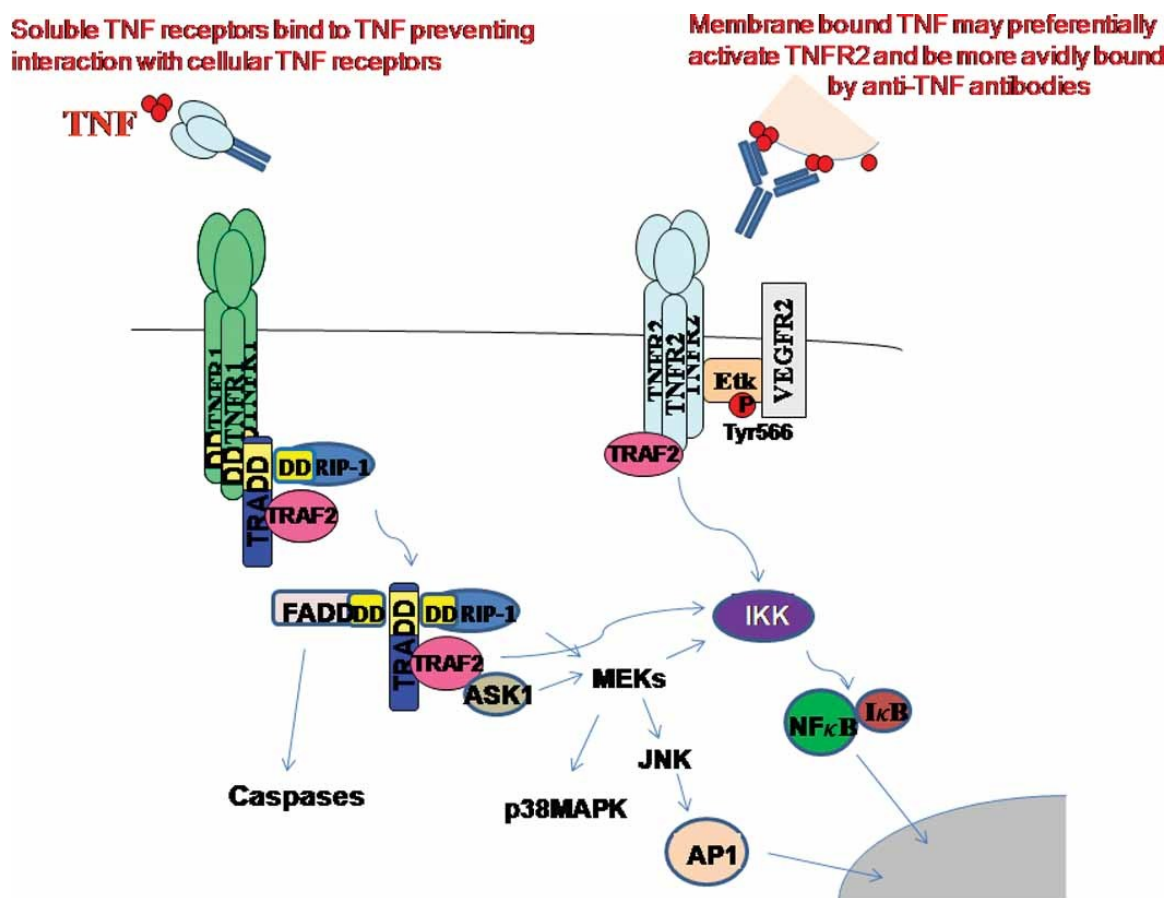
Figure 3 - TNF signaling pathways. Used with permission from [49]

Many inflammatory effects of TNF result from its effects on the vascular endothelium and endothelial leukocyte interactions. Following stimulation by TNF, endothelial cells present adhesion molecules for leukocytes such as E-selectin, intercellular adhesion molecule 1 (ICAM1) and vascular adhesion molecule 1 (VCAM1).[49] TNF furthermore induces the secretion of chemokines such as monocyte chemoattractant protein-1 (MCP1),[51] interleukin-8 (IL8),[52] and C-X-C motif chemokine 10 (CXCL10).[49, 53, 54] The adhesion molecules and chemokines facilitate the recruitment of different leukocyte populations. TNF furthermore induces cyclooxygenase-2 (COX2; synonym Cyclooxygenase-2, Prostaglandin G/H synthase 2, PTGS2), leading to increased levels of prostaglandin I2 (PGI2; synonym prostacyclin). This

causes vasodilation and thereby an increased local blood flow, leading to the inflammation features redness (rubor) and heat (calor). TNF causes swelling (tumor) through enhancing vascular permeability, leading to an increased trans-endothelial passage of fluids and macromolecules from blood vessels into the tissue. TNF is crucial for the normal immune response to bacteria, viruses and parasites, but an excessive production can be harmful.[49]

Naïve CD4+ T cells in secondary lymphoid organs (Peyer's patches and mesenteric lymph nodes) differentiate to regulatory T cells (e.g. T_{reg}) or helper T cells (e.g. T_H1, T_H2, T_H17), depending on the surrounding cytokine milieu and the antigen-presenting cells (see Figure 4). The activated CD4+ cells then migrate to the lamina propria, where they execute effector functions.[19] A transformation imbalance between T_H17 and T_{reg} cells may play an important role in IBD. T_{reg} cells effectively repair damaged tissue, but in the presence of interleukin-6 and/or interleukin-23 they transform into T_H17 cells. A transformation of T_H17 to T_{reg} cells was not reported. The colon mucosa of UC and CD patients shows increased T_H17 cells[55] and higher levels of the T_H17 cytokine interleukin-17, but decreased T_{reg} cells.[56]

T_H17 cells are crucial mediators in both UC and CD.[57] The interleukin-23–T_H17 pathway is central to the function of T_H17 cells and contributes to the sustained inflammation. Activated antigen-presenting cells (primarily dendritic cells, monocytes and macrophages) secrete interleukin-23.[19, 58] Interleukin-23 is as a member of the interleukin-12-type cytokine family and is a heterodimer of interleukin-12 subunit beta (IL12B, which it has in common with interleukin-12) and interleukin-23 subunit alpha (IL23A). Interleukin-23 signals to T_H17 cells by triggering the dimerization of interleukin-12 receptor subunit beta-1 (IL12RB1) and interleukin-23 receptor (IL23R), which together form the interleukin-23 receptor complex.[59] The interleukin-23 receptor complex signals through the JAK-STAT pathway.[19, 59] Most Signal Transducer and Activator of Transcription (STAT)-activating cytokine receptors do not have intrinsic tyrosine kinase activity. They instead recruit proteins from the Janus kinase (JAK) family that provide them with tyrosine kinase activity. In mammalian cells the JAK family contains the four proteins JAK1, JAK2, JAK3, and Tyrosine kinase 2 (TYK2).[60] The interleukin-23 receptor complex recruits JAK2 (Interleukin-23 receptor) and TYK2 (Interleukin-12 receptor subunit beta-1), which enables it to phosphorylate predominantly STAT3 and to a lesser degree STAT1, STAT4, and STAT5.[59] Phosphorylated STATs form homo- and heterodimers, which are rapidly transported to the nucleus. There they bind to target promoters and increase the transcription from those promoters.[60] The transcriptional activation results in increased proliferation and/or survival of T_H17 cells[61] and mediates intestinal inflammation.[19] T_H17 cells attract neutrophils and induce the release of anti-microbial peptides from epithelial cells. Among the T_H17 cytokines are interleukin-17A, interleukin-17F, interleukin-22, and TNF-α.[26] In addition to its functions related to T_H17 cells, interleukin-23 also restrains the activity of regulatory T cells and reduces their immunosuppressive pathways.[62]
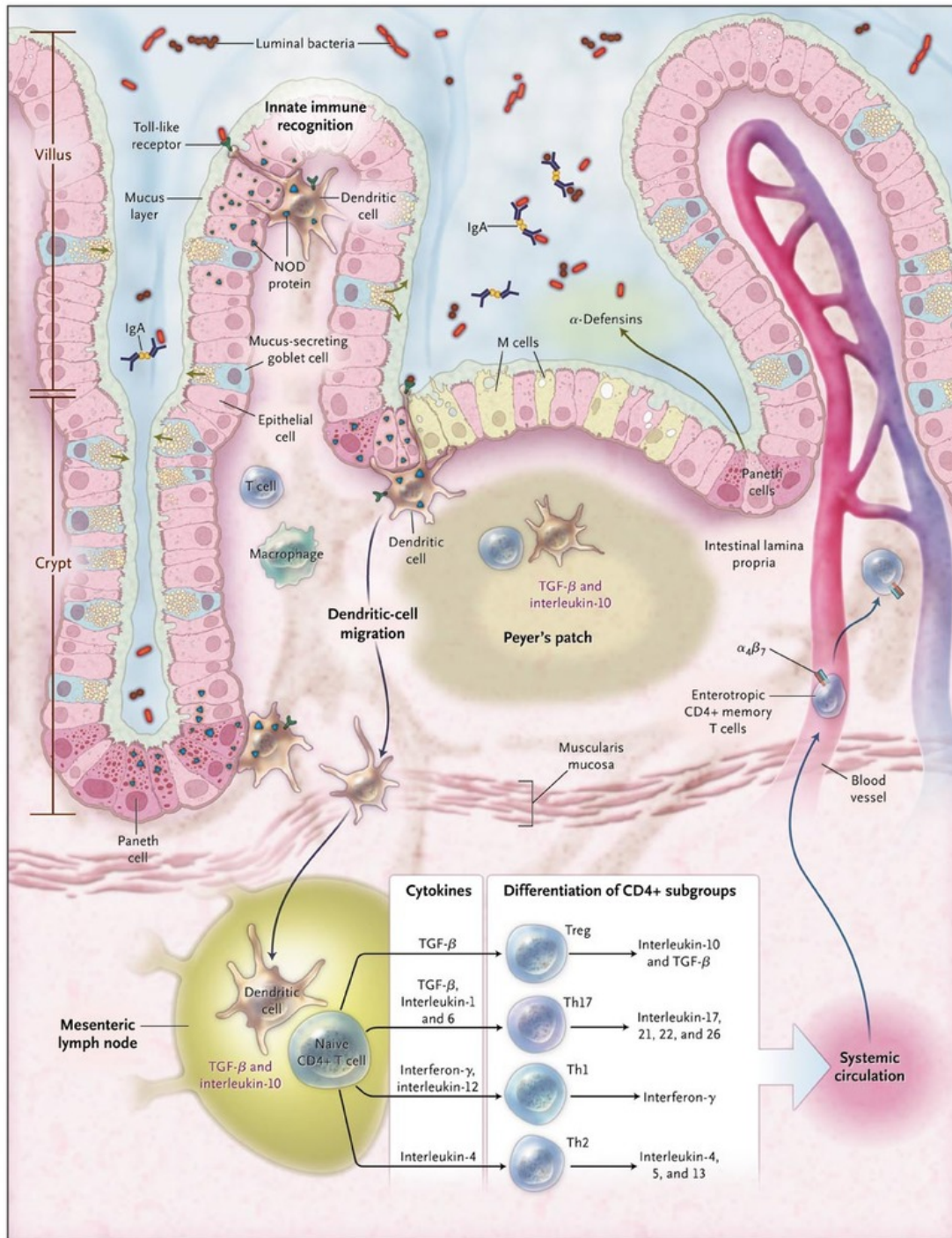
Figure 4 - The intestinal immune system and the role of T cell differentiation. Used with permission from [19]

Besides T_H2 and T_H17 cells, Natural Killer T cells that secrete interleukin-13 and T_H9 cells that secrete interleukin-9 contribute to the UC pathophysiology.[11, 63, 64] The contribution of the various cell types to the UC pathophysiology is subject to ongoing debate and research.

Nitric oxide (NO) is a diffusible messenger for cell-cell communication throughout the body. In the gastrointestinal mucosa, NO regulates perfusion, microvascular and epithelial

permeability, and immune processes. It has been associated with the initiation and maintenance of inflammation in IBD. Nitric oxide synthases (NOSs) produce NO from arginine. The constitutively expressed NOSs endothelial NOS and neuronal NOS produce NO in nanomolar amounts, while the inducible NOS (NOS2) produces micromolar NO quantities. Specific cytokines, microbes, and bacterial compounds induce iNOS.[65] The microbes that induce NOS2 include the enteroinvasive bacteria *Escherichia coli*, *Salmonella* and *Shigella*[65, 66], and NO has a direct antimicrobial effect.[67] This suggest a role of NO in the antimicrobial defense. However, sustained NOS2 overexpression and the resulting NO increase may be detrimental.[65] Increased levels of NOS2 and NO have been detected in UC tissue.[68-72]

Histamine is another small molecule messenger that is produced from an amino acid[73] and is a pro-inflammatory mediator.[74] Histidine decarboxylase produces histamine through α-decarboxylation from L-histidine.[73] In addition to its role in inflammation, histamine is involved in gastric acid secretion, bone loss, sleep regulation, food intake control, and schizophrenia.[73, 75-78] The known histamine receptors H_1, H_2, H_3 and H_4 are G protein-coupled receptors.[73] H_1, H_2 and H_3 can be found in the gut.[74] Histamine N-methyltransferase (HNMT) and diamine oxidase (amiloride binding protein 1, ABP1, AOC1) degrade histamine. A polymorphism in ABP1 is associated with UC severity[79] and histamine content is increased in UC colon mucosa.[80, 81] These findings indicate a role of histamine in the UC pathophysiology. Gut mast cells are probably a major source of the histamine in UC colon mucosa.[79, 81, 82]

The most prominent feature of UC is the intestinal inflammation, but the pathophysiology comprises a wide range of further aspects. Peroxisome proliferator-activated receptor gamma (PPAR-γ) is reduced in colonic epithelial cells of UC patients.[11, 83] PPAR-γ is a member of the nuclear receptor superfamily and a ligand dependent transcription factor. Its ligands include hormones, vitamins, endogenous metabolites, dietary compounds and synthetic drugs.[84] Four isoforms of PPAR-γ exist. PPAR-γ1 is expressed in ubiquitously, PPAR-γ2 is found mainly in adipose tissue, PPAR-γ3 is expressed in macrophages, large intestine, and white adipose tissue, and PPAR-γ4 can be found in endothelial cells.[84] PPAR-γ controls the expression of very many genes and is a key transcriptional regulator of lipid and glucose metabolism.[84-86] PPAR-γ agonists furthermore reduce the activation of macrophages[87] and decrease the cytokine secretion from myeloid cells[88] in vitro, and they attenuate colitis in mice in vivo.[89, 90] In macrophages PPAR-γ activation reduces the expression of inducible nitric oxide synthase (NOS2), matrix metalloproteinase-9 (MMP9, synonym Gelatinase B), and scavenger receptor class A (SR-A, synonyms CD204, Macrophage scavenger receptor types I and II, MSR1), partly by antagonizing the transcriptional induction by NF-κB, AP-1 and STAT1.[87] A mechanism for this was shown in Caco-2 cells. In these cells, the PPAR-γ agonist 15d-PGJ₂ inhibits the immune-response induced degradation of NF-κB inhibitor alpha (IκBα). This results in an inhibition of the promotor activity of nuclear factor-κB (NF-κB), because IκBα inhibits the nuclear translocation and subsequent DNA-binding of NF-κB.[90] PPAR-γ is of

interest to dietary research, because of findings that the anti-inflammatory properties of PPAR-γ ligands in food may be beneficial to human health.[84]

An energy deficiency of colonocytes seems to occur in colonocytes in UC. Short chain fatty acid (SCFA), especially butyrate[91], are the major energy source for colonocytes. Bacteria generate SCFA by fermentation from dietary fibers.[92] The oxidation of butyrate to carbon dioxide and ketone bodies is decreased in UC. The energy generation in UC is shifted to increased glucose and glutamine oxidation.[91] Monocarboxylate transporter 1 (SLC16A1, MCT1) is a transporter of butyrate and its expression is decreased in HT-29 upon treatment with the pro-inflammatory cytokines TNF interferon-γ. This indicates that the impairment is rather a result than the initial cause of the inflammation. In addition to its nutritional role, butyrate promotes the formation of Treg cells, which suppress inflammatory processes.[92] Its metabolite beta-hydroxybutyrate, a ketone body, furthermore acts as an endogenous histone deacetylase inhibitor. The resulting increased gene transcription of FOXO3A, encoding for Forkhead box protein O3, and MT2, encoding for metallothionein 2, increases the protection from oxidative stress.[93]

The involvement of metallothioneins (MTs) in IBD is not limited to metallothionein 2 and has been suspected for several members of the MT superfamily. Even though MTs are rapidly upregulated in inflammation and are hence classified as acute phase proteins, most studies report a decreased abundance of MTs in UC compared to healthy controls.[94] Human metallothioneins are heavy metal binding proteins that have a low molecular weight, are cysteine-rich, and a majority is clustered on chromosome 16. They bind heavy metals of the International Union of Pure and Applied Chemistry (IUPAC) groups of chemical elements 11, e.g. Cd^{2+}, Zn^{2+}, Hg^{2+}, and 12, e.g. Cu^+, Ag^+ and Au^+. MTs regulate the heavy metal metabolism and protect against toxic heavy metal ions.[95] Zinc ions have various roles in inflammation, including anti-inflammatory properties. MTs, together with other compounds like vitamin E, ascorbate and superoxide dismutase (SOD1), also protect against reactive oxygen species (ROS). Neutrophils and macrophages produce ROS as a defense mechanism to kill bacteria and parasites, but ROS are also cytotoxic to host cell. ROS cause oxidative DNA damage, oxidative protein damage and lipid peroxidation, and they can destabilize tight junctions, thereby increasing permeability.[94]

Further contributing factors in UC colon mucosa include, but are not limited to:

- collagen degradation by matrix metalloproteinases (MMPs)[96]
- decreased carbonic anhydrase abundances[97]
- epithelial ER stress and the related unfolded protein response (UPR)[98]
- mitochondrial dysfunction[99]
- decrease of xenobiotic metabolism[100]
- decrease of bile acid transporter expression[101]

- increased noradrenaline content[102]
- changed expression of steroid metabolism enzymes[103, 104]

1.1.4 Symptoms and clinical presentation of Ulcerative Colitis

Blood in the stool is the hallmark symptom of UC and is reported by more than 90% of the patients.[5, 105] A symptom of extensive active UC is chronic diarrhea with rectal bleeding or at least visible blood. UC symptoms further comprise crampy abdominal pain, rectal urgency[105], incontinence, increased frequency of bowel movements, mucus discharge and nocturnal defecations.[6, 11] Irritation from diarrhea can cause skin tags and anal fissures in UC.[11] Systemic symptoms of UC include weight loss, fever, tachycardia, nausea, vomiting[105], anemia and fatigue.

Extraintestinal manifestations (EIMs) occur in approximately one third of the UC patients and can precede an IBD diagnosis.[11, 106] Anatomical structures affected by EIMs of UC include skin, joints, eyes, mouth, liver and lung.[11] The most common EIMs in UC are peripheral arthritis (21%, joints), aphthous stomatitis (4%, mouth), uveitis (4%, eye), and primary sclerosing cholangitis (4%, liver).[107]

The clinical course in UC comprises phases of exacerbation and remission, which occur either spontaneous or as a response to treatment or intercurrent illnesses.[5] Up to 15% of the patients present initially with severe illness.[105]

UC is classified based on the disease extend (see Figure 5).[108, 109] The clinical presentation tends to vary depending on this. Urgency and tenesmus (a sensation of incomplete evacuation) seem to predominate in patients with proctitis, whereas bloody diarrhea and abdominal pain are probably more prominent in pancolitis. Paradoxical constipation occurs in up to 10% of the patients with proctitis or left-sided colitis.[11]
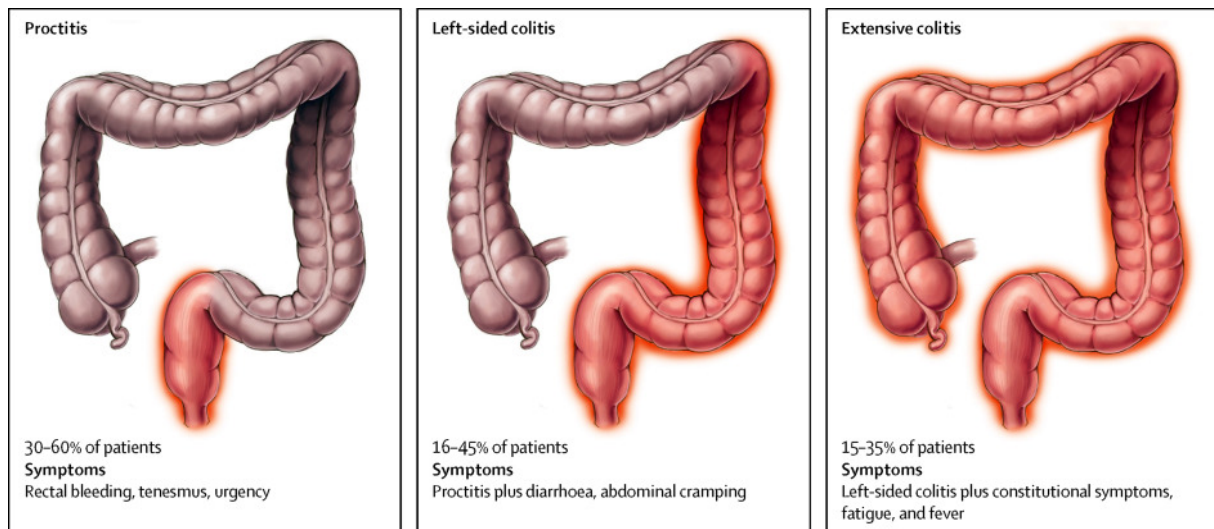
Figure 5 - Disease extend of UC according to Montreal Classification. Illustration by Jill K Gregory, used with permission from ©Mount Sinai Health System

Further classification is based on the disease severity. The various clinical disease activity indices usually return the classifications *mild*, *moderate*, and *severe*. Examples are the Mayo Score[110], Lichtiger Score[111], Simple Clinical Colitis Activity Index[112], Ulcerative Colitis Endoscopic Index of Severity (UCEIS)[11, 113], and Ulcerative Colitis Disease Activity Index (UCDAI).[114] These scores assess the frequency and severity of symptoms (including EIMs) and endoscopic findings.[11] The Geboes Index[115], Nancy index[116], Robarts Histopathology index (RHI)[117], and Modified Riley Score (MRS)[118] assess histological features.

Clostridium difficile infections among UC patients are a growing problem, because they can precipitate flares and are associated with increased mortality, hospitalization rates and resource utilization.[11, 119, 120]

1.1.5 Treatment

The treatment aim in UC is to achieve remission. Remission can be defined as the resolution of clinical symptoms and as the endoscopic mucosa healing, which do not necessarily correlate well with each other.[11, 121-123] The resolution of clinical symptoms is important for the patients' quality of life. However, endoscopic mucosa healing must be assessed independently of this, because mucosal healing is associated with improved long-term clinical remission, corticosteroid-free clinical remission and decreased risk of colectomy.[124] The treatment options comprise topical and systemic administration of drugs as well as surgery. Optimized treatment approach algorithms in the form of flow charts help in choosing an appropriate treatment (see Figures 6 and 7).
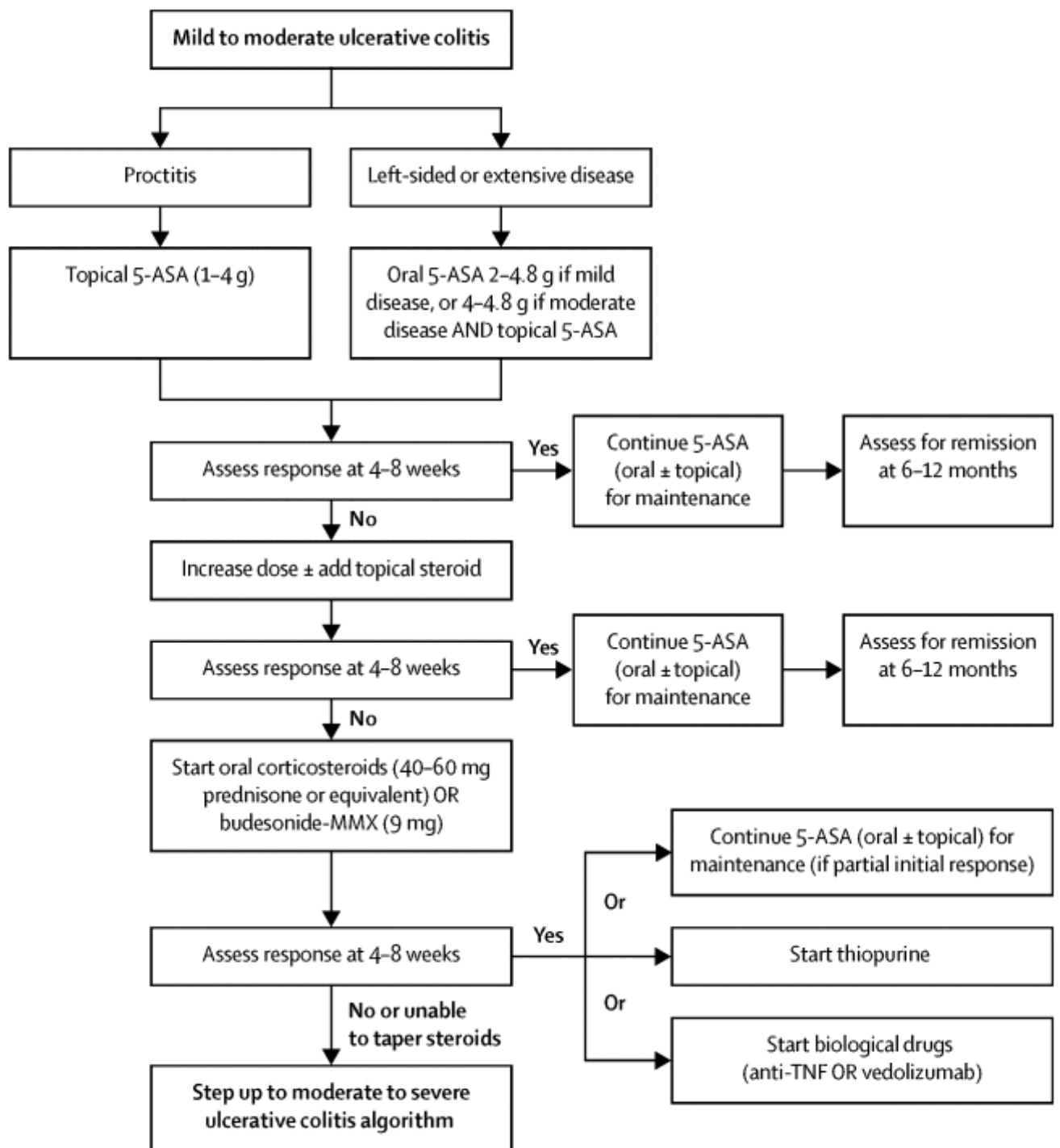
Figure 6 - Exemplary treatment approach algorithm for mild to moderate UC. Used with permission from [11]
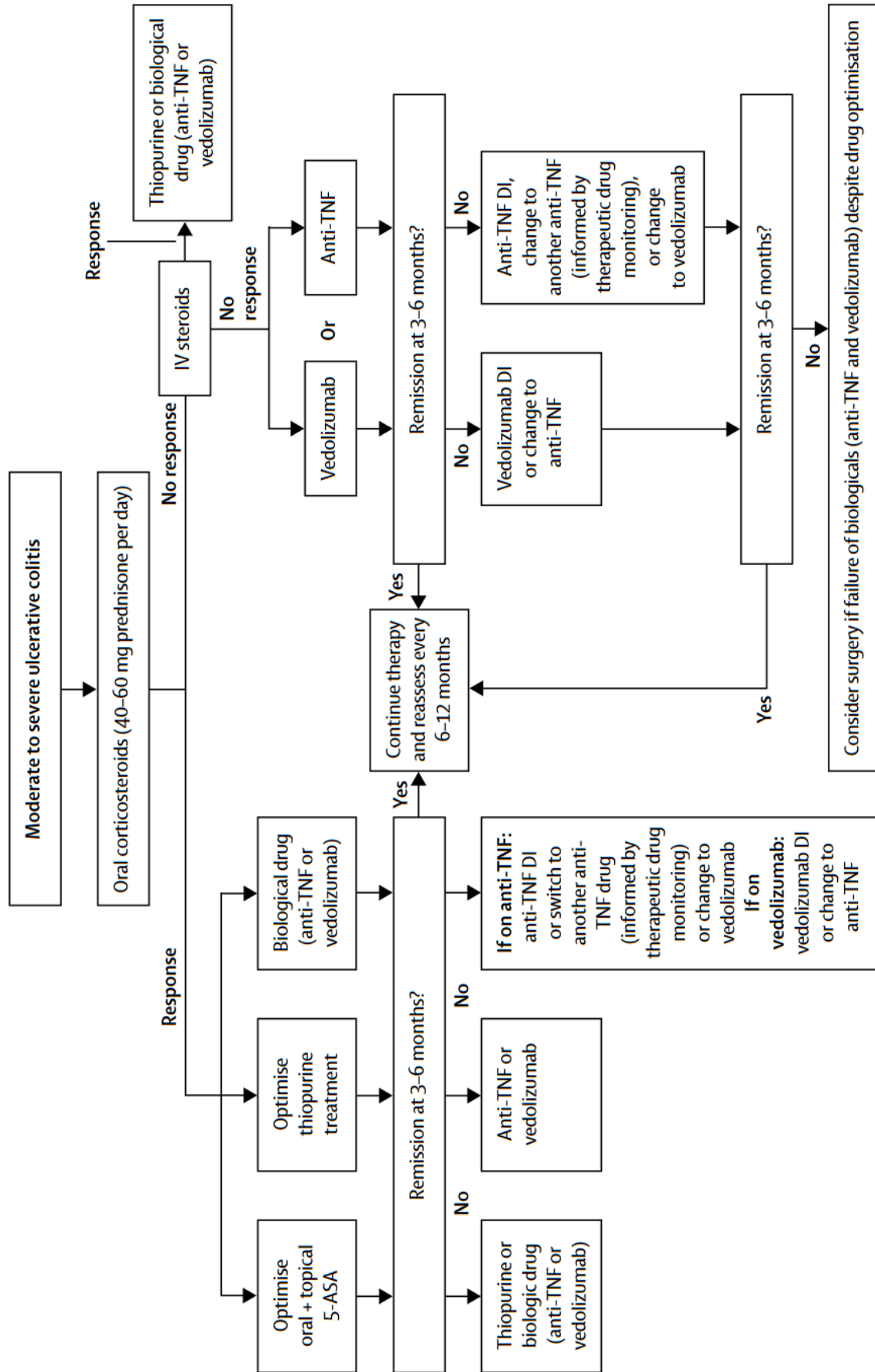
Figure 7 - Exemplary treatment approach algorithm for moderate to severe UC. Used with permission from [11]

21

5-aminosalicylic acid

5-aminosalicylic acid (5-ASA, synonyms Mesalazine, Mesalamine) drugs are the first-line treatment of mild and moderate ulcerative colitis.[11] The mechanism of action of 5-ASA is poorly understood. It is believed that it decreases the production of pro-inflammatory prostaglandins and leukotrienes from the cyclooxygenase and lipoxygenase pathways.[125-127] 5-ASA probably exerts further effects over PPAR-γ, as it was shown that 5-ASA increases PPAR-γ expression, promotes its translocation from the cytoplasm to the nucleus, and changes its conformation, which increases the binding of coactivators and the activation of a peroxisome-proliferator response element-driven gene.[128] 5-ASA may further have antioxidant properties and act as a radical scavenger.[129] 5-ASA efficacy is not dependent on systemic absorption and redistribution to the colon mucosa, but rather on the effects resulting from topical concentrations. However, after oral ingestion 5-ASA is rapidly absorbed by the small intestine; hence, approaches were developed to enhance topical drug concentrations in the colon.[130] Common approaches to achieve high 5-ASA concentrations in the colon are tablet and microgranule formulations that release the 5-ASA at the desired sites (often by coating), prodrugs that contain an azo bond and release 5-ASA upon cleavage by bacterial azoreductases in the colon, and the use of suppositories or enemas. Enemas may be gels, liquids, or aerosols.[131]

Glucocorticoids

Glucocorticoids (GCs) can be given to patients with mild to moderate disease who do not respond or do not reach remission under 5-ASA treatment.[11] Intravenous GCs are the mainstay of conventional therapy of severe ulcerative colitis of any extend.[132, 133]

Oral GCs are readily absorbed and transported with the blood, where they are mainly bound to corticosteroid binding globulin (CBG, SERPINA6) and to a lesser extend to albumin (ALB). GCs diffuse passively through cell membranes and bind to the intracellular glucocorticoid receptor (GR). GR is, like the previously discussed PPAR-γ, a nuclear receptor. Accordingly, it primarily functions as transcription factor.[134] These genomic effects are the classical actions of glucocorticoid signaling. GRs can homodimerize upon binding of a ligand and subsequently bind to glucocorticoid response elements of GC target genes. The bound GR modulates the transcription of the respective gene.[135] In addition to this, GR can change the gene transcription in several other ways.[135-138] One mechanism of significance for the use of GCs as anti-inflammatory agents is that GR interact ("tether") with other transcription factors, which are DNA-bound to their respective response elements. This can cause enhancement or repression of transcription. GR interacts in such way with NF-κB and AP-1 and decreases the expression of their target genes.[136, 138] This contributes to the anti-inflammatory effects of GCs.[135] Genomic effects from GR would be expected after a few hours, because the transcription and translation are time-consuming. However, evidence

suggests that GCs have effects that occur within minutes and non-genomic mechanisms are proposed for those. One of these mechanisms is an interaction of GR with signal transduction pathways.[135] Another hypothesis is that rapid GC effects result from physicochemical interactions with cell membranes.[139]

GCs affect all major systems of the body.[140] The diversity of effects reflects in diverse side effects, which are a concern in GC treatments with a duration of more than one week.[141] The side effects of prolonged GC treatment include weight gain, hyperglycemia, osteoporosis, hypertension, dyslipidemia, glaucoma, and psychiatric and cognitive disturbances.[142]

GCs can exert their effects either systemically or topically. The administration for topical effects can be either rectal, e.g. in the form of foams or enemas, or oral. Topical effects of oral GCs are achieved by using modified-release formulations of GCs with high first-pass liver metabolism, such as budesonide-multimatrix and prolonged release beclomethasone dipropionate. They are a first-line treatment of patients with mild to moderate UC who do not reach remission with 5-ASA.[11]

Thiopurines

The thiopurines azathioprine (AZA) and 6-mercaptopurine (MP) are used for remission maintenance in UC patients.[11]

The first step in the metabolism of AZA is its conversion to MP. However, both AZA and MP are prodrugs. Multiple enzymes are involved in the metabolism to their active metabolite 6-thioguanine nucleotide (6TGN).[143] 6TGN exerts its immunosuppressive effects mainly by inactivating Rac1, which is a small GTPase and an important intracellular mediator with impact on the fate of T cells. The inactivation of Rac1 by 6TGN induces T cell apoptosis and impairs the interaction of T cells with antigen-presenting cells. In addition to this mechanism, a small proportion of 6TGN is incorporated into the DNA instead of guanine. This activates the mismatch repair system and eventually leads to cell death.[144]

Biologics

The biologics applied in the UC treatment are antibodies targeting TNF and an integrin, respectively.[11, 145] They are effective at inducing and maintaining remission in moderate and severe UC.[11]

The TNF antibodies infliximab, adalimumab and golimumab are used in the treatment of UC.[11, 146] They bind to TNF and thereby block the interactions with its receptors. As described earlier, TNF is a major mediator of inflammation in UC, accordingly the inhibition of its effects result in decreased inflammation. However, the exact molecular mechanisms

involved in anti-TNF treatment are still a matter of debate, because these molecules have effects in addition to blocking TNF.[146] In addition to the soluble TNF acting as a ligand, its precursor form transmembrane TNF (mTNF) can act as a receptor that is activated upon binding of the TNF antibodies. This is believed to contribute further to the anti-inflammatory effects of TNF antibodies.[146, 147]

The antibody vedolizumab is directed against α4β7 integrin. α4β7 integrin is a glycoprotein that B and T cells present on their cell surface.[148] It interacts with mucosal addressin-cell adhesion molecule 1 (MAdCAM-1), which is preferentially expressed in the intestinal tract and associated lymphoid tissues, but is not detected in the majority of other tissues, including those with mucosal surfaces.[149] The interaction between α4β7 integrin and MAdCAM-1 facilitates the lymphocyte recruitment to intestinal tissue.[148, 149] The binding of vedolizumab to α4β7 integrin prevents this.[148]

The development of biologics improved the health outcomes in IBD drastically, and accordingly their use has been increasing. For instance, one US study showed that the proportion of UC outpatients using biologics increased from 5.1% in 2007 to 16.2% in 2015. However, with an increased use not only in UC (for instance the proportion of CD patients using biologics increased in the same study from 21.8% to 43.8%), the expensive biologics have a considerable impact on the health care costs of IBD.[150]

Surgery

Surgery is indicated in patients with refractory UC, perforation, uncontrolled bleeding, steroid dependence, colorectal carcinoma, or dysplastic lesions that are not amenable to endoscopic removal.[11, 151] It can also be necessary when the disease impairs the patient's quality of life significantly or when the patient is not compliant.[151] Despite the progresses in medical treatment, surgery is needed in 15%-35% of the UC patients.[151-153] The gold standard and most commonly performed surgery is restorative proctocolectomy with ileal pouch-anal anastomosis (RP-IPAA), which comprises a total resection of the colon and upper rectum, a construction of a pouch from the end of the small intestine, and the attachment of that pouch to the anus.[11, 151, 154] The most common postoperative issue after RP-IPAA is pouchitis. Pouchitis is an inflammatory condition of the ileal pouch, of which at least one episode occurs in up to 46% of the RP-IPAA patients.[11]

1.1.6 Clinical outcome

The clinical outcomes in IBD are not clearly defined and there is little agreement on the endpoints.[155] One of the well-known studies on outcomes in UC is the IBSEN study. The IBSEN study applied four predefined curves to describe the disease progression (Figure 8).[156] 59% of the 420 patients in the study who did not undergo surgery experienced a decline of disease severity (Figure 8, curve 1). 31% of these patients experienced chronic relapsing

symptoms (Figure 8, curve 4) and 9% chronic continuous symptoms (Figure 8, curve 3). The disease severity increased for only 1% of the patients (Figure 8, curve 2).
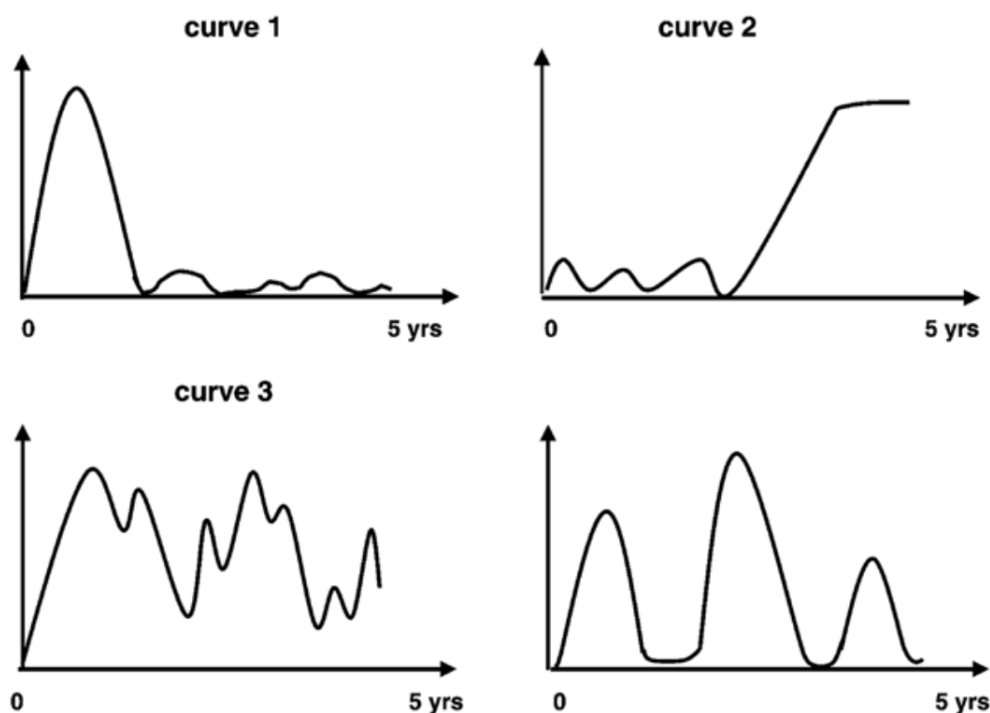
Figure 8 - Predefined curves describing the disease progression in the IBSEN study. Used with permission from [156]

The UC treatment according to the medical step-up approach uses 5-ASA and local GCs in mild disease manifestations and additional oral GC, immunomodulator and biological therapy in moderate to severe disease.[157] Conversely, the top-down approach uses immunomodulators and anti-TNF agents early in the disease. While no studies are published on the top-down therapy in UC, it can induce long term clinical remission of CD.[158]

Good biomarkers for predicting the disease progression at UC onset would support clinicians in tailoring individualized therapies. For instance, the top-down treatment approach may be most beneficial in patients with a predicted severe disease course.[159] An optimized, personalized treatment starting at disease onset may change the course of the disease.[160] Accordingly, the demand is high for biomarkers that can predict the disease progression in IBD. Some progress has been made in this area. Biomarkers have been suggested for predicting outcomes in CD.[159] It was furthermore shown that normalized mucosal TNF mRNA can predict an increased duration of remission in both UC and CD.[161, 162] However, there still is a lack of good biomarkers for predicting clinical outcomes in UC at disease onset.

1.2 Proteomics

Proteomics is the scientific discipline that investigates the proteome. The proteome can be defined as the entire protein content of a cell, tissue or organism.

1.2.1 Value of proteomics

Proteins are the major functional units of every cell. Hence, knowledge about the abundance and the state of proteins are crucial for our understanding of physiological and pathophysiological processes. However, the determination of protein abundances is challenging.

The measurement of messenger RNA (mRNA) abundances in great depth is comparably easy, because mRNA can be amplified. mRNA is the template for protein translation. Hence, mRNA abundances are routinely measured in transcriptomics as surrogates for protein abundances. The use of this approach however is limited, because the abundances of mRNA and protein correlate poorly.[163-165] This poor correlation results from varying rates of the translation from a given amount of mRNA and of the degradation of proteins, see Figure 9.[165]
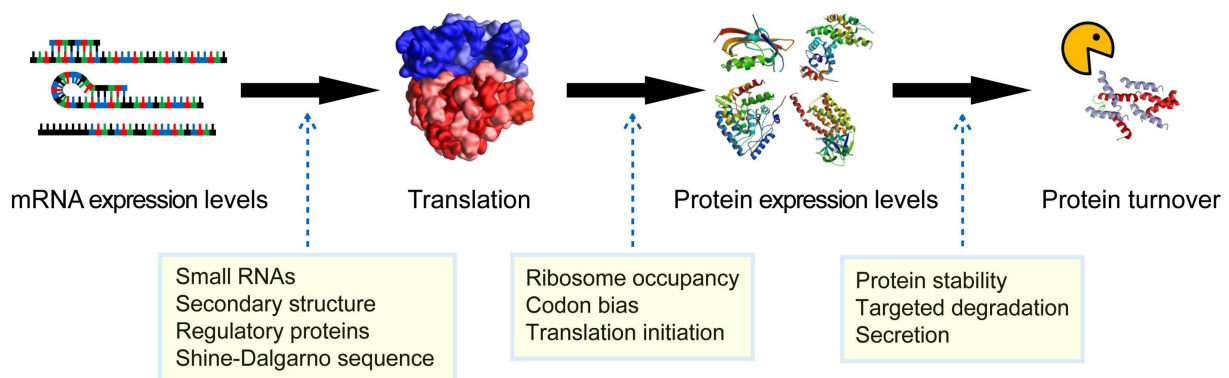
Figure 9 - The translation and degradation of proteins, and factors that contribute to a poor correlation between mRNA and protein abundances. Used and modified with permission from [165]

The translational efficiency is defined as the number of completed protein translations per mRNA molecule and time. Translational efficiencies differ between mRNA molecules, which directly influences the mRNA-protein correlation.[165] One factor that influences the translation rate are the properties of the mRNA itself. The secondary mRNA structure is variable in eukaryotes and influences translation.[166, 167] The mRNA codon sequence in dependence of transfer RNA (tRNA) levels can further influence the transcription rate.[168] The distribution of the mRNA within the cell, e.g. localization in the nucleus, changes its access to ribosomes for translation.[165] Small RNA (sRNA) can induce target mRNA cleavage and destabilization without cleavage, inhibit protein translation, and induce histone modifications that silence further transcription.[169, 170] While mRNA degradation or reduced transcription do not cause poor correlation between mRNA and protein levels because the mRNA levels decrease as well, inhibition of translation can contribute to decreased correlation. The

translation is facilitated by ribosomes, which are built from ribosomal RNA (rRNA), and modulated by further proteins. Most such effects result from regulation at the stages of initiation.[171-173] Regulation of elongation and termination contribute as well.[174, 175] A study in yeast found a hundred-fold range of translational efficiency in the same sample between different genes, and changes in translational efficiency upon starvation were found for roughly one-third of the genes.[176]

The individual rates of protein degradation are the most important post-translational influence on the mRNA-protein correlation. The half-life of a protein depends on factors including its intrinsic stability, posttranslational modifications (PTMs, e.g. ubiquitination), its last N-terminal amino acid (N-end rule), and protein localization.[165, 177] Protein secretion further decreases the correlation, because secreted protein that leave the sample tissue escape the quantification, while proteins translated and secreted elsewhere can enter the sample tissue.[165]

Beside the advantage of measuring the abundance of proteins instead of a surrogate, proteomics can identify and quantify PTMs. PTMs are changes of covalent bonds in proteins after translation. They are not encoded in the mRNA and for this reason not accessible to transcriptomics. Most PTMs are introduced enzymatically by the respective cell or organism. The term PTM summarizes diverse chemical reactions. These include the covalent attachment of various inorganic and organic groups/molecules (including other proteins and lipids), formation of disulfide bridges between cysteines, other chemical modifications on amino acids (e.g. deamidation of glutamine/asparagine, hydroxylation of proline, isomerization), and proteolytic cleavage at peptide bonds. Protein phosphorylytion can arguably be considered one of the most important regulatory mechanisms. PTMs in proteins can influence their activity state, localization, and stability, can modify the interaction with other proteins, and make major structural contributions.[178]

1.2.2 History of proteomics

The first studies that can be categorized as proteome investigation were published in 1975.[179-182] These beginnings of proteomics were dominated by two-dimensional gel electrophoresis (2-DE) of proteins. These techniques separated and visualized proteins, but the proteins could not be identified due to the lack of sensitive sequencing techniques, which were necessary because of the limited loading capacity in gel electrophoresis. The first major method for protein sequencing was the Edman degradation[183], but the identification of the electroblotted proteins became only possible in the 1980s with the emergence of microsequencing procedures.[179, 184]

The proteomics field has since experienced major developments. Sequencing has essentially been replaced by mass spectrometry (MS) approaches. Most commonly, bottom-up approaches

are applied. In bottom-up approaches, the proteins are enzymatically cleaved into peptides, which are subsequently identified by high-resolution tandem-MS.

Electrophoresis of the proteins is today only an optional step to reduce sample complexity and commonly omitted. Gel electrophoresis is no longer mandatory, because the peptides resulting from the protein digestion are usually separated by low-pH reversed-phase liquid chromatography (LC) before they are subjected to mass spectrometry analysis. Furthermore, when an additional reduction of the sample complexity is desired, other fractionation approaches such as high pH reversed-phase, strong cation exchange (SCX) or hydrophilic interaction chromatography (HILIC) are often preferred over gel electrophoresis.

1.2.3 Basic sample preparation workflow in bottom-up proteomics

The bottom-up proteome analysis of a given sample consists of three major steps: sample preparation, liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, and computational analyses, see Figure 10.
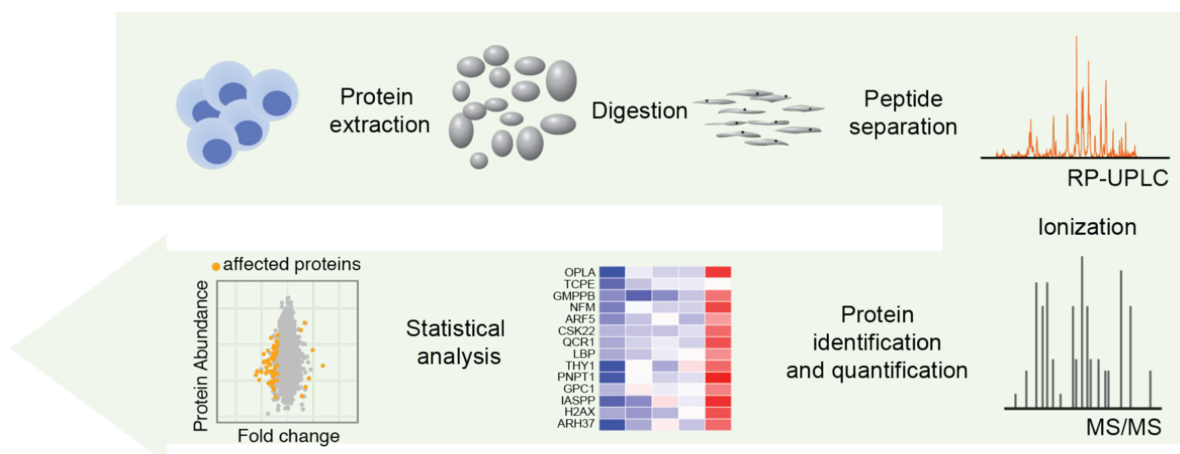
Figure 10 - Typical workflow for quantitative bottom-up proteomics. Used with permission from Joanna Kirkpatrick, Core Facility Proteomics, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI)

The purpose of the sample preparation in a bottom-up proteomics study is to convert the protein-containing sample of interest into a peptide solution for LC-MS/MS analysis. The first step of the sample preparation facilitates the lysis and homogenization of the cell or tissue sample as well as the solubilization and denaturation of the proteins. This can be accomplished with reagents and mechanical force. Among the reagents for this purpose are detergents and urea (see below). The mechanical disruption methods include grinding in liquid nitrogen[185], sonication, cell scraping, bead mills, gentle rocking, and French pressing.[186] The macroscopic appearance and properties of the sample help in selecting an appropriate mechanical disruption method.

Disulfide bonds occur between protein cysteines and can contribute to the tertiary structure of proteins. These bonds add additional complexity, the interpretation of the resulting mass spectrometry data is challenging, and the disulfide bonds can rearrange (so-called reshuffling).[187] The resulting data from linked peptides is furthermore not compatible with the standard computational approaches for bottom-up proteomics. Adjusted protocols can tackle these challenges and extensive research efforts are directed at disulfide proteomes.[187-189] However, disulfide bonds are undesirable in general bottom-up proteomics. Hence, reducing agents like dithiothreitol (DTT) are added to reduce disulfide bonds to thiols, which are then alkylated with a reagent like iodoacetamide (IAA). This step serves to break the disulfide bonds and to prevent their reformation

However, IAA can alkylate peptide N-termini and side-chain nitrogens in addition to the more reactive cysteine thiols, if given enough incubation time. The alkylation of these amino groups with IAA is termed overalkylation.[190] Overalkylation increases the sample complexity and peptides that are modified in such way would not be identified in computational analysis, unless specified as possible modification. Overalkylation furthermore introduces a possible bias with regard to peptide quantification. Quenching of excess IAA with additional DTT after alkylation is a suitable measure to avoid overalkylation.[191]

The proteins are then enzymatically digested into peptides with enzymes, which cut at specific cleavage sites. Trypsin is the most commonly applied enzyme and cleaves the amino acid chain after the basic amino acids lysine and arginine. Lys-C cuts proteins after lysine as well, but tolerates higher urea concentrations. Hence, it can be applied for protein predigestion under harsher solubilization conditions, without introducing additional cleavage sites compared to tryptic digestion alone. See below for more detail.

An optional step is the isobaric labelling of the resulting peptides. Isobaric labelling is a chemical peptide modification that attaches chemical groups of identical structure and total masses, but with different distributions of heavy and light isotopes. Each sample for one LC-MS/MS run is labelled with a different mass tag and the samples are then mixed. The identical chemical properties and total masses of these mass tags make them indistinguishable in both the liquid chromatography and the mass spectrometer prior to fragmentation. The fragmentation in the collision cell however releases reporter ions of distinct masses. The relative intensities of the respective reporter ions reflect the proportion of peptide molecules originating from each sample. This approach can reduce the possible bias resulting from changes in the LC-MS/MS system between runs.

After the digestion, and if applicable isobaric labelling, substances that would interfere with the LC-MS/MS analysis must be removed from the peptide solution. These substances include detergents and salts. The detergent removal methods vary depending on their chemical properties, and some need to be performed before the digestion (see below). Solid-phase

extraction methods are most commonly applied to remove hydrophilic contaminants and salts that are not compatible with mass spectrometry.

In solid-phase extractions, the peptide solution flows over a solid stationary resin that retains the peptides. Salts are not retained and are washed away with the sample solvent and washing solutions. Elution buffer release the peptides from the resin. The desalting can either be a separate step or be included in a sample fractionation workflow. Pipette tips that contain a small bed of C18 sorbent are commonly applied when no fractionation is desired.

Before the sample clean up with C18 pipette tips, the sample solution must be acidified with an ion-pairing agent like trifluoroacetic acid (TFA) or heptafluorobutyric acid (HFBA). The peptides get protonated under acidic conditions and carry positive charges. The conjugated bases of the ion pairing agents (i.e. trifluoroacetate and heptafluorobutyrate ions, respectively) are negatively charged. Ion pairing of peptides with TFA and HFBA, respectively, removes the overall positive charge of the peptides; hence, it increases the peptide hydrophobicity and thereby their affinity to the C18 sorbent. The C18 sorbent must be conditioned before use with an acetonitrile-water solution to improve peptide binding. Subsequent flushing with TFA (or HFBA) solution removes residual organic solvent. The peptide solution is then applied to the sorbent by repeated aspiration. During that, the peptides bind to the sorbent. The bound peptides are then repeatedly washed with TFA (or HFBA) solution and eventually eluted. The organic solvent in the elution buffer breaks the interactions between the peptides and the sorbent and thereby releases the peptides. Evaporation facilitates the removal of the organic solvent in the peptide solution. Reconstitution of the peptides in MS compatible acid solution and dilution to the desired peptide concentration give the peptide solution for LC-MS/MS analysis.

1.2.4 Detergents for protein denaturation and solubilization

Detergents are applied frequently in proteomics experiments because they denaturize proteins and increase their solubilization in lysis buffers. These properties improve the enzymatic digestion of sample proteins. The presence of detergents is especially beneficial for the digestion of hydrophobic proteins, e.g. transmembrane proteins.[192] Detergents can decrease the activity of proteolytic enzymes, which limits the applicable concentrations.

Detergent molecules have both hydrophilic and lipophilic moieties. Hence, they can facilitate the interaction between lipophilic protein moieties and the water as solvent, which in turn increases the protein solubilization. Detergents can be further classified as anionic, cationic, zwitterionic and non-ionic according to their electrical charge. Anionic detergents carry negatively charged groups such as carboxylate (e.g. sodium deoxycholate, sodium laurate) or sulfate functions (e.g. sodium dodecyl sulfate). The positively charged groups in cationic detergents are most commonly quaternary ammonium groups (e.g. cetyltrimethylammonium bromide). Zwitterionic detergents carry equal numbers of negative and positive charges, resulting in a net zero charge. Non-ionic detergents contain no charged groups.

Detergents interfere with the liquid chromatography; hence, they need to be removed or degraded before injection into an LC-MS/MS system. Ionic detergents can have a fixed charge or their charge can be pH dependent. The removal of pH-dependent charges from detergents decreases their hydrophilicity and as a result their water solubility. Those detergents with pH-dependent charges that show sufficiently low water solubility after charge removal, precipitate (i.e. transition to solid state), and present with higher density than water can be effectively removed by pH manipulation and subsequent centrifugation. Under these conditions, the respective detergent forms a pellet and the peptide solution can be separated from the detergent by pipetting. This procedure allows the removal of detergents with pH dependent charges after digestion, which is advantageous especially for membrane proteins.[192] If compatibility of the detergent with tandem mass tag (TMT) labelling is given, this furthermore allows for detergent removal after the differently labelled peptides are mixed. Alterations of the sample composition due to the detergent removal step would then not change the detected isotope ratios, except for particular cases such as contaminations with other TMT labelled peptides.

Commonly applied methods for the removal of detergents with fixed charges need to be performed prior to protein digestion, e.g. protocols depending on filter-aided sample preparation (FASP)[193], protein precipitation[194], or polyacrylamide gel electrophoresis (PAGE). These methods have further disadvantages besides the loss of the potentially beneficial effects of detergents on the digestion. The additional, partially tedious, working steps prior to enzymatic digestion increase the risk of contaminations. Contaminations with proteins, e.g. keratin, from these steps undergo crucial parts of the sample preparation, i.e. reduction, alkylation, digestion, and labelling. Hence, these contaminants would appear as proteins of the respective samples. Furthermore, especially complex additional steps before mixing of the differently labelled peptides introduce the risk of producing artificial differences between samples. However, alternatives exist that allow the removal of detergents with fixed charges after digestion, e.g. Pierce Detergent Removal Spin Columns, which are packed with a proprietary resin.

Many detergents have been applied in proteomics sample preparation. Anionic detergents are the most common choice. Due to its role in two-dimensional PAGE, sodium dodecyl sulfate (SDS) has been an important detergent in proteomics from the early days. SDS has this role in two-dimensional PAGE because it binds to all proteins in the ratio 1.4 g SDS / 1 g protein under the conditions used in electrophoresis. This results in the loss of all protein specificity in two-dimensional PAGE, with the mobility becoming a measure of the molecular size alone.[195] An unbranched 12-C hydrocarbon tail constitutes the lipophilic part of SDS, while a sulfate group serves as the hydrophilic moiety. When in-solution digestion protocols gained increasing relevance, not the least due to progress in LC-MS/MS technology, the development of new detergent removal methods such as FASP[193] secured SDS a continuing significant role in proteomics sample preparations.

Besides detergents, urea has been commonly used for protein denaturation for more than a century.[196] Urea denatures proteins probably by direct interaction with the backbone and other hydrophobic protein moieties.[197, 198] An indirect effect from strengthening the water structure might contribute as well, probably to a minor degree.[198]

The detergent choice is an important factor in optimizing both the reproducibility and the identification numbers in proteomics experiments. Even though general trends regarding the preferability of detergents exist, the different compositions of biological samples warrant individual optimization efforts.

1.2.5 The endoproteinases trypsin and Lys-C in proteomics

Endoproteinases are proteins that cleave peptide bonds within other proteins. Endoproteinases can be categorized according to their catalytic residue. The catalytic residues often are often in the side-chains of amino acids (e.g. alcoholic function in serine and threonine, cysteine thiols etc.). Metalloproteases utilize metal ions, most commonly zinc ions, as their catalytic residue. The most commonly applied endoproteinases in proteomics are serine proteases.

Endoproteinases cleave peptide bonds usually either before or after one or several specific amino acids. Digest specificity, efficiency and reproducibility are prerequisites for reliable quantifications and identifications. The specificity is important in bottom-up proteomics because protein identifications in these experiments are based on the comparison of measured mass spectra with theoretical spectra that are derived from in-silico cleavages of known protein sequences. These in-silico cleavages are performed according to rules defined depending on the used endoproteinase. High digest efficiency decreases the sample complexity because it removes additional, often redundant, peptides that would result from missed cleavages. High digest efficiency can also contribute to high reproducibility. High digest reproducibility is important for quantification, because protein intensities in bottom-up experiments are derived from peptide intensities.

The serine protease trypsin is the most commonly applied endoproteinase in proteomics. By the end of 2014, 96% of the deposited data at Global Proteome Machine Database (GPM DB) were from tryptic digestions.[199] Trypsin is a protease of the digestive system that appeared early in evolution. Trypsin stems from a common ancestor of both prokaryotes and eukaryotes.[200] Species-specific adaptations of trypsins, for factors such as temperature and substrates, increase the fitness of the respective organisms. In animals, trypsin can be found in both vertebrates and invertebrates.[201] The trypsin cleavage pattern is commonly simplified to cutting after (i.e. on the carboxyl side of) lysine and arginine, but not before proline. However, violations of this "Keil rule" occur with commercially available trypsin.[202] The tryptic cleavage pattern yields peptides in a favorable mass range for sequencing. The basic properties of arginine and lysine furthermore add a positive charge to all peptide C-termini, except for the protein C-terminal peptides. Due to this additional charge, the C-terminal fragments become accessible to mass

spectrometry as y-ions. Together with the b-ions from the N-terminal peptide fragments, this gives information-rich, easily interpretable peptide fragment spectra.[203]

However, trypsin has some shortcomings. Tightly packed proteins can be resistant to tryptic cleavage, and the use of reagents to facilitate unfolding is limited because many such reagents inhibit trypsin activity. Furthermore, tryptic digestion is usually incomplete, with missed cleavages occurring especially at theoretical cleavage sites after lysine. The use of Lys-C in addition to trypsin can avoid these shortcomings.[204]

Lys-C originates from *Lysobacter enzymogenes*[205] and is, like trypsin, a serine endoproteinase. Lys-C retains its activity in high urea concentrations (6-8 M). This allows a predigestion with Lys-C under maximized protein denaturation. After dilution, the resulting peptides can then be digested further with trypsin to yield the favorable tryptic cleavage pattern. This tandem-digestion approach results in improved proteolytic efficiency compared to the use of trypsin alone.[206]

1.2.6 Data handling

The first step in the computational analysis of mass spectrometry data in proteomics experiments is the protein identification and quantification. The protein identification is facilitated by matching the in-silico digestion of protein sequences, in accordance with the cleavage pattern of the applied protease, with the acquired mass spectra. The peptide bonds along the protein backbone break randomly in the collision cell of the mass spectrometer. The fragments resulting from collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD) appear as b-ions (from the N-terminal side) and corresponding y-ions (from the C-terminal side). The peptide sequences can be determined from the mass differences between the various b-ions and y-ions, respectively, which correspond to the characteristic molecular weights of the respective amino acids. Leucine and isoleucine however are indistinguishable with this method.

Multiple steps are necessary to identify differently abundant proteins from raw protein abundances. Perseus[207] is a comprehensive software solution for these tasks. Its flexibility allows to process proteomics data from very different experimental designs, and even non-proteomics data. Perseus offers multiple data visualization tools.

Enrichment analyses are one way to approach the challenging interpretation of comprehensive lists of protein abundance changes from proteomics experiments. Enrichment analyses determine which gene ontology (GO) terms are overrepresented among a specific group of proteins in comparison to the entire dataset. The gene ontology terms comprise for instance biological and molecular functions, cellular localization, pathways, and protein families. The enrichment of such terms among the differently abundant proteins can help to understand which role the respective ontologies play in the disease pathophysiology. Many of the challenges in

the interpretation of proteomics results arise from the high numbers of differently abundant proteins that a modern proteomics approach can identify especially in strong phenotypes. Alternative approaches to enrichment analyses include the focus on only the proteins with the highest abundance changes and the manual interpretation of extensive protein lists. However, the focus on only the highest abundance changes disregards large amounts of data and does not detect moderate changes that accumulate in one function or compartment. The manual interpretation of extensive protein lists is prone to be arbitrary and to be influenced by the background of the interpreting researcher.[208]

In network analyses, interactions or functions that are retrieved from databases serve to create networks. These networks consist of nodes and edges that connect the nodes. Proteins in such analyses can have either role. For instance, when proteins serve as nodes they can be connected by edges that indicate known protein-protein interactions. When gene ontology terms serve as nodes, proteins can serve as edges. Network analyses can help to understand and visualize the interplay of proteins and biological functions that are subject to abundance changes. Cytoscape[209] and its plugin ClueGO[210] create and visualize networks from protein lists. ClueGO uses gene ontology terms as nodes and builds networks with the respective proteins as edges.

1.2.7 Proteomics and protein biomarkers in Ulcerative Colitis

Common aims of proteomics in the context of diseases are the elucidation of the disease pathophysiology and the finding of biomarkers. In a clinical setting, an improved understanding of the pathophysiology is advantageous because it can lead to new and improved treatment options and can help to understand the pathogenesis, which may help to prevent the disease.

Biomarkers are measurable indicators of specific biological states. Clinically they are used in the diagnosis of diseases, in monitoring their activity, for guiding the therapy, to assess the therapeutic response, or to predict the disease progression.[211, 212] Efforts of biomarker discovery in UC with diagnostic purpose aimed to differentiate UC from CD or healthy controls. In the context of treatment the prediction and monitoring of responses to medication, especially biologics, are of interest.[212] Prediction of disease progression is of interest with regard to the intestinal inflammation, EIMs, complications, and relapses. Most potential clinical applications of proteomics are believed to require the assessment of multiple proteins. The rationale behind this is that a single protein has multiple functions in a cell and interacts with other proteins to exert these functions, and that as a result a change in the abundance of one protein may only be relevant when other protein abundances change as well.[212]

The assessment of intestinal inflammation allows clinicians to adjust the treatment. Proteins that are measured in this assessment are often mediators of inflammation, e.g. cytokines and eicosanoids, or other compounds produced as a response to inflammation.[213] Such biomarkers of inflammation in UC are C-reactive protein (CRP), fecal lactotransferrin (LTF,

synonym: lactoferrin), and fecal calprotectin.[11, 214] CRP is useful in acute severe UC. A CRP over 45 mg/l on day 3 after hospital admission for severe colitis together with 3-8 stools/day is predictive for colectomy.[105] Fecal biomarkers have a higher specificity for inflammation located in the intestines.[213] Fecal calprotectin seems to have the highest sensitivity and specificity for active inflammation.[11, 215, 216] It correlates with increased neutrophils in the intestines.[11] Calprotectin is a complex of protein S100A8 and protein S100A9. These markers are not specific for UC, because they only indicate the presence of inflammation (CRP) and intestinal inflammation (fecal biomarkers), respectively.[11, 105] However, fecal calprotectin and CRP can essentially exclude IBD in patients with irritable bowel syndrome (IBS) symptoms. The probability of having IBD is ≤ 1% in patients with low CRP or fecal calprotectin.[214] While many proteins used as biomarkers of inflammation are increased in this state, the serum biomarkers albumin, transferrin, α2 macroglobulin and Factor XII are decreased in the acute phase response.[213] Hypoalbuminemia occurs in severe UC and is a predictor for colectomy and poor response to biologics.[11, 217]

The sample type choice for proteomics of UC follows usually one of three approaches. Colon mucosa is the main site of the disease manifestation and candidate biomarkers may be more concentrated, which reduces the chance of false discoveries.[212] Sample types from the site of inflammation are colon mucosa biopsies, the mucosal-luminal interface, and "proximal fluid".[218-221] The second approach is the analysis of easily accessible body liquids. These sample types are most interesting for clinical application, because they are easy to obtain. For this reason, a blood test is usually the ultimate aim in biomarker discovery.[212] Such samples comprise blood, plasma, serum, and urine.[212, 222-225] The third approach is measuring of bacterial proteins. The bacteria population in the gut is involved in the UC pathogenesis and pathophysiology. Sample types for this purpose are feces and the mucosal-luminal interface.[212, 221, 226]

Several studies show the potential of proteomics to elucidate pathophysiological processes. Different protein abundances found in UC compared to controls implicate roles of inflammation and tissue repair[227], mitochondrial dysfunction[218] and energy metabolism[228], amino acid metabolism (arginine and tryptophan)[229], cellular stress, and signal transduction.[212, 228]

Further results implicate that proteomics can facilitate biomarker discovery. The proteomes of UC patients and healthy controls show clear differences, which indicates feasibility of biomarker development for this purpose.[212, 228, 230] However, validation for clinical applications is necessary.[212] A panel of proteins can differentiate between UC and CD.[219] The prediction of the disease course and response to biologics or the monitoring of treatment response were not addressed by proteomics studies in UC.[212] While the obtained results are promising, many research questions remain unaddressed and none of the proteomic biomarkers is implemented in clinical daily use.[212, 231]

The studies applied different study designs and analytical approaches. This makes comparisons between studies difficult.[212] However, with different designs/approaches and their characteristics taken into account, e.g. ratio compression in TMT labelled samples compared to label-free MS analysis[232], the confirmation of results across different designs and analytical approaches can validate findings.

1.3   Aims of the thesis

The overall hypothesis of this thesis is that through elucidating the properties of the proteome in ulcerative colitis, an improvement of our understanding of the pathophysiology in the active and the remission state will lead to further development through new treatment regimens and treatment options. We also hypothesize that a deep proteomic investigation of treatment naïve patients can lead to the identification of potential biomarker candidates.

We pursued these hypotheses with the following intermediate aims:

- Optimization of the sample preparation of colon mucosa biopsies for bottom-up proteomics
- Identification of the most differently abundant proteins between active UC and healthy controls
- Functional enrichment analyses of differently abundant proteins between UC and healthy controls
- Characterization of the proteome in remission from UC as compared to active UC and healthy controls
- Development of a prediction model for the 1-year outcome of UC severity

2 Materials and Methods

2.1 Patient characteristics

The colon mucosa biopsies from treatment-naïve and newly diagnosed UC patients were collected in Norway. Diagnosis of UC was based on criteria defined in the European Crohn and Colitis Organization (ECCO) guidelines.[105] Subjects undergoing a cancer screening examination without intestinal disease were included as healthy controls. The inflammation was evaluated during colonoscopy based on the ulcerative colitis disease activity index (UCDAI).[114] TNF mRNA was measured by real-time PCR. Geboes index values were determined during histological examination.[115]

The patient characteristics can be summarized as follows:

Table 1 - Characteristics of patients and healthy controls in paper I [1]

	Number of Subjects	Average age (SD)	Female/male	UCDAI score (SD)	Endoscopic score (SD)	Average TNF (SD)
UC onset active	6	48 (14)	1/5	9.5 (2)	2.6 (0.5)	14,188 (12390)
Healthy controls	6	54 (10)	1/5	–	0	2933 (1084)

Table 2 - Characteristics of patients and healthy controls in paper II [2] and III

	Number of Subjects	Average Age (SD)	Female/Male	Average years since UC debut (SD)	Median TNF (IQR)	Average UCDAI Score (SD)	Average Geboes index (SD)
UC remission	14	46.8 (16.0)	5/9	7.4 (4.5)	4800 (3000)	0.4 (0.95)	1.15 (1.7)
UC onset active	17	39.4 (16.7)	4/13	0	14350 (16725)	8.7 (2.2)	7.9 (3.8)
Healthy Control	15	51.9 (14.3)	5/10	-	4500 (2400)	-	-

Table 3 - Characteristics of patients and healthy controls in paper IV

1-year outcome	Number of Subjects	Average age (Range)	Sex (F/M)	Average TNF (Range)	Average Geboes index (Range)
Mild/Moderate	14	40.3 (20-62)	4/10	11,315 (4,600-31,700)	6.5 (3-13)
Severe	5	40.0 (18-68)	1/4	37,880 (17,200-96,000)	12.0 (8-16)

2.2 Ethical approval and participation consent

The study and the storage of biological material were approved under the number REK NORD 2012/1349 by the Regional Committee of Medical Ethics of North Norway and the Norwegian Social Science Data Services. All enrolled subjects signed informed and written consent forms.

2.3 Sample preparation

Paper 1 (Method development) – Colon mucosa biopsies were homogenized in water. Protein concentration was determined with a bicinchoninic acid (BCA) assay. The homogenate was subsequently added to lysis buffers of the following compositions:

- Sodium deoxycholate (5%), 8 M urea, 100 mM TEAB
- Sodium dodecyl sulfate (1%), 8 M urea, 100 mM TEAB
- Sodium laurate (5%), 5.7 M urea, 100 mM TEAB
- Rapigest (0.25%), 8 M urea, 100 mM TEAB
- Cetyltrimethylammonium bromide (1%), 8 M urea, 100 mM TEAB
- *N*-methyldioctylamine (5%), 6 M urea, 100 mM TEAB
- 8 M urea, 100 mM TEAB

Protein solubilization was supported with MagNA Lyser treatment. The proteins were reduced with DTT and alkylated with IAA. Sodium dodecyl sulfate, cetyltrimethylammonium bromide and *N*-methyldioctylamine were removed by modified FASP methods before digestion. Rapigest was removed by acid induced cleavage after digestion. Sodium deoxycholate and sodium laurate were removed after digestion by acid precipitation. Predigestion was performed with Lys-C (enzyme:protein ratio 1:100) for 8 h, followed by digestion with trypsin (enzyme:protein ratio 1:20) for 16 h. The samples were desalted with C18 pipette tips and reconstituted in 0.1% formic acid.

Paper 1 (Clinical samples) – Colon mucosa biopsies were added to the optimized lysis buffer (sodium deoxycholate 5%, 8 M urea. 100 mM TEAB) and homogenized by MagNA Lyser treatment. The protein content for each sample was determined by BCA assay. The proteins were reduced, alkylated, and digested as described above. A standard was produced by combining equal peptide amounts from each sample. The standard and the samples were labelled with TMTsixplex and afterwards combined. Desalting and reconstitution were performed as described above.

Papers 2, 3 and 4 – The samples were treated as described for the clinical samples of paper 1 up until the labelling step. A new standard was produced from all samples included in this experiment. After the samples and the standard were combined, the combined samples were subjected to high pH reversed-phase fractionation, yielding eight fractions each.

2.4 LC-MS/MS analyses

Paper 1 – The peptides were analyzed with an EASY-LC system coupled to a qExactive mass spectrometer. Peptides were separated on a C18 column (2 μm, 100 Å, 50 μm, 50 cm) with a 4-40% gradient of acetonitrile in 0.1% formic acid (FA) over 1 h. The first fixed mass was set to 120 m/z for the clinical samples. These were further subjected to two additional 1 h LC-MS/MS analyses with inclusion lists with the purpose of acquiring better mass spectra for low confidence identifications and to identify and quantify peptides that were identified in one other samples as well.

Paper 2, 3 and 4 – The fractions were analyzed with the instrument described above with a fixed first mass of 120 m/z and a gradient of 0% to 5% over 19 min, further to 30% at 180 min and to 100% at 200 min.

2.5 Data handling

Protein identification and quantification were performed with MaxQuant[233] by searching against fasta files from UniProt[234] that contain the human proteins. Methionine oxidation and N-term acetylation were set as variable modifications. The quantification method was label-free for the method development and set to TMTsixplex for all clinical samples.

The statistical analyses were performed and visualized with Perseus.[207, 235]

For paper 1, the protein intensities were log2(x+1) transformed in Perseus, and batch corrected with the combat method.[236] The entries were filtered for the labels "potential contaminant", "reverse" and "only identified by site". The data was further analyzed by principal component analysis (PCA), and with the limma package, Cluster Profiler and revigo.[237, 238]

For paper 2, the protein intensities were log(2) transformed and the standard intensities were subtracted from the corresponding sample intensities. The resulting intensities were Z-score transformed (matrix access: columns). Significantly different proteins were determined with a two-sample test. For functional analyses gene ontology terms were added from GOBP (biological process), GOMF (molecular function), GOCC (cellular component),[239] KEGG,[240] GSEA,[241] UniProt keywords,[234] InterPro,[242], Reactome,[243] and PROSITE.[244] Enrichment of proteins with increased and decreased, respectively, abundance was determined with a Fisher exact test. The network analyses were performed with the Cytoscape app ClueGO, applying the WikiPathways database.[245]

For paper 3, the protein intensities were normalized as described for paper 2. ANOVA testing with subsequent post-hoc testing determined significant differences and their direction (i.e. increased or decreased abundance) between active UC, healthy controls, and UC remission. The proteins were assigned to groups according to this pattern. The GO annotations described

above were added and categorical enrichments within these groups were determined with a Fisher exact test. Network analyses were performed with the Cytoscape app ClueGO, for which the GOBP database was applied.

For paper 4, the protein intensities were normalized as described for papers 2 and 3. All proteins with missing data were removed from the dataset. Statistical analyses and model development were performed with SIMCA P+ 15.0 (Umetrics, Sweden). Network analysis was performed with StringDB.[246]

3 Summary of results

The Proteome of Ulcerative Colitis in Colon Biopsies from Adults - Optimized Sample Preparation and Comparison with Healthy Controls [1]

Armin Schniers, Endre Anderssen, Christopher Graham Fenton, Rasmus Goll, Yvonne Pasing, Ruth Hracky Paulssen, Jon Florholmen and Terkel Hansen

In this study, we investigated seven lysis buffers to optimize the sample preparation of colon mucosa biopsies. A lysis buffer of 5% sodium deoxycholate, 8 M urea, and 100 mM was superior in the tandem-digestion with Lys-C and trypsin, with high reproducibility and identification numbers. Interestingly, the proportion of identified transmembrane proteins was highest with the two cationic detergents. Our results indicate that sodium laurate and urea inhibit tryptic enzyme activity synergistically.

For the investigation of clinical samples, 6 healthy controls and 6 UC patients at disease onset were included. A PCA separates clearly between the groups. 68 proteins were up- and 100 proteins down-regulated for $p < 0.05$ and fold change (log base 2) $> |0.7|$.

Proteins related to the immune system are of increased abundance in UC. Proteins related to anion transport, epithelial cell adhesion, and mucus production are decreased in active UC.

Seven out of 11 quantified S100 proteins were significantly different abundant, with both increases and decreases in active UC compared to healthy controls. S100A8 and S100A9, which form the heterodimer calprotectin, S100A11, S100A12, and S100P are of increased abundance in UC. The abundances of S100A14 and S100A16 are decreased.
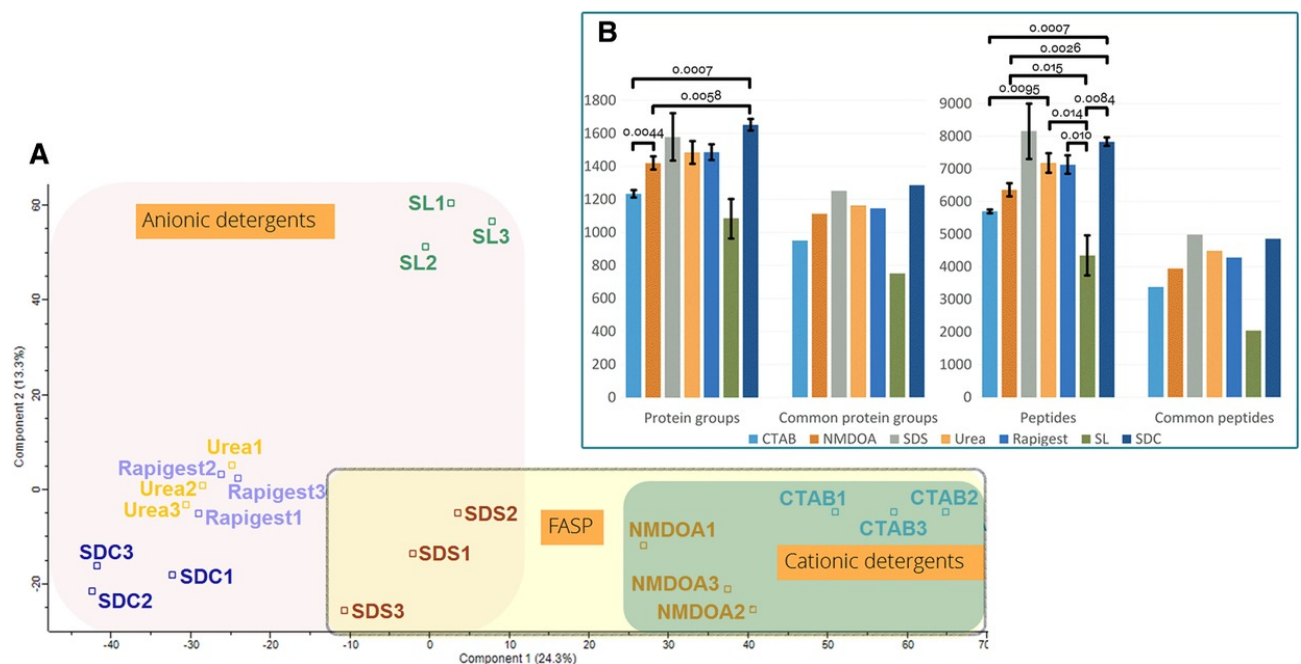
Figure 11 - Comparison of detergents in proteomics of colon mucosa biopsies. (A) PCA shows grouping according to detergent properties and detergent removal method. CTAB = cetyltrimethylammonium bromide, NMDOA = N-methyldioctylamine, SDS = sodium dodecyl sulfate, SL = sodium laurate, SDC = sodium deoxycholate (B) Comparison of protein and peptide ID numbers. Used with permission from [1].

Paper II

Ulcerative colitis: functional analysis of the in-depth proteome [2]

Armin Schniers, Rasmus Goll, Yvonne Pasing, Sveinung Wergeland Sørbye, Jon Florholmen and Terkel Hansen

In this study, we investigated the functional changes in the proteomes of 17 UC patients at debut compared to 15 healthy controls.

The proteome depth achieved in the analyses was enhanced compared to paper I by high pH reversed-phase fractionation of the samples and increased analysis time. This approach resulted in the deepest proteome of colon mucosa biopsies published to-date. 8562 proteins were identified, of which 6818 proteins were quantified in at least 70% of the samples. 321 proteins were of decreased and 275 increased abundance in active UC compared to healthy controls for S0=2 and false discovery rate (FDR) = 0.01 of the normalized values.

Hierarchical clustering of the significantly different proteins revealed 3 protein clusters. The decreased proteins group in one cluster, while the increased proteins group in a major cluster of 247 and a minor cluster of 20 proteins. The minor cluster presents with a more pronounced abundance increase in a subgroup of UC patients. The proteins in this minor cluster include the calprotectin complex proteins S100A8 and S100A9 and further proteins related to neutrophils.

The functional analysis of the dataset shows that proteins associated with metabolism of nutrients, energy, steroids, xenobiotics and carbonate are of decreased abundance in active UC. Decreased in active UC are furthermore metallothioneins, PPAR-inducible proteins, fibrillar collagens and proteins involved in bile acid transport. Increased abundances were found for proteins involved in immune response and protein processing in the endoplasmic reticulum, e.g. unfolded protein response and signal peptidase complex proteins.
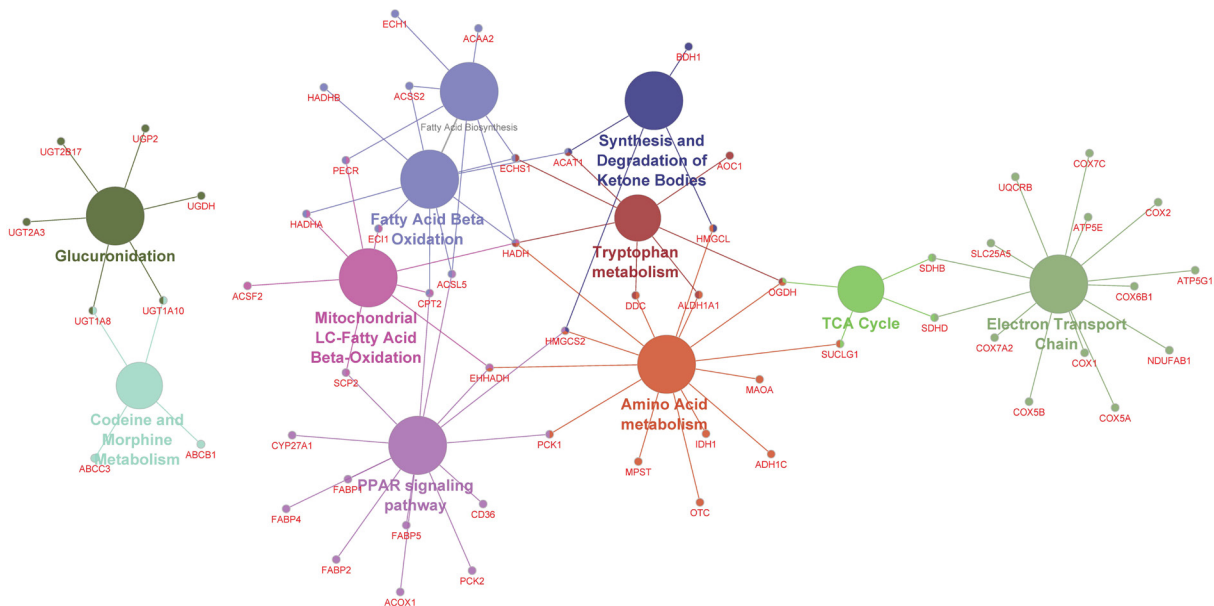
Figure 12 - Functional network of proteins with significantly decreased abundance in active UC compared to healthy controls. Modified from [2]

Paper III

The proteome of ulcerative colitis in remission – functional differences in comparison with healthy controls and the active disease

Armin Schniers, Rasmus Goll, Sveinung Wergeland Sørbye, Jon Florholmen and Terkel Hansen

In this study, we investigated how protein abundances in UC remission relate to those in healthy controls and active UC. We included 14 patients in deep remission from UC, 17 patients with active UC, and 15 healthy controls. 6775 proteins were quantified in over 70% of the samples. 631 proteins were differently abundant in remission compared to active UC, healthy controls, or both. 39 (6%) of these proteins were higher abundant in remission than in healthy controls and active UC. 359 (57%) proteins showed similar abundances in remission as in healthy controls. 82 proteins (13%) had a similar abundance in remission as in active UC. 151 (24%) of the protein abundances in remission lay in between healthy control and active UC abundances. These results show that the UC remission proteome is in most parts a combination of the proteomes of active UC and of healthy controls, with more similarity to healthy controls (see Figure 13).

The functions that return in remission to abundance levels as in healthy controls include the increased abundances of immune processes (except for immunoglobulins) and protein processing in the ER observed in active UC, as well as the decreased levels of fibrillary collagens. The abundances of proteins that are involved in nutrient, energy and xenobiotics metabolism or are associated with PPAR signaling increase in remission significantly compared to active UC. However, they remain lower than in healthy controls.

Metallothioneins remain in remission at similarly decreased levels as in active UC, and most immunoglobulins remain at similarly increased abundances. These persistent changes may contribute to symptoms and relapses in UC remission.
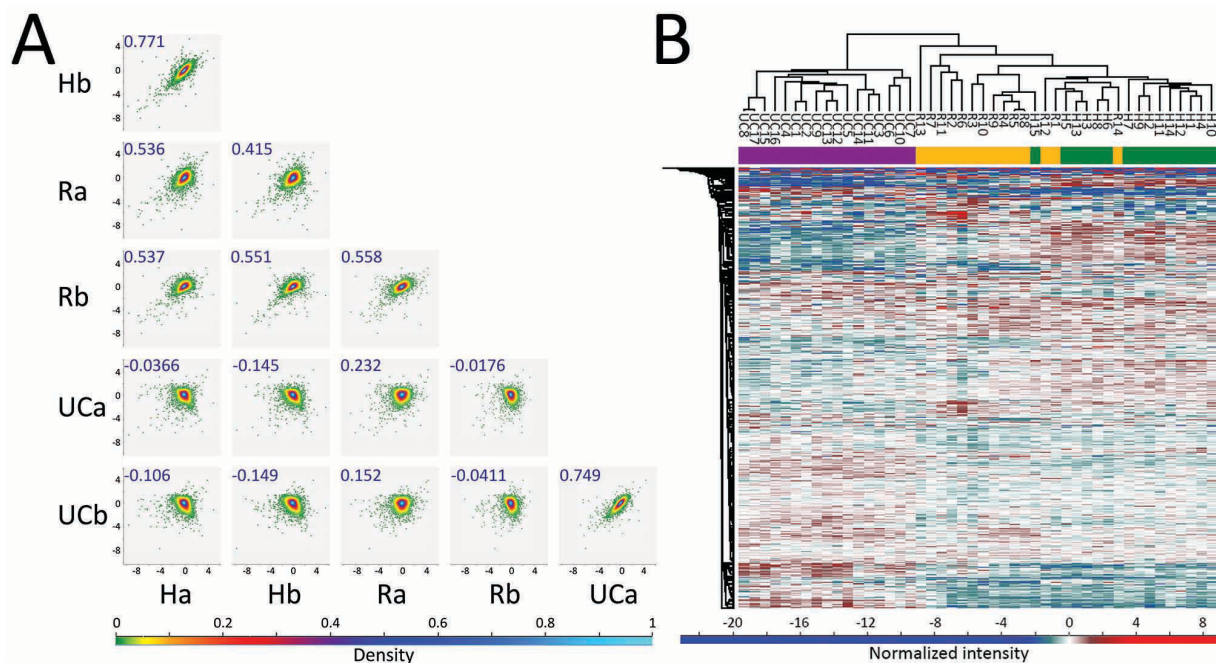
Figure 13 - Exploratory analysis of the proteome in UC remission, active UC, and healthy controls. (A) The scatter plots of averaged abundances and corresponding Pearson correlations show that the proteome in remission is more similar to healthy controls than to active UC. (B) The separation of remission (yellow) and healthy (green) from active UC (violet) and the overlap between remission UC and healthy controls after hierarchical clustering confirm that finding.

Paper IV

Prediction of the 1-year outcome for Ulcerative Colitis from the proteomic profile of treatment naïve patients

Armin Schniers, Rasmus Goll, Sveinung Wergeland Sørbye, Hans Stenlund, Jon Florholmen and Terkel Hansen

In this article, we present a model to predict the one-year outcome for UC patients from biopsies obtained at disease onset. This final model takes into account the 50 most important proteins and separates clearly between mild/moderate and severe outcomes.

Fibrinogens (FGA, FGB and FGG) are at the core in a network of the 50 most important proteins. Fibrinogens have pro-inflammatory properties in several diseases, e.g. multiple sclerosis, Alzheimer's disease, rheumatoid arthritis, kidney fibrosis and several types of cancer.[247] The network core is connected to clusters with functions in splicing, protein translation and mitochondria. An independent network has functions in gut permeability and contains two Rab proteins, Rab13 and Rab22A. Both proteins are involved in inflammation.[248] Mistargeting of Rab13 to basolateral sites has been shown in patients with CD.[249]

The prediction of the disease development at disease onset would be useful in the clinical decision-making. Knowledge that the disease will take a severe course may for instance warrant an early treatment escalation. However, the prediction model needs further validation. To establish clinical applications it would furthermore be necessary to show a benefit of the respective treatment adjustments.

In conclusion, our study indicates that a protein network with fibrinogens at its core is important for the course of UC. This improves our understanding of the disease pathophysiology and could emerge as a useful tool for clinicians.
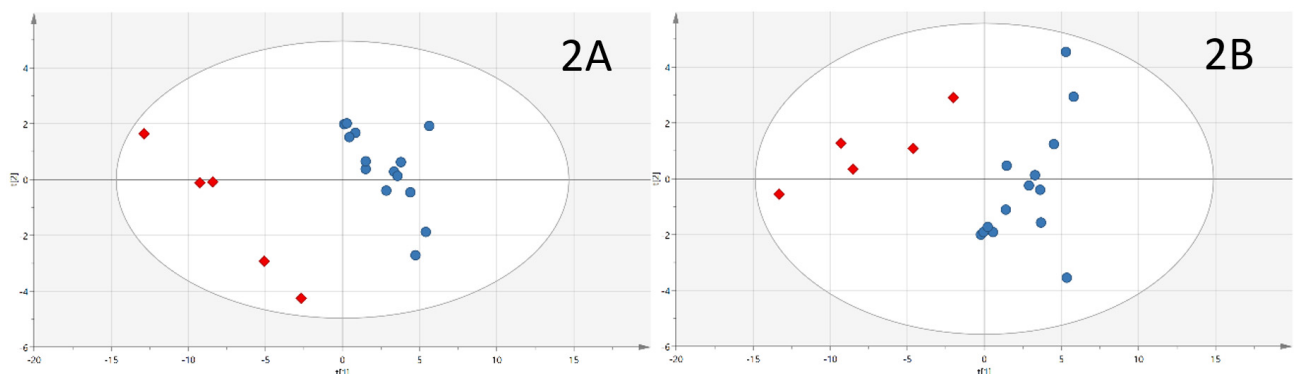
Figure 14 - Partial least squares regression analysis (PLS) (2A) and principal component analysis (PCA) (2B) of samples taking into account the 50 most important proteins for separating mild or moderate outcome from severe outcome. Blue circles represent samples with mild or moderate outcome after one year; red diamonds represent samples with severe outcome after one year.

4 Discussion

4.1 Advanced Systems Biology Analysis of Early events and Clinical Outcomes in Inflammatory Bowel Disease (ASIB) study

The present project is part of the ASIB study. The ASIB study is an international multicenter and multiregional study of IBD. It is a multi-omics study and comprises projects that investigate genomic, transcriptomic, proteomic, metabolomic and lipidomic patterns in IBD patients. The ASIB study is coordinated by Professor Jon Florholmen and funded by UiT – The Artic University of Norway and Helse Nord RHF.

The study includes a biobank, which currently holds colon mucosa biopsies from about 100 UC patients at disease debut and further samples from patients in UC remission, from CD patients, and from healthy controls. The study aims to include in total 340 IBD patients.

The clinical samples are central to the project. One strength of the study is that all omics in this study investigate samples from the same patients and healthy controls, which makes the direct comparison of different omics in the same patients possible. In this regard, for instance the correlation between mRNA and protein abundance changes or between enzymes and their respective metabolites are of interest. The enrolled patients and the disease manifestation are well characterized. The same evaluators perform the characterization of a particular patient over time, because changing evaluators could entail differences in the characterization. The biobank is centralized in one department, which maximizes the control over the sample handling. A current limitation of the ASIB biobank is the lack of other sample types than colon mucosa biopsies. This limitation has been recognized and the collection of additional sample types such as feces and blood is being planned.

The biobank is relatively large for the characteristics of containing biopsies and of originating from untreated patients, but not particularly large when considering only the total number of patients. The clinicians optimize the treatment depending on their evaluation for each patient. Accordingly, the treatment is not randomized. This is necessary to ensure best possible care for each patient, but complicates the evaluation of the study results. The biopsy collection within Norway entails that the enrolled patients show little spread of ethnicity. While this has some advantages because the limitation of non-disease related differences simplifies the interpretation of the results, the lack of different ethnicities is also a major limitation of the study.

Despite some important limitations, the ASIB study design is well tailored for answering a range of so far unresolved issues regarding IBD and the combination of omics disciplines is central to it.

4.2 Patients

In this project, we investigated proteomes in biopsies from UC patients and healthy controls. We included patients with active inflammation at disease onset to investigate the proteome properties of the disease without treatment. In this group, we included patients with different 1-year outcomes to build a model to predict the disease progression. We further included samples from patients in remission from UC to investigate how the remission state compares to the active disease and healthy controls. The control samples were taken from healthy subjects in cancer screening examinations. The subject characterizations are comprehensive and include clinical, endoscopic and histologic scores.

We aimed to select samples with age and gender subject characteristics that were similar between groups we aimed to compare. However, the age is skewed towards a higher age in controls compared to UC patients, which reflects that colon cancer screening is less common in young subjects. The average age of patients in remission is higher than in patients at disease onset, reflecting that the disease onset of the patients in remission was on average 7.4 years before the biopsy was obtained. It was not possible to avoid age differences between the groups with the available samples entirely. However, the ages between the groups overlap strongly and the highest difference in average age between two groups (12.5 years) is moderate. Additionally, as opposed to the disease/healthy states, we did not observe grouping according to age in PCAs, which indicates that age has comparably little effect on the proteome. While a reduction of the differences in the average age would have been desirable, the present differences are acceptable.

The gender ratios female/male are similar between the compared groups. However, the number of male subjects exceeds the number of female subjects. An equal distribution between the genders would have been preferable, but was not achievable with the samples available in the biobank without creating larger differences between the groups. Considering the study aims, we deemed similarity between the groups to have a higher priority and had to accept the low female/male ratio.

The biopsies were obtained in Norwegian hospitals. Accordingly, the ethnicities and life-styles of the subjects reflect the Norwegian society. This can in principal entail results that are not transferable to other populations. This consideration would favor a more diverse study group. On the other hand, a comparably uniform study group, such as the one we investigated, may reduce noise in the data and lead to more results that are meaningful. While the present biobank did not leave a choice between these options, it can also be argued that the advantages of a low spread in factors such as ethnicity and life-styles outweigh the disadvantages for the aims of the present study. However, possible differences in UC manifestations depending on ethnicity or life-style would be an attractive topic for future studies.

The average TNF mRNA and average Geboes index at disease onset were considerably lower for patients with a mild or moderate 1-year outcome than for those with a severe outcome. This was not avoidable in planning the experiments. In line with the previously described characteristics, more similarity between both groups would be advantageous. However, both TNF mRNA and Geboes index values overlap between the outcomes, and the model described in paper IV separates between the outcomes regardless of the TNF mRNA and Geboes index values.

In conclusion, the selection of biopsies was suitable and optimized for answering the questions raised by the hypotheses of the present work, but came with some unavoidable trade-offs.

4.3 Sample preparation of colon mucosa biopsies for proteomics

Our sample preparation approach aimed to achieve optimal protein denaturation, solubilization and extraction. This resulted in quite high concentration of both urea and detergent. A possible drawback of this approach is that denaturing conditions can also impair the digestion efficiency of enzymes. The use of a tandem digestion with Lys-C followed by trypsin helps to tackle this, because Lys-C can tolerate denaturing conditions better than trypsin. The idea behind the tandem digestion is that it allows the predigestion under strongly denaturing conditions, e.g. 6 M urea, to introduce first cleavages by Lys-C into proteins, which yields peptides that are soluble under less denaturing conditions. Subsequent sample dilution to less denaturing condition, e.g. ≤1 M urea, restores conditions compatible with tryptic digestion.

Sodium laurate (SL) has a 12-C carbon tail and a terminal carboxylate function; hence, it can be viewed as the carboxylate analog of SDS. It is the sodium salt of lauric acid, a fatty acid. The carboxylate function remains deprotonated under pH conditions common for enzymatic digestion, but can be quantitatively protonated by acidification. Protonation causes a loss of the negative charge and hence decreases the hydrophilicity, causing the precipitation from aqueous solutions. SL can subsequently be removed by centrifugation, which allows the detergent removal after enzymatic digestion. SL has been studied as a detergent for sample preparation in proteomics. SL has some favorable solubilization properties, but 1% SL inhibits trypsin activity already by 20%.[250]

Sodium deoxycholate (SDC) has repeatedly been shown to be a favorable detergent in lysis buffers for proteomics experiments.[251] SDC is the sodium salt of the bile acid deoxycholic acid. As such, SDC has a sterane skeleton that contributes to its hydrophobic properties. The major hydrophilic moiety of SDC is the carboxylate function in the side-chain at 24-position. Little decrease of trypsin activity is observed for 1% SDC, and an activity of 77.4% was still determined in 10% SDC solution.[252]

Rapigest is an anionic, acid cleavable detergent. The negative charge in Rapigest rests on a sulfonate group. This detergent contains a cyclic acetal. The acetal is stable under the lysis and digestion conditions, but degrades under acidic condition. The degradation products do not interfere with LC-MS/MS analysis. Rapigest is a patented surfactant and marketed at comparably high cost specifically for the use in proteomics. However, other detergents repeatedly outperformed Rapigest in sample preparation studies for proteomics.[251, 253]

Cationic detergents like cetyltrimethylammonium bromide (CTAB) have been applied in PAGE as an alternative to SDS.[254] However, unlike SDS the cationic detergents did not gain a major role in in-solution digestion protocols for proteomics. Many cationic detergents are quaternary ammonium ions. CTAB is the bromide salt of a quaternary ammonium. Its nitrogen carries a positive charge, as it has four bonds – 3 to methyl groups and one to a hexadecyl group.

Furthermore, primary, secondary, or tertiary amines can be protonated to acquire positive charges and thereby detergent properties.

The use of 5% SDC and 8 M urea in the lysis buffer gave the best results in terms of identification numbers and reproducibility. Especially the reproducibility of the method with regard to both the identity and the amount of extracted proteins is essential in projects concerning quantitative analyses from clinical samples. Our finding confirms previous studies that identify SDC as a preferable detergent. However, these studies used SDC in other settings, i.e. without urea and/or not in the context of Lys-C trypsin tandem digestion.[251, 252, 255]

SDS was inferior to SDC, both in reproducibility and average protein identification number. For FASP methods, the detergent removal before digestion removes the possible advantage of increased solubilization due to the detergent. The more tedious sample processing furthermore introduce a potential source for contaminations and bias, which may explain the decreased reproducibility. On the other hand, these samples still benefit from the denaturation introduced to the sample during lysis and additionally from the removal of potentially interfering compounds. Beside this, the respective detergent can no longer interfere with the enzyme activities. The evaluation of the FASP method is further complicated by the variability of available FASP filters in material and shape, which is believed to influence the results and may contribute to the performance differences of FASP between laboratories.

The high numbers of missed cleavages and the low identification numbers with SL confirm previous findings that SL impairs tryptic activity. However, the additional presence of urea seems to potentiate this effect. It is clear from our results that SL is not suitable for an approach that applies urea in addition. However, it seems to be inferior to SDC in other protocols as well.[251]

Cationic detergents were clearly inferior to SDC. However, they had the interesting property of identifying a higher proportion of transmembrane proteins than any of the other tested detergent (not a higher absolute number though). The applied FASP method and filter material might be a cause for the inferiority. Considering that transmembrane proteins are of great biological interest, the development of an alternative approaches for the removal of cationic detergents could be promising. Spin columns with an appropriate resin may be a promising approach.

An interesting observation is that in a PCA not only the samples treated with the same detergent grouped together, but that they were also closer together the more similar the applied detergents were chemically and in their removal method. Accordingly, cationic detergents group separately from anionic detergent, and detergents removed with FASP before digestion group separately from detergents removed by acid induced cleavage or acid precipitation.

51

The study shows that the additional use of a detergent compared to urea alone can improve the Lys-C trypsin tandem digestion significantly. The study however does not compare the performance in the use of detergents with or without urea. Preliminary results comparing SDC with or without urea showed a higher identification number with urea. However, the difference was not as pronounced as in the comparison of urea with or without SDC. We did not confirm this with further replicates and did not acquire results for the performance of other detergents without urea. Future studies could elucidate these issues.

In conclusion, SDC is overall superior to the other tested detergents. The enhanced proportion of transmembrane proteins with the use of cationic detergents is of potential interest, but further optimization is necessary.

4.4 Increased protein abundances in active UC

The method developed in paper I made it possible to extract and analyze proteins with good yields in a reproducible manner. The developed protocol was further used to investigate the proteome of colon biopsies at different stages of UC, however with the addition of a pre-fractionation step by high pH reversed-phase spin cartridges.

Proteins with increased abundance in active UC compared to healthy controls show, as expected, an enrichment for immune system processes and, more surprisingly, protein processing in the endoplasmic reticulum (ER).

Most proteins of increased abundance in active UC return in remission fully (219 proteins) to levels as in healthy controls (healthy controls ≈ remission < active UC). 16 return only partially (healthy controls < remission < active UC) and 32 proteins remain at levels as in active UC (healthy controls < remission ≈ active UC). Immunoglobulin domains are predominant among these 48 proteins that do not return to levels as observed in healthy controls.

4.4.1 Immune system

The inflammation reflects in the proteome mainly with increased abundances of immune system proteins, but also decreases of anti-inflammatory processes can promote inflammation. Proteins of increased abundance that participate in inflammation are associated with phagocytes, T cells, and B cells. These proteins include inducible nitric oxide synthase (NOS2) that produces nitric oxide, metalloproteinases that degrade collagen, and immunoglobulins. Decreased abundances we found that might contribute to the inflammation are those of histamine degrading enzymes. A lack of these enzymes likely contributes to the increased histamine levels observed in UC tissue.[80] These findings are in agreement with current knowledge on the UC pathophysiology.

Remarkable is that a group of 20 proteins (minority cluster) related to neutrophils follows a different abundance pattern than most other proteins with increased abundance in UC. Interestingly, these proteins include the biomarker proteins of the calprotectin complex (S100A8, S100A9) and lactotransferrin, which are widely applied as fecal biomarkers. The abundance of these proteins shows a marked increase in a subset of UC samples compared to other UC samples (see Figure 15A-1). This abundance pattern is representative for only few proteins and the abundances between active UC and healthy controls overlap for these proteins, which is not the case for several proteins of the majority clusters (see Figure 15B). These findings suggest that assessing only the proteins from the minority cluster gives an incomplete picture and that an additional assessment of the majority cluster proteins such as CD38, SLC26A2 or HMGCS2 (see Figure 15A-2,3 and 15B) may be sensible.
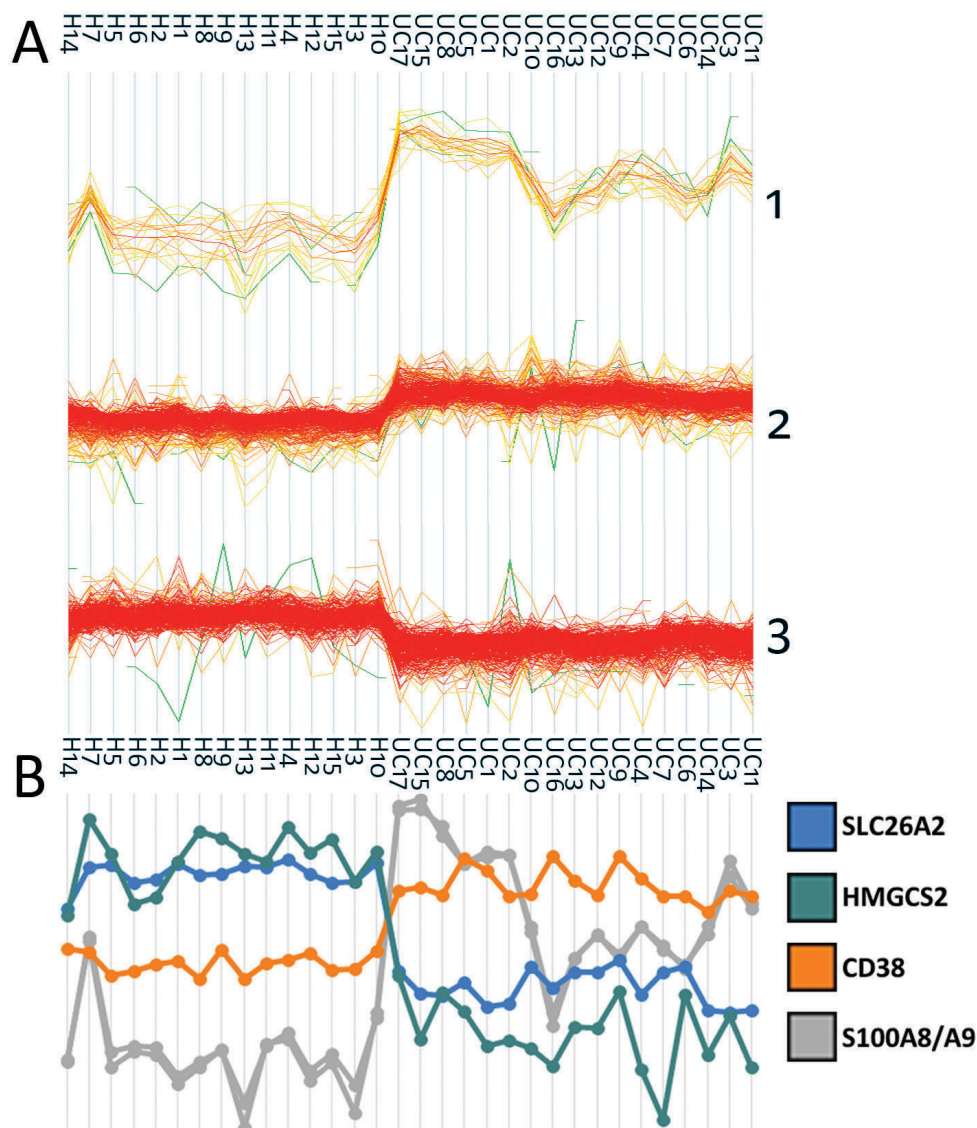
Figure 15 - Profile plots of relative abundances for significantly different abundant proteins in active UC (UC) compared to healthy controls (H). (A) Hierarchical clustering returned the protein clusters 1-3. The proteins of increased abundance in active UC form the minor cluster 1, which includes the biomarkers calprotectin and lactotransferrin, and the major cluster 2, which contains the majority of proteins of increased abundance. Cluster 3 contains proteins of decreased abundance in active UC compared to healthy controls. (B) Abundance changes of proteins from groups 2 (CD38) and 3 (HMGCS2, SLC26A2), selected based on magnitude of fold change and p-value, in comparison with the calprotectin complex proteins S100A8 and S100A9. The calprotectin complex proteins present with high average abundance changes, but the relative abundances overlap between active UC and healthy controls (e.g. H7 presents with higher calprotectin abundance than UC16). The higher consistency of abundance changes for the presented group 2 and 3 proteins could be appealing for clinical applications. Modified from [2]

The abundances of the minority protein cluster correlate with the histological neutrophil assessment, but not strongly with fecal calprotectin levels. A factor that may contribute to the low correlation between biopsy calprotectin with fecal calprotectin is that biopsies represent only a small fraction of the colon, while fecal calprotectin can accumulate from different areas. Biopsy calprotectin furthermore is within the colon mucosa, while fecal calprotectin has left the tissue.

In deep remission from UC, most immune system proteins return to levels as observed in healthy controls. The exception are immunoglobulins, which remain increased in remission compared to healthy controls. Most immunoglobulins remain at levels as in active UC. Only a minority decreases their abundance in UC remission significantly compared to active UC, but also their abundances remain higher than in healthy controls. These findings indicate that the immune system activity is mostly contained in the state of deep remission, with the exception of immunoglobulins.

4.4.2 Protein processing

Proteins related to protein processing in the endoplasmic reticulum (ER) are of increased abundance in active UC compared to healthy controls. The increased abundances observed for protein processing comprise chaperones that are related to unfolded protein response (UPR), e.g. disulfide isomerases that rearrange S-S bonds in proteins, and signal peptidase complex proteins, which cleave the N-terminal signal peptides from maturing proteins.

The UPR leads to an increase in protein folding capacity, increased protein degradation in the ER, and a decreased global protein synthesis.[256] As the term indicates the UPR is induced by unfolded proteins, but as shown in yeast it also has housekeeping functions in the absence of stress conditions in accommodating increased synthesis of cytokines.[257] Unfolded proteins, e.g. mucin-2 (MUC2), and the UPR in immune cells and intestinal epithelial cells are believed to be involved in the IBD pathophysiology. The UPR has ambiguous effects. In intestinal epithelial cells, the UPR effectors can support the cell function and prevent inflammation.[256] On the other hand, the UPR modulates cytokine production, e.g. by activating the pro-inflammatory transcription factors NF-κB and AP-1.[258]

Interestingly, protein processing in the ER was the only major function beside the immune system we found strongly enriched among proteins of increased abundance. This may give ground for giving it increased attention in its function in the UC pathophysiology. One important limitation of our studies with regard to UPR is that they assess the protein abundances in the complete biopsies. For UPR proteins, the cell type-resolved abundances are of interest due to the ambiguous functions of the UPR in epithelial cells and immune cells. Without such data, it is not possible to determine how the total UPR comes into place, i.e. whether immune cell UPR, epithelial cell UPR, or both are of increased abundance in active UC.

The abundance of protein processing proteins decreases under remission to abundances similar to healthy controls. With regard to UPR in immune cells, this is consistent with the finding of immune cell proteins returning in remission to abundances as in healthy controls. A return of epithelial cell UPR could result from a decrease of unfolded proteins, but could also imply a decreased effect of its anti-inflammatory effects in these cells. However, as we did not design our studies to differentiate between epithelial cell UPR and immune cell UPR, further research is necessary to assess the abundance changes in different cell types.

4.5 Decreased protein abundances in active UC

We observe decreased abundances in active UC compared to healthy controls predominantly in proteins related to metabolism, binding, and transport of small chemical compounds. Changes in these functions do not present with an equally obvious phenotype as the inflammation associated with increased immune system protein abundances. However, most of the significant abundance changes we found between active UC and healthy controls are in fact abundance decreases in UC (321 abundances were decreased in UC of in total 596 differently abundant proteins). This shows that the UC pathophysiology has manifold aspects beyond the obvious immune system activation.

Interestingly, most decreased abundances observed in active UC do not return in remission fully to levels as observed in healthy controls. In our studies, 50 protein abundances remain in remission at a similar level as in active UC (healthy controls > remission ≈ active UC), while 135 abundances increase significantly in remission, but show abundances between healthy controls and active UC (healthy controls > remission > active UC). 140 proteins return to levels as in healthy controls (healthy controls ≈ remission > active UC). This is remarkable, because most proteins of increased abundance in active UC return in remission to levels as in healthy controls, with the exception of immunoglobulins.

In conclusion, our findings imply that decreased protein abundances have a major role in the pathophysiology of active UC, and that they could be dominant in the pathophysiology of UC in remission.

4.5.1 Nutrient and energy metabolism

Proteins involved in the utilization of nutrients are of decreased abundance in active UC compared to healthy controls in processes ranging from the transport within the cell, over degradation processes and the citric acid cycle, up to the electron transport chain in the mitochondria. The nutrients with associated decreased metabolic protein abundances are fatty acids including the SCFA butyrate, ketone bodies, and amino acids (valine, leucine, isoleucine, and tryptophan). Fatty acids depend, due to their low solubility, on fatty acid binding proteins (FABPs) for their intracellular transport. We observe decreased abundances in active UC compared to healthy controls for FABP1, FABP2, FABP4, and FABP5. Our results suggest that the decreased transport and metabolic capacities jointly impair the utilization of fatty acids.

Butyrate is the major energy source for colonocytes. Previous studies observed an impaired utilization of butyrate. Based on this, an energy deficiency was suggested as a factor in the UC pathophysiology. Besides, colonocytes convert a fraction of the butyrate to the ketone body β-hydroxybutyrate;[91] hence, the butyrate metabolism is intertwined with the ketone body metabolism. β-hydroxybutyrate is like other ketone bodies an energy source, but it also acts as an endogenous histone deacetylase inhibitor. The histone deacetylase inhibition by

β-hydroxybutyrate leads to increased acetylation at the promoters of FOXO3A and MT2 and to increased transcription of these genes. Accordingly, treatment of mice with β-hydroxybutyrate provides protection from oxidative stress.[93] Previous studies applied SCFA including butyrate to treat patients with UC, but the results are inconclusive.

We believe that treatment with β-hydroxybutyrate could be a more promising approach, because our results suggest that the conversion of butyrate to β-hydroxybutyrate is impaired. Furthermore, our data shows decreased levels of the above-mentioned metallothionein-2 (MT2) in active UC. MT2 is among 50 proteins that remain in remission at similarly decreased levels as in active UC (H > R ≈ UC), whereas most other proteins are of significantly different abundance in remission compared to active UC. Our results indicate that MT2 abundances are non-responsive to current treatment in remission. Treatment with β-hydroxybutyrate could be an elegant approach to restore MT2 abundances to levels as in healthy controls and simultaneously improve the colonocyte nutrition. β-hydroxybutyrate has previously shown promising results and safety in the treatment of other pathological conditions.[259, 260] However, it should be noted and carefully considered that the role of metallothioneins in IBD is highly controversial[94], see the section on metallothioneins below.

The citric acid cycle (synonyms tricarboxylic acid cycle, TCA cycle, and Krebs cycle) is a metabolic pathway that oxidizes acetyl-CoA to two equivalents CO_2 and in turn reduces three equivalents NAD^+ to NADH and one equivalent FAD to $FADH_2$, and converts one equivalent GDP to GTP. Acetyl-CoA for the TCA cycle is supplied by the degradation of a wide range of nutrients, including fatty acids, ketone bodies, carbohydrates, and amino acids. Accordingly, the TCA cycle is central for the utilization of these nutrients. The NADH and $FADH_2$ produced in the TCA cycle are further processed in the electron transport chain (ETC) to produce easily usable chemical energy in form of ATP. The decreased levels of proteins associated with the TCA cycle and the ETC suggest impaired ATP production capabilities with little nutrient specificity, which occur in addition to the impairments of more nutrient specific metabolic pathways.

In remission, we observe increased abundances compared to active UC for proteins that are involved in the nutrient and energy metabolism. However, the abundances remain significantly below those in healthy controls (healthy controls > UC remission > active UC). This suggests that the proposed energy deficiency in active UC is only partially resolved in the remission state. For this reason, it seems likely that it continues to influence the clinical presentation and development of UC in the remission state.

4.5.2 Peroxisome proliferator-activated receptor

Peroxisome proliferator-activated receptors (PPARs) are nuclear receptors that are stimulated by small lipophilic ligands. They have major regulatory roles in energy homeostasis and energy metabolism. PPARs have protective or detrimental effects in a range of different

pathophysiological conditions, including anti-inflammatory properties of PPAR-α and PPAR-γ.[84] Colonic epithelial cells express PPAR-γ. Activation of these receptors by the agonist 15d-PGJ2 inhibits the activation of nuclear factor-κB in an IκB-α–dependent way. It thereby decreases the expression of cytokines in colon cancer cell lines. Inflammation in IBD mouse models decreased upon treatment with thiazolidinedione agonists of PPAR-γ.[90] 5-ASA, which is widely used in IBD, is believed to exert some of its effects through PPAR-γ.[128]

Our results show decreases of proteins related to PPAR signaling in active UC compared to healthy controls. This includes proteins that are involved in the transport of ligands to PPARs, especially the previously mentioned FABPs, and proteins whose expression PPARs increase. These findings suggest an impaired PPAR signaling in active UC. Accordingly, our findings support that PPAR agonists could be promising agents for the treatment of active UC.[90]

We further show that proteins associated with PPAR signaling increase significantly in remission compared to active UC, but remain at decreased abundances compared to healthy controls (healthy controls > UC remission > active UC). This implies that the impaired PPAR signaling is not fully resolved in remission and that accordingly PPAR agonists could be of interest for maintenance treatment.

4.5.3 Metallothioneins

Metallothioneins (MTs) are small, cysteine-rich, heavy metal binding proteins. With their heavy metal binding properties, they function in the detoxification of toxic heavy metals, the utilization of essential heavy metals, and protect from oxidative stress.[94, 261] They furthermore modulate the activity of NF-κB and of immune cells.[94]

Our experiments show that MTs are of decreased abundance in active UC compared to healthy controls. Even more interestingly, as opposed to most other proteins, their abundances do not increase in remission and instead remain at the levels observed in active UC. This may contribute to symptoms observed in the remission state or to relapses.

Zinc homeostasis by redistribution of the intracellular zinc pool is an extensively investigated function of MTs. The MTs have opposite effects on the availability on zinc. They can increase the intracellular zinc levels to enable reactions of the acute phase response and they transport zinc within the cell to release it at zinc-requiring proteins, but they can also sequester zinc to increase the activity of enzymes that are inhibited by zinc.[261, 262] Zinc is essential for the normal development and function of neutrophils, macrophages, NK cells, B cells, and T cells. Zinc is necessary for the activation of the lipopolysaccharide induced NF-κB pathway.[263] On the other hand, zinc has anti-inflammatory and anti-oxidant effects. For instance, zinc induces protein A20 (TNFAIP3) that inhibits the pro-inflammatory NF-κB, and zinc deficiency can lead to an increased production of pro-inflammatory cytokines.[262, 264] Both the effects

of zinc on immune processes and the effects of MTs on zinc availability are ambiguous. Accordingly, the effects that altered zinc binding resulting from decreased MT abundances in UC has on the inflammation process are not clear.

The antioxidant effects of MTs alter inflammatory processes through the sequestration of harmful oxygen and nitrogen intermediates and through modulation of NF-κB activity.[94] They are exerted by cysteine residues, which get oxidized in the process.[265] Increased levels of reactive oxygen species (ROS) and of oxidative damage in proteins and RNA are present in IBD.[266] Neutrophils and macrophages produce ROS (superoxide anion, hydrogen peroxide, nitric oxide) to kill bacteria and parasites during infections. However, these agents are cytotoxic to host cells. MTs protect host cells from these compounds. Thereby, they prevent damage and allow survival and growth in inflammation.[94]

The role of MTs in IBD is obscure. Opposite findings on the expression of MTs in IBD patients exist, but most find decreased abundances in IBD compared to healthy controls. Results from animal models are similarly inconclusive. They include findings of increased, unchanged, and decreased abundances of MTs in response to induced colitis.[94] Two studies on MTnull and wild type mice found no protective or even unfavorable effects of MTs in DSS-induced colitis.[94, 267, 268] On the other hand, MTs showed favorable effects in other inflammation animal models, including *Helicobacter pylori*-induced gastritis [269], lipopolysaccharide (LPS)-induced acute lung injury [270], collagen-induced arthritis [271], and LPS-induced lethal shock.[272] On the other hand, MTs sensitize for TNF-induced lethal shock.[94, 273]

An improved understanding of MTs in IBD could help in the development of new treatment strategies. MTs and some associated functions are compelling treatment targets, because they are readily addressable with existing compounds. For instance, zinc and β-hydroxybutyrate can induce MTs[93, 274], chelation of toxic heavy metals can contain their effects[275], supplementation can increase the availability of beneficial polyvalent metal ions[276], and antioxidants[277] could potentially replace decreased anti-oxidative capabilities from MTs.

Waeytens et al. raised the question whether the regulation of MTs in IBD is dependent or independent on inflammation.[94] Our finding that the abundances remain at similarly decreased abundances in the absence of inflammation in remission suggests that the abundance changes are independent from acute inflammation.

4.5.4 Prostaglandin metabolism

The cyclooxygenase (COX) isoforms COX1 and the inducible COX2 generate prostaglandins from arachidonic acid. Prostaglandins are important mediators of inflammation, but can also have anti-inflammatory effects. They appear to participate in both the acute inflammation and the resolution of inflammation. COX2-KO mice show little resolution of the inflammation and higher numbers of immune cells than wild type and COX1-KO mice 7 days after induction of

inflammation.[278] Inhibitors of COX2 in rats decreased the initial inflammatory response at 3 hours when injected one hour before induction of inflammation, but exacerbated inflammation after 48 hours when injected 24 hours after induction of inflammation. Accordingly, the study concludes that COX2 may be pro-inflammatory in the early phase of inflammation, but that it may help to resolve inflammation in later phases.[279]

Our data showed that prostaglandin metabolizing enzymes are enriched among the proteins that remain at similarly decreased abundances in remission as in active UC (healthy controls > UC remission ≈ active UC). The proteins aldo-keto reductase family 1 member C3 (AKR1C3) and peroxisomal acyl-coenzyme A oxidase 1 (ACOX1) are involved in the degradation of prostaglandins. 15-hydroxyprostaglandin dehydrogenase (HPGD) is involved in the formation of prostaglandin $F_{2\alpha}$ from prostaglandin D_2 and prostaglandin H.[280, 281] Prostaglandin H_2 is formed from arachidonic acid and a common intermediate in the formation of prostaglandin D_2, prostaglandin E_2, prostaglandin F_2, prostaglandin I_2, and thromboxane A_2 (TXA_2). Prostaglandin D2 has pro-inflammatory and anti-inflammatory effects, and it may contribute to the resolution of inflammation. Prostaglandin $F_{2\alpha}$ has a prominent role in the female reproductive system, but also has pro-inflammatory effects.[282]

It should be noted that further identified metabolic proteins of prostaglandins have other abundance patterns. For instance, prostaglandin E synthase (PTGES) is similarly abundant in UC remission and healthy controls, but significantly higher abundant in active UC (healthy controls ≈ UC remission < active UC). PTGES forms prostaglandin E_2 from prostaglandin H_2.[283] Prostaglandin reductase 2 (PTGR2) is significantly higher abundant in healthy controls than in UC remission and significantly higher abundant in UC remission than in active UC (healthy controls > UC remission > active UC). PTGR2 is involved in the inactivation of prostaglandins.[284]

Prostaglandins have long been known as a factor in IBD. Also their involvement in the remission state of IBD has been recognized previously.[285] Our findings show that several prostaglandin degradation pathways are decreased in UC remission and active UC compared to healthy controls. We furthermore observe differential abundances of enzymes converting between different prostaglandins. These results support that prostaglandins have a role in both active UC and UC remission and they provide additional explanations for the differential abundances of prostaglandins.

4.6 Prediction of 1-year outcome

We present the first proteomic prediction model that separates mild/moderate disease courses from severe outcomes one year after diagnosis. Prediction of the disease course would be a useful tool for clinicians. For instance, it would allow for a "hit hard and early" treatment strategy for patients that would present a severe disease course, which has been hypothesized to be beneficial.[286]

The prediction model takes the 50 most important proteins into account. These proteins form a major network with fibrinogens at its core (Figure 16, network A) that further comprises proteins that are involved in splicing and protein translation (Figure 16, network B), mitochondrial proteins (Figure 16, network C) and collagen metabolism (MMP2, SERPINH1). A smaller network contains proteins involved in gut permeability, including RAB13 and RAB22A (Figure 16, network D).

Several of the 50 most important proteins have roles in inflammation, but our data suggests that proteins with directly inflammatory functions are only one of multiple factors determining the disease course. This is in line with our observation that inflammatory proteins are only one facet in the pathophysiology of both active UC and UC remission.
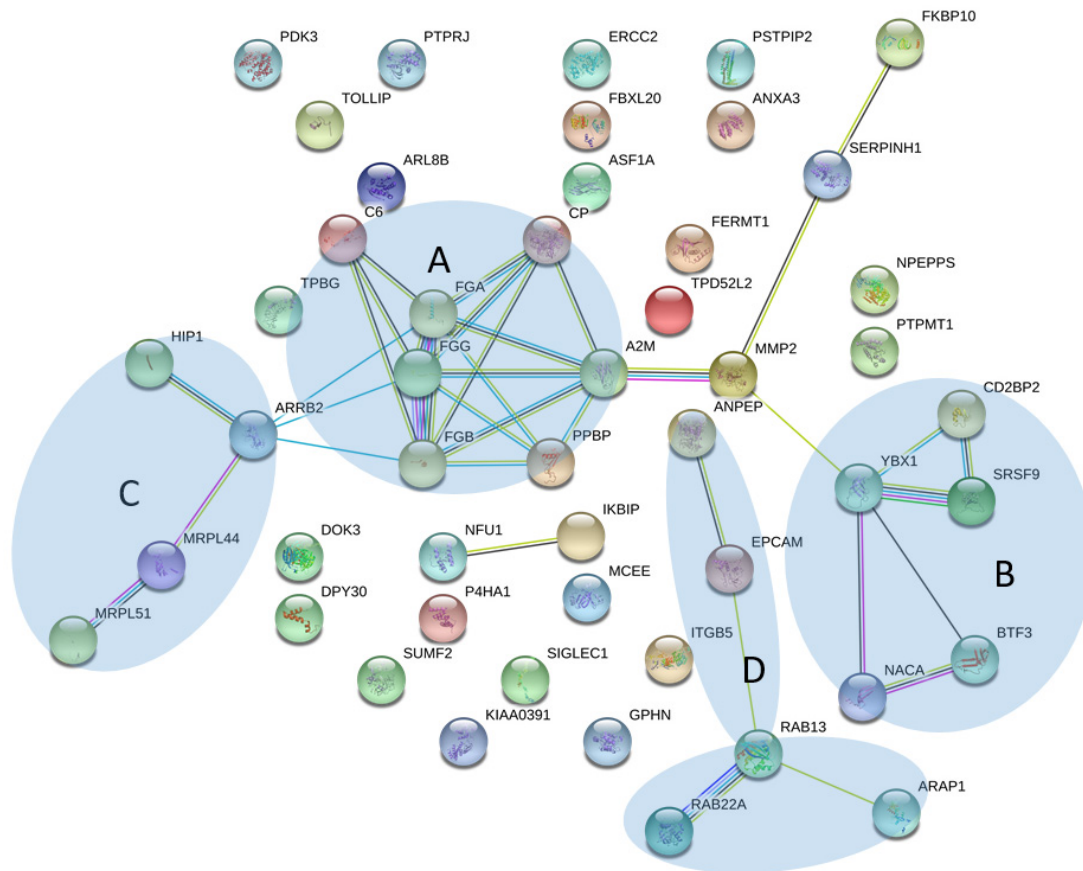
Figure 16 - Network of the most important proteins for separating mild or moderate from severe 1-year-outcomes from biopsies obtained at disease debut. From manuscript IV.

Despite the potential use of the model, further studies are necessary before establishing it in clinical applications. These studies are necessary to validate the model and characterize it in terms of sensitivity and specificity. A further measure to enhance the clinical use would be to optimize the treatment adjustment in dependence of the predictions from the model. Changes of the analytical method (e.g. to ELISA based approaches) and sample material (e.g. blood instead of colon mucosa biopsies) would benefit the usability.

In conclusion, we present the first proteomic model to predict the 1-year outcome at UC debut. The model could become a useful tool, but more research is necessary to establish its use in clinical practice.

4.7 Potential clinical utility

Our results show that immune processes in the deep remission state of UC are contained, whereas several other functions remain unresolved or only partially resolved. Addressing these functions may help in both reaching and maintaining the remission state. The impaired nutrient utilization capacities of colonocytes imply that supplementation of readily usable energy sources such as β-hydroxybutyrate could be beneficial. β-hydroxybutyrate may exert further beneficial effects by inducing metallothioneins, even though the effects of metallothioneins in the context of IBD warrant further evaluation. PPAR pathways are overlapping with the nutrient utilization, because the fatty acid-binding FABPs are involved in PPAR signaling and because PPAR regulates metabolic pathways. Our data indicate that the previous hypothesis that PPAR-γ agonists could ameliorate colitis [89, 90] may be applicable to UC.

We observe two distinct abundance patterns among proteins with increased abundance in active UC. Only a minority of proteins of increased abundances in active UC follows the abundance pattern represented by the inflammation biomarkers lactotransferrin and calprotectin (proteins S100A8 and S100A9). Proteins with this abundance pattern are associated with neutrophils and the abundance of these proteins correlates with histological findings on neutrophil infiltration. The observation that they represent only a minority of proteins of increased abundance suggests that an additional measurement of proteins, which represent the majority of proteins of increased abundance, may carry further useful information. The same applies to proteins of decreased abundances. While the abundance changes in the minority group of proteins are among the highest average fold changes, the fold changes of the majority proteins are more consistent within the conditions active UC and healthy controls, respectively (see Figure 15). Knowledge on the abundance changes of these proteins in other diseases could further enhance their use in diagnostics. However, biopsies represent only a small area of the colon mucosa. A common sample type for calprotectin and lactotransferrin measurements is feces, which accumulates protein from a greater area and is easier to sample. Measurement in sample types such as feces or blood furthermore depend on the proteins leaking out of the tissue in question. The findings from biopsies may accordingly not reflect in feces or blood.

The prediction of the 1-year outcome could be of great clinical use. Validation of the model is necessary before implementation in clinical practice. A subsequent step should be to optimize the treatment adjustment according to the prediction. The hypothesis that a top-down treatment approach could benefit patients that are prone to a severe disease course offers a straightforward approach to this.

5   Conclusion and hypothesis evaluation

In this thesis the properties of the proteome in ulcerative colitis were elucidated, which resulted in a great improvement of our understanding of the pathophysiology in the active and the remission state of the disease. The present work suggests approaches to improve diagnostics and treatment, but their implementation in the clinical practice demands further research.

The findings might help to identify and to prioritize potential treatment targets. We also showed that it is possible to predict the 1-year outcome of patients from the proteomic profile of treatment naïve patients. However, validation of this method is still needed and potentially transferring the biomarkers to easier accessible sample types such as feces or blood would further enhance the possible usefulness.

Citations

1. Schniers A, Anderssen E, Fenton Christopher G, Goll R, Pasing Y, Paulssen Ruth H, Florholmen J, Hansen T: **The Proteome of Ulcerative Colitis in Colon Biopsies from Adults - Optimized Sample Preparation and Comparison with Healthy Controls**. *PROTEOMICS - Clinical Applications* 2017, **11**(11-12):1700053.

2. Schniers A, Goll R, Pasing Y, Sorbye SW, Florholmen J, Hansen T: **Ulcerative colitis: functional analysis of the in-depth proteome**. *Clin Proteomics* 2019, **16**:4.

3. Danese S, Fiocchi C: **Ulcerative Colitis**. 2011, **365**(18):1713-1725.

4. Fakhoury M, Negrulj R, Mooranian A, Al-Salami H: **Inflammatory bowel disease: clinical aspects and treatments**. *Journal of inflammation research* 2014, **7**:113-120.

5. da Silva BC, Lyra AC, Rocha R, Santana GO: **Epidemiology, demographic characteristics and prognostic predictors of ulcerative colitis**. *World journal of gastroenterology* 2014, **20**(28):9458-9467.

6. Baumgart DC, Sandborn WJ: **Inflammatory bowel disease: clinical aspects and established and evolving therapies**. *The Lancet* 2007, **369**(9573):1641-1657.

7. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, Panaccione R, Ghosh S, Wu JCY, Chan FKL *et al*: **Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies**. *The Lancet* 2017, **390**(10114):2769-2778.

8. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, Benchimol EI, Panaccione R, Ghosh S, Barkema HW *et al*: **Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review**. *Gastroenterology* 2012, **142**(1):46-54.e42; quiz e30.

9. Danese S, Sans M, Fiocchi C: **Inflammatory bowel disease: the role of environmental factors**. *Autoimmunity Reviews* 2004, **3**(5):394-400.

10. Kamm MA: **Rapid changes in epidemiology of inflammatory bowel disease**. *The Lancet* 2017, **390**(10114):2741-2742.

11. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF: **Ulcerative colitis**. *Lancet* 2017, **389**(10080):1756-1770.

12. Cosnes J, Gower–Rousseau C, Seksik P, Cortot A: **Epidemiology and Natural History of Inflammatory Bowel Diseases**. *Gastroenterology* 2011, **140**(6):1785-1794.e1784.

13. Loftus EV: **Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences**. *Gastroenterology* 2004, **126**(6):1504-1517.

14. Bernstein CN, Rawsthorne P, Cheang M, Blanchard JF: **A population-based case control study of potential risk factors for IBD**. *Am J Gastroenterol* 2006, **101**(5):993-1002.

15. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Philip Schumm L, Sharma Y, Anderson CA *et al*: **Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease**. *Nature* 2012, **491**:119.

16. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T *et al*: **Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations**. *Nat Genet* 2015, **47**(9):979-986.

17. Straub RH, Besedovsky HO: **Integrated evolutionary, immunological, and neuroendocrine framework for the pathogenesis of chronic disabling inflammatory diseases**. *The FASEB Journal* 2003, **17**(15):2176-2183.

18. Sopori M: **Effects of cigarette smoke on the immune system**. *Nature Reviews Immunology* 2002, **2**:372.

19. Abraham C, Cho JH: **Inflammatory Bowel Disease**. 2009, **361**(21):2066-2078.

20. Taurog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, Fernández-Sueiro JL, Balish E, Hammer RE: **The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats**. 1994, **180**(6):2359-2364.

21. Wijmenga C: **Expressing the Differences between Crohn Disease and Ulcerative Colitis**. *PLOS Medicine* 2005, **2**(8):e230.

22. Kim DH, Cheon JH: **Pathogenesis of Inflammatory Bowel Disease and Recent Advances in Biologic Therapies**. *Immune Network* 2017, **17**(1):25-40.

23. Helander HF, Fandriks L: **Surface area of the digestive tract - revisited**. *Scand J Gastroenterol* 2014, **49**(6):681-689.

24. McGhee JR, Fujihashi K: **Inside the Mucosal Immune System**. *PLOS Biology* 2012, **10**(9):e1001397.

25. Sartor RB: **Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis**. *Nat Clin Pract Gastroenterol Hepatol* 2006, **3**(7):390-407.

26. Kempski J, Brockmann L, Gagliani N, Huber S: **TH17 Cell and Epithelial Cell Crosstalk during Inflammatory Bowel Disease and Carcinogenesis**. 2017, **8**(1373).

27. Dalziel TK: **Chronic Interstitial Enteritis**. *The British Medical Journal* 1913, **2**(2756):1068-1070.

28. Hansen R, Thomson JM, El-Omar EM, Hold GLJJoG: **The role of infection in the aetiology of inflammatory bowel disease**. 2010, **45**(3):266-276.

29. Gradel KO, Nielsen HL, Schønheyder HC, Ejlertsen T, Kristensen B, Nielsen H: **Increased Short- and Long-Term Risk of Inflammatory Bowel Disease After Salmonella or Campylobacter Gastroenteritis**. *Gastroenterology* 2009, **137**(2):495-501.

30. Hugot J-P, Alberti C, Berrebi D, Bingen E, Cézard J-P: **Crohn's disease: the cold chain hypothesis**. *The Lancet* 2003, **362**(9400):2012-2015.

31. Elliott DE, Urban JJ, Argo CK, Weinstock JV: **Does the failure to acquire helminthic parasites predispose to Crohn's disease?** *Faseb j* 2000, **14**(12):1848-1855.

32. Weinstock JV, Elliott DE: **Helminths and the IBD hygiene hypothesis**. *Inflamm Bowel Dis* 2009, **15**(1):128-133.

33. Sartor RB: **Microbial Influences in Inflammatory Bowel Diseases**. *Gastroenterology* 2008, **134**(2):577-594.

34. Erichsen K, Milde AM, Arslan G, Helgeland L, Gudbrandsen OA, Ulvik RJ, Berge RK, Hausken T, Berstad A: **Low-dose oral ferrous fumarate aggravated intestinal inflammation in rats with DSS-induced colitis**. *Inflamm Bowel Dis* 2005, **11**(8):744-748.

35. Zhang Y-Z, Li Y-Y: **Inflammatory bowel disease: pathogenesis**. *World journal of gastroenterology* 2014, **20**(1):91-99.

36. Wehkamp J, Salzman NH, Porter E, Nuding S, Weichenthal M, Petras RE, Shen B, Schaeffeler E, Schwab M, Linzmeier R *et al*: **Reduced Paneth cell α-defensins in ileal Crohn's disease**. 2005, **102**(50):18129-18134.

37. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nuñez G, Flavell RA: **Nod2-Dependent Regulation of Innate and Adaptive Immunity in the Intestinal Tract**. 2005, **307**(5710):731-734.

38. Hisamatsu T, Suzuki M, Reinecker H-C, Nadeau WJ, McCormick BA, Podolsky DK: **CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells**. *Gastroenterology* 2003, **124**(4):993-1000.

39. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J *et al*: **A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1**. *Nature Genetics* 2006, **39**:207.

40. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D *et al*: **Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility**. *Nat Genet* 2007, **39**(7):830-832.

41. The UKIBDGC, Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A, Wesley E, Parnell K, Zhang H *et al*: **Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region**. *Nature Genetics* 2009, **41**:1330.

42. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM *et al*: **Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease**. *Nature genetics* 2008, **40**(8):955-962.

43. Di Sabatino A, Lenti MV, Giuffrida P, Vanoli A, Corazza GR: **New insights into immune mechanisms underlying autoimmune diseases of the gastrointestinal tract**. *Autoimmunity Reviews* 2015, **14**(12):1161-1169.

44. Sands BE, Kaplan GG: **The Role of TNFα in Ulcerative Colitis**. *The Journal of Clinical Pharmacology* 2007, **47**(8):930-941.

45. Clark IA: **How TNF was recognized as a key mechanism of disease**. *Cytokine & growth factor reviews* 2007, **18**(3):335-343.

46. Breese EJ, Michie CA, Nicholls SW, Murch SH, Williams CB, Domizio P, Walker-Smith JA, Macdonald TT: **Tumor necrosis factor α-producing cells in the intestinal mucosa of children with inflammatory bowel disease**. *Gastroenterology* 1994, **106**(6):1455-1466.

47. Wéra O, Lancellotti P, Oury C: **The Dual Role of Neutrophils in Inflammatory Bowel Diseases**. *Journal of clinical medicine* 2016, **5**(12):118.

48. Műzes G, Molnár B, Tulassay Z, Sipos F: **Changes of the cytokine profile in inflammatory bowel diseases**. *World journal of gastroenterology* 2012, **18**(41):5848-5861.

49. Bradley JR: **TNF-mediated inflammatory disease**. *The Journal of Pathology* 2008, **214**(2):149-160.

50. Tsujimoto M, Yip YK, Vilcek J: **Tumor necrosis factor: specific binding and internalization in sensitive and resistant cells**. *Proceedings of the National Academy of Sciences* 1985, **82**(22):7626.

51. Rollins BJ, Yoshimura T, Leonard EJ, Pober JS: **Cytokine-activated human endothelial cells synthesize and secrete a monocyte chemoattractant, MCP-1/JE**. *The American journal of pathology* 1990, **136**(6):1229-1233.

52. Osawa Y, Nagaki M, Banno Y, Brenner DA, Asano T, Nozawa Y, Moriwaki H, Nakashima S: **Tumor necrosis factor alpha-induced interleukin-8 production via NF-kappaB and phosphatidylinositol 3-kinase/Akt pathways inhibits cell apoptosis in human hepatocytes**. *Infection and immunity* 2002, **70**(11):6294-6301.

53. Harris DP, Bandyopadhyay S, Maxwell TJ, Willard B, DiCorleto PE: **Tumor Necrosis Factor (TNF)-α Induction of CXCL10 in Endothelial Cells Requires Protein Arginine Methyltransferase 5 (PRMT5)-mediated Nuclear Factor (NF)-κB p65 Methylation**. *Journal of Biological Chemistry* 2014, **289**(22):15328-15339.

54.   Strieter RM, Kunkel SL, Showell HJ, Remick DG, Phan SH, Ward PA, Marks RM: **Endothelial cell gene expression of a neutrophil chemotactic factor by TNF-alpha, LPS, and IL-1 beta**. *Science* 1989, **243**(4897):1467-1469.

55.   Fujino S, Andoh A, Bamba S, Ogawa A, Hata K, Araki Y, Bamba T, Fujiyama Y: **Increased expression of interleukin 17 in inflammatory bowel disease**. *Gut* 2003, **52**(1):65-70.

56.   Huang Y, Chen Z: **Inflammatory bowel disease related innate immunity and adaptive immunity**. *American journal of translational research* 2016, **8**(6):2490-2497.

57.   lvez J: **Role of Th17 Cells in the Pathogenesis of Human IBD**. *J ISRN Inflammation* 2014, **2014**:14.

58.   Hue S, Ahern P, Buonocore S, Kullberg MC, Cua DJ, McKenzie BS, Powrie F, Maloy KJ: **Interleukin-23 drives innate and T cell-mediated intestinal inflammation**. *The Journal of experimental medicine* 2006, **203**(11):2473-2483.

59.   Floss DM, Schroder J, Franke M, Scheller J: **Insights into IL-23 biology: From structure to function**. *Cytokine & growth factor reviews* 2015, **26**(5):569-578.

60.   Aaronson DS, Horvath CM: **A road map for those who don't know JAK-STAT**. *Science* 2002, **296**(5573):1653-1655.

61.   McGeachy MJ, Cua DJ: **Th17 Cell Differentiation: The Long and Winding Road**. *Immunity* 2008, **28**(4):445-453.

62.   Izcue A, Hue S, Buonocore S, Arancibia-Carcamo CV, Ahern PP, Iwakura Y, Maloy KJ, Powrie F: **Interleukin-23 restrains regulatory T cell activity to drive T cell-dependent colitis**. *Immunity* 2008, **28**(4):559-570.

63.   Lai LJ, Shen J, Ran ZH: **Natural killer T cells and ulcerative colitis**. *Cellular Immunology* 2018.

64.   Gerlach K, Hwang Y, Nikolaev A, Atreya R, Dornhoff H, Steiner S, Lehr H-A, Wirtz S, Vieth M, Waisman A *et al*: **TH9 cells that express the transcription factor PU.1 drive T cell–mediated colitis via IL-9 receptor signaling in intestinal epithelial cells**. *Nature Immunology* 2014, **15**:676.

65.   Kolios G, Valatas V, Ward SG: **Nitric oxide in inflammatory bowel disease: a universal messenger in an unsolved puzzle**. *Immunology* 2004, **113**(4):427-437.

66.   Witthoft T, Eckmann L, Kim JM, Kagnoff MF: **Enteroinvasive bacteria directly activate expression of iNOS and NO production in human colon epithelial cells**. *The American journal of physiology* 1998, **275**(3 Pt 1):G564-571.

67.   Fang FC: **Perspectives series: host/pathogen interactions. Mechanisms of nitric oxide-related antimicrobial activity**. *The Journal of clinical investigation* 1997, **99**(12):2818-2825.

68.   Middleton SJ, Shorthouse M, Hunter JO: **Increased nitric oxide synthesis in ulcerative colitis**. *The Lancet* 1993, **341**(8843):465-466.

69.   Lundberg JON, Lundberg JM, Alving K, Hellström PM: **Greatly increased luminal nitric oxide in ulcerative colitis**. *The Lancet* 1994, **344**(8938):1673-1674.

70.   Godkin AJ, Belder AJD, Villa L, Wong A, Beesley JE, Kane SP, Martin JF: **Expression of nitric oxide synthase in ulcerative colitis**. *European Journal of Clinical Investigation* 1996, **26**(10):867-872.

71.   Rachmilewitz D, Stamler JS, Bachwich D, Karmeli F, Ackerman Z, Podolsky DK: **Enhanced colonic nitric oxide generation and nitric oxide synthase activity in ulcerative colitis and Crohn's disease**. *Gut* 1995, **36**(5):718-723.

72. Leonard N, Bishop AE, Polak JM, Talbot IC: **Expression of nitric oxide synthase in inflammatory bowel disease is not affected by corticosteroid treatment**. *Journal of Clinical Pathology* 1998, **51**(10):750-753.

73. Moya-García AA, Pino-Angeles A, Gil-Redondo R, Morreale A, Sánchez-Jiménez F: **Structural features of mammalian histidine decarboxylase reveal the basis for specific inhibition**. *British journal of pharmacology* 2009, **157**(1):4-13.

74. He S-H: **Key role of mast cells and their major secretory products in inflammatory bowel disease**. *World journal of gastroenterology* 2004, **10**(3):309-318.

75. Jorgensen EA, Knigge U, Warberg J, Kjaer A: **Histamine and the regulation of body weight**. *Neuroendocrinology* 2007, **86**(3):210-214.

76. Haas HL, Sergeeva OA, Selbach O: **Histamine in the nervous system**. *Physiological reviews* 2008, **88**(3):1183-1241.

77. Ohtsu H: **Progress in allergy signal research on mast cells: the role of histamine in immunological and cardiovascular disease and the transporting system of histamine in the cell**. *Journal of pharmacological sciences* 2008, **106**(3):347-353.

78. Schubert ML, Peura DA: **Control of gastric acid secretion in health and disease**. *Gastroenterology* 2008, **134**(7):1842-1860.

79. García-Martin E, Mendoza JL, Martínez C, Taxonera C, Urcelay E, Ladero JM, de la Concha EG, Díaz-Rubio M, Agúndez JAG: **Severity of ulcerative colitis is associated with a polymorphism at diamine oxidase gene but not at histamine N-methyltransferase gene**. *World journal of gastroenterology* 2006, **12**(4):615-620.

80. Baenkler HW, Lux G, Gunthner R, Kohlhaufl M, Matek W: **Biopsy histamine in ulcerative colitis and Crohn's disease**. *Hepato-gastroenterology* 1987, **34**(6):289-290.

81. Raithel M, Matek M, Baenkler HW, Jorde W, Hahn EG: **Mucosal Histamine Content and Histamine Secretion in Crohn's Disease, Ulcerative Colitis and Allergic Enteropathy**. *International Archives of Allergy and Immunology* 1995, **108**(2):127-133.

82. Nolte H, Spjeldnaes N, Kruse A, Windelborg B: **Histamine release from gut mast cells from patients with inflammatory bowel diseases**. *Gut* 1990, **31**(7):791.

83. Dubuquoy L, Jansson EÅ, Deeb S, Rakotobe S, Karoui M, Colombel J-F, Auwerx J, Pettersson S, Desreumaux P: **Impaired expression of peroxisome proliferator-activated receptor γ in ulcerative colitis**. *Gastroenterology* 2003, **124**(5):1265-1276.

84. Tyagi S, Gupta P, Saini AS, Kaushal C, Sharma S: **The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases**. *Journal of advanced pharmaceutical technology & research* 2011, **2**(4):236-240.

85. Ye J: **Regulation of PPARgamma function by TNF-alpha**. *Biochemical and biophysical research communications* 2008, **374**(3):405-408.

86. Schoonjans K, Martin G, Staels B, Auwerx J: **Peroxisome proliterator-activated receptors, orphans with ligands and functions**. *Current Opinion in Lipidology* 1997, **8**(3).

87. Ricote M, Li AC, Willson TM, Kelly CJ, Glass CK: **The peroxisome proliferator-activated receptor-γ is a negative regulator of macrophage activation**. *Nature* 1998, **391**:79.

88. Jiang C, Ting AT, Seed B: **PPAR-γ agonists inhibit production of monocyte inflammatory cytokines**. *Nature* 1998, **391**:82.

89. Desreumaux P, Dubuquoy L, Nutten S, Peuchmaur M, Englaro W, Schoonjans K, Derijard B, Desvergne B, Wahli W, Chambon P *et al*: **Attenuation of Colon Inflammation through Activators of the Retinoid X Receptor (Rxr)/Peroxisome Proliferator–Activated Receptor γ (Pparγ) Heterodimer**. *A Basis for New Therapeutic Strategies* 2001, **193**(7):827-838.

90.   Su CG, Wen X, Bailey ST, Jiang W, Rangwala SM, Keilbaugh SA, Flanigan A, Murthy S, Lazar MA, Wu GD: **A novel therapy for colitis utilizing PPAR-γ ligands to inhibit the epithelial inflammatory response**. *The Journal of Clinical Investigation* 1999, **104**(4):383-389.

91.   Roediger WEW: **The colonic epithelium in Ulcerative-Colitis - An Energy-Deficiency Disease**. *Lancet* 1980, **2**.

92.   Cushing K, Alvarado DM, Ciorba MA: **Butyrate and Mucosal Inflammation: New Scientific Evidence Supports Clinical Observation**. *Clinical And Translational Gastroenterology* 2015, **6**:e108.

93.   Shimazu T, Hirschey MD, Newman J, He W, Shirakawa K, Le Moan N, Grueter CA, Lim H, Saunders LR, Stevens RD *et al*: **Suppression of Oxidative Stress by β-Hydroxybutyrate, an Endogenous Histone Deacetylase Inhibitor**. *Science* 2013, **339**(6116):211-214.

94.   Waeytens A, De Vos M, Laukens D: **Evidence for a Potential Role of Metallothioneins in Inflammatory Bowel Diseases**. *Mediators of Inflammation* 2009, **2009**.

95.   Karin M, Eddy RL, Henry WM, Haley LL, Byers MG, Shows TB: **Human metallothionein genes are clustered on chromosome 16**. *Proceedings of the National Academy of Sciences* 1984, **81**(17):5494.

96.   Koelink PJ, Overbeek SA, Braber S, Morgan ME, Henricks PAJ, Roda MA, Verspaget HW, Wolfkamp SC, te Velde AA, Jones CW *et al*: **Collagen degradation and neutrophilic infiltration: a vicious circle in inflammatory bowel disease**. 2014, **63**(4):578-587.

97.   Giovanni RF, Renzo L, Giuseppe C, Adriana F, Sambuy MY: **Carbonic Anhydrase I Reduction in Colonic Mucosa of Patients with Active Ulcerative Colitis**. *Digestive Diseases and Sciences*, **43**(9):2086-2092.

98.   Cao SS: **Epithelial ER Stress in Crohn's Disease and Ulcerative Colitis**. *Inflamm Bowel Dis* 2016, **22**(4):984-993.

99.   Novak EA, Mollen KP: **Mitochondrial dysfunction in inflammatory bowel disease**. *Frontiers in Cell and Developmental Biology* 2015, **3**(62).

100.   Langmann T, Moehle C, Mauerer R, Scharl M, Liebisch G, Zahn A, Stremmel W, Schmitz G: **Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes**. *Gastroenterology* 2004, **127**(1):26-40.

101.   Jahnel J, Fickert P, Hauer AC, Hogenauer C, Avian A, Trauner M: **Inflammatory bowel disease alters intestinal bile acid transporter expression**. *Drug metabolism and disposition: the biological fate of chemicals* 2014, **42**(9):1423-1431.

102.   Penttila O, Kyosola K, Klinge E, Ahonen A, Tallqvist G: **Studies of rectal mucosal catecholamines in ulcerative colitis**. *Annals of clinical research* 1975, **7**(1):32-36.

103.   Stegk JP, Ebert B, Martin HJ, Maser E: **Expression profiles of human 11β-hydroxysteroid dehydrogenases type 1 and type 2 in inflammatory bowel diseases**. *Molecular and Cellular Endocrinology* 2009, **301**(1–2):104-108.

104.   Takahashi K-I, Fukushima K, Sasano H, Sasaki I, Matsuno S, Krozowski ZS, Nagura H: **Type II 11β-Hydroxysteroid Dehydrogenase Expression in Human Colonic Epithelial Cells of Inflammatory Bowel Disease**. *Digestive Diseases and Sciences* 1999, **44**(12):2516-2522.

105.   Maaser C, Magro F, Mantzaris G, Lindsay JO, Colombel J-F, Geboes K, Eliakim R, Vermeire S, Travis S, Reinisch W *et al*: **Second European evidence-based consensus on the diagnosis and management of ulcerative colitis Part 1: Definitions and diagnosis**. *Journal of Crohn's and Colitis* 2012, **6**(10):965-990.

106.   Vavricka SR, Rogler G, Gantenbein C, Spoerri M, Prinz Vavricka M, Navarini AA, French LE, Safroneeva E, Fournier N, Straumann A *et al*: **Chronological Order of Appearance of Extraintestinal Manifestations Relative to the Time of IBD Diagnosis in the Swiss Inflammatory Bowel Disease Cohort**. *Inflamm Bowel Dis* 2015, **21**(8):1794-1800.

107.   Vavricka SR, Brun L, Ballabeni P, Pittet V, Prinz Vavricka BM, Zeitz J, Rogler G, Schoepfer AM: **Frequency and Risk Factors for Extraintestinal Manifestations in the Swiss Inflammatory Bowel Disease Cohort**. *The American Journal Of Gastroenterology* 2010, **106**:110.

108.   Satsangi J, Silverberg MS, Vermeire S, Colombel JF: **The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications**. *Gut* 2006, **55**(6):749-753.

109.   Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, Caprilli R, Colombel JF, Gasche C, Geboes K *et al*: **Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology**. *Canadian journal of gastroenterology = Journal canadien de gastroenterologie* 2005, **19 Suppl A**:5a-36a.

110.   Schroeder KW, Tremaine WJ, Ilstrup DM: **Coated Oral 5-Aminosalicylic Acid Therapy for Mildly to Moderately Active Ulcerative Colitis**. 1987, **317**(26):1625-1629.

111.   Lichtiger S, Present DH, Kornbluth A, Gelernt I, Bauer J, Galler G, Michelassi F, Hanauer S: **Cyclosporine in Severe Ulcerative Colitis Refractory to Steroid Therapy**. 1994, **330**(26):1841-1845.

112.   Walmsley RS, Ayres RCS, Pounder RE, Allan RN: **A simple clinical colitis activity index**. *Gut* 1998, **43**(1):29.

113.   Travis SPL, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel J-F, Feagan BG, Hanauer SB, Lémann M, Lichtenstein GR *et al*: **Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS)**. *Gut* 2012, **61**(4):535.

114.   Sutherland LR, Martin F, Greer S, Robinson M, Greenberger N, Saibil F, Martin T, Sparr J, Prokipchuk E, Borgen L: **5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis**. *Gastroenterology* 1987, **92**(6):1894-1898.

115.   Geboes K, Riddell R, Öst A, Jensfelt B, Persson T, Löfberg R: **A reproducible grading scale for histological assessment of inflammation in ulcerative colitis**. *Gut* 2000, **47**.

116.   Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, Bastien C, Cahn V, Cadiot G, Diebold M-D, Danese S, Reinisch W, Schreiber S *et al*: **Development and validation of the Nancy histological index for UC**. *Gut* 2017, **66**(1):43.

117.   Mosli MH, Feagan BG, Zou G, Sandborn WJ, Haens G, Khanna R, Shackelton LM, Walker CW, Nelson S, Vandervoort MK *et al*: **Development and validation of a histological index for UC**. *Gut* 2017, **66**(1):50.

118.   Riley SA, Mani V, Goodman MJ, Dutt S, Herd ME: **Microscopic activity in ulcerative colitis: what does it mean?** *Gut* 1991, **32**(2):174-178.

119.   Ananthakrishnan AN, McGinley EL, Binion DG: **Excess hospitalisation burden associated with Clostridium difficile in patients with inflammatory bowel disease**. *Gut* 2008, **57**(2):205-210.

120.   Nguyen GC, Kaplan GG, Harris ML, Brant SR: **A national survey of the prevalence and impact of Clostridium difficile infection among hospitalized inflammatory bowel disease patients**. *Am J Gastroenterol* 2008, **103**(6):1443-1450.

121. Jharap B, Sandborn WJ, Reinisch W, D'Haens G, Robinson AM, Wang W, Huang B, Lazar A, Thakkar RB, Colombel JF: **Randomised clinical study: discrepancies between patient-reported outcomes and endoscopic appearance in moderate to severe ulcerative colitis**. *Aliment Pharmacol Ther* 2015, **42**(9):1082-1092.

122. Falvey JD, Hoskin T, Meijer B, Ashcroft A, Walmsley R, Day AS, Gearry RB: **Disease activity assessment in IBD: clinical indices and biomarkers fail to predict endoscopic remission**. *Inflamm Bowel Dis* 2015, **21**(4):824-831.

123. Regueiro M, Rodemann J, Kip KE, Saul M, Swoger J, Baidoo L, Schwartz M, Barrie A, Binion D: **Physician assessment of ulcerative colitis activity correlates poorly with endoscopic disease activity**. *Inflamm Bowel Dis* 2011, **17**(4):1008-1014.

124. Shah SC, Colombel JF, Sands BE, Narula N: **Mucosal Healing Is Associated With Improved Long-term Outcomes of Patients With Ulcerative Colitis: A Systematic Review and Meta-analysis**. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2016, **14**(9):1245-1255.e1248.

125. Ye B, van Langenberg DR: **Mesalazine preparations for the treatment of ulcerative colitis: Are all created equal?** *World journal of gastrointestinal pharmacology and therapeutics* 2015, **6**(4):137-144.

126. Ligumsky M, Karmeli F, Sharon P, Zor U, Cohen F, Rachmilewitz D: **Enhanced thromboxane A2 and prostacyclin production by cultured rectal mucosa in ulcerative colitis and its inhibition by steroids and sulfasalazine**. *Gastroenterology* 1981, **81**(3):444-449.

127. Hawkey CJ, Boughton-Smith NK, Whittle BJ: **Modulation of human colonic arachidonic acid metabolism by sulfasalazine**. *Dig Dis Sci* 1985, **30**(12):1161-1165.

128. Rousseaux C, Lefebvre B, Dubuquoy L, Lefebvre P, Romano O, Auwerx J, Metzger D, Wahli W, Desvergne B, Naccari GC *et al*: **Intestinal antiinflammatory effect of 5-aminosalicylic acid is dependent on peroxisome proliferator-activated receptor-gamma**. *The Journal of experimental medicine* 2005, **201**(8):1205-1215.

129. Ahnfelt-Ronne I, Nielsen OH, Christensen A, Langholz E, Binder V, Riis P: **Clinical evidence supporting the radical scavenger mechanism of 5-aminosalicylic acid**. *Gastroenterology* 1990, **98**(5 Pt 1):1162-1169.

130. Williams C, Panaccione R, Ghosh S, Rioux K: **Optimizing clinical use of mesalazine (5-aminosalicylic acid) in inflammatory bowel disease**. *Therapeutic advances in gastroenterology* 2011, **4**(4):237-248.

131. Iacucci M, de Silva S, Ghosh S: **Mesalazine in inflammatory bowel disease: a trendy topic once again?** *Canadian journal of gastroenterology = Journal canadien de gastroenterologie* 2010, **24**(2):127-133.

132. Dignass A, Lindsay JO, Sturm A, Windsor A, Colombel JF, Allez M, D'Haens G, D'Hoore A, Mantzaris G, Novacek G *et al*: **Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 2: current management**. *Journal of Crohn's & colitis* 2012, **6**(10):991-1030.

133. Truelove SC, Jewell DP: **Intensive intravenous regimen for severe attacks of ulcerative colitis**. *Lancet* 1974, **1**(7866):1067-1070.

134. Sever R, Glass CK: **Signaling by nuclear receptors**. *Cold Spring Harbor perspectives in biology*, **5**(3):a016709-a016709.

135. Kadmiel M, Cidlowski JA: **Glucocorticoid receptor signaling in health and disease**. *Trends in pharmacological sciences* 2013, **34**(9):518-530.

136. Kassel O, Herrlich P: **Crosstalk between the glucocorticoid receptor and other transcription factors: Molecular aspects**. *Molecular and Cellular Endocrinology* 2007, **275**(1):13-29.

137. King EM, Chivers JE, Rider CF, Minnich A, Giembycz MA, Newton R: **Glucocorticoid repression of inflammatory gene expression shows differential responsiveness by transactivation- and transrepression-dependent mechanisms**. *PloS one* 2013, **8**(1):e53936-e53936.

138. Ratman D, Vanden Berghe W, Dejager L, Libert C, Tavernier J, Beck IM, De Bosscher K: **How glucocorticoid receptors modulate the activity of other transcription factors: A scope beyond tethering**. *Molecular and Cellular Endocrinology* 2013, **380**(1):41-54.

139. Song I-H, Buttgereit F: **Non-genomic glucocorticoid effects to provide the basis for new drug developments**. *Molecular and Cellular Endocrinology* 2006, **246**(1):142-146.

140. McKay L, Cidlowski J: **Physiologic and Pharmacologic Effects of Corticosteroids**. In: *Holland-Frei Cancer Medicine.* Edited by Kufe D, Pollock R, Weichselbaum R, 6th edn. Hamilton (ON): BC Decker; 2003.

141. Becker DE: **Basic and clinical pharmacology of glucocorticosteroids**. *Anesthesia progress* 2013, **60**(1):25-32.

142. Alan IS, Alan B: **Side Effects of Glucocorticoids**. In: *Pharmacokinetics and Adverse Effects of Drugs.* Edited by Malangu N: IntechOpen; 2017.

143. González-Lama Y, Gisbert JP: **Monitoring thiopurine metabolites in inflammatory bowel disease**. *Frontline gastroenterology* 2016, **7**(4):301-307.

144. Meijer B, de Boer NKH, van Bodegraven AA, Lopez A, Peyrin-Biroulet L, Jharap B, Sanderson JD, Atreya I, Neurath MF, Barclay ML *et al*: **Thiopurines in Inflammatory Bowel Disease: New Findings and Perspectives**. *Journal of Crohn's and Colitis* 2017, **12**(5):610-620.

145. Rawla P, Sunkara T, Raj JP: **Role of biologics and biosimilars in inflammatory bowel disease: current trends and future perspectives**. *Journal of inflammation research* 2018, **11**:215-226.

146. Billmeier U, Dieterich W, Neurath MF, Atreya R: **Molecular mechanism of action of anti-tumor necrosis factor antibodies in inflammatory bowel diseases**. *World journal of gastroenterology* 2016, **22**(42):9300-9313.

147. Pallai A, Kiss B, Vereb G, Armaka M, Kollias G, Szekanecz Z, Szondy Z: **Transmembrane TNF-alpha Reverse Signaling Inhibits Lipopolysaccharide-Induced Proinflammatory Cytokine Formation in Macrophages by Inducing TGF-beta: Therapeutic Implications**. *J Immunol* 2016, **196**(3):1146-1157.

148. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel J-F, Sandborn WJ, Van Assche G, Axler J, Kim H-J, Danese S *et al*: **Vedolizumab as Induction and Maintenance Therapy for Ulcerative Colitis**. *New England Journal of Medicine* 2013, **369**(8):699-710.

149. Briskin M, Winsor-Hines D, Shyjan A, Cochran N, Bloom S, Wilson J, McEvoy LM, Butcher EC, Kassam N, Mackay CR *et al*: **Human mucosal addressin cell adhesion molecule-1 is preferentially expressed in intestinal tract and associated lymphoid tissue**. *The American journal of pathology* 1997, **151**(1):97-110.

150. Yu H, MacIsaac D, Wong JJ, Sellers ZM, Wren AA, Bensen R, Kin C, Park KT: **Market share and costs of biologic therapies for inflammatory bowel disease in the USA**. *Alimentary pharmacology & therapeutics* 2018, **47**(3):364-370.

151. Sofo L, Caprino P, Sacchetti F, Bossola M: **Restorative proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis: A narrative review**. *World journal of gastrointestinal surgery* 2016, **8**(8):556-563.

152. Biondi A, Zoccali M, Costa S, Troci A, Contessini-Avesani E, Fichera A: **Surgical treatment of ulcerative colitis in the biologic therapy era**. *World J Gastroenterol* 2012, **18**(16):1861-1870.

153. Feuerstein JD, Cheifetz AS: **Ulcerative colitis: epidemiology, diagnosis, and management**. *Mayo Clinic proceedings* 2014, **89**(11):1553-1563.

154. Parks AG, Nicholls RJ: **Proctocolectomy without ileostomy for ulcerative colitis**. *British medical journal* 1978, **2**(6130):85-88.

155. Williet N, Sandborn WJ, Peyrin–Biroulet L: **Patient-Reported Outcomes as Primary End Points in Clinical Trials of Inflammatory Bowel Disease**. *Clinical Gastroenterology and Hepatology* 2014, **12**(8):1246-1256.e1246.

156. Henriksen M, Jahnsen J, Lygren I, Sauar J, Kjellevold O, Schulz T, Vatn MH, Moum B: **Ulcerative colitis and clinical course: results of a 5-year population-based follow-up study (the IBSEN study)**. *Inflammatory bowel diseases* 2006, **12**(7):543-550.

157. Danese S, Siegel CA, Peyrin-Biroulet L: **Review article: integrating budesonide-MMX into treatment algorithms for mild-to-moderate ulcerative colitis**. *Alimentary pharmacology & therapeutics* 2014, **39**(10):1095-1103.

158. D'Haens GR: **Top-down therapy for IBD: rationale and requisite evidence**. *Nature reviews Gastroenterology & hepatology* 2010, **7**(2):86-92.

159. Siegel CA, Horton H, Siegel LS, Thompson KD, Mackenzie T, Stewart SK, Rice PW, Stempak JM, Dezfoli S, Haritunians T *et al*: **A validated web-based tool to display individualised Crohn's disease predicted outcomes based on clinical, serologic and genetic variables**. *Aliment Pharmacol Ther* 2016, **43**(2):262-271.

160. Siegel CA: **Refocusing IBD Patient Management: Personalized, Proactive, and Patient-Centered Care**. 2018, **113**(10):1440-1443.

161. Olsen T, Goll R, Cui G, Christiansen I, Florholmen J: **TNF-alpha gene expression in colorectal mucosa as a predictor of remission after induction therapy with infliximab in ulcerative colitis**. *Cytokine* 2009, **46**(2):222-227.

162. Rismo R, Olsen T, Cui G, Paulssen EJ, Christiansen I, Johnsen K, Florholmen J, Goll R: **Normalization of mucosal cytokine gene expression levels predicts long-term remission after discontinuation of anti-TNF therapy in Crohn's disease**. *Scand J Gastroenterol* 2013, **48**(3):311-319.

163. Vogel C, Marcotte EM: **Insights into the regulation of protein abundance from proteomic and transcriptomic analyses**. *Nature Reviews Genetics* 2012, **13**:227.

164. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast**. *Molecular and cellular biology* 1999, **19**(3):1720-1730.

165. Maier T, Güell M, Serrano L: **Correlation of mRNA and protein in complex biological samples**. *FEBS Letters* 2009, **583**(24):3966-3973.

166. Kramer MC, Gregory BD: **Does RNA secondary structure drive translation or vice versa?** *Nature Structural & Molecular Biology* 2018, **25**(8):641-643.

167. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E *et al*: **Landscape and variation of RNA secondary structure across the human transcriptome**. *Nature* 2014, **505**(7485):706-709.

168. Quax TEF, Claassens NJ, Söll D, van der Oost J: **Codon Bias as a Means to Fine-Tune Gene Expression**. *Molecular cell* 2015, **59**(2):149-161.

169. Pais H, Moxon S, Dalmay T, Moulton V: **Small RNA discovery and characterisation in eukaryotes using high-throughput approaches**. *Advances in experimental medicine and biology* 2011, **722**:239-254.

170. Hannon GJ: **RNA interference**. *Nature* 2002, **418**:244.

171. Kimball SR: **Eukaryotic initiation factor eIF2**. *Int J Biochem Cell Biol* 1999, **31**(1):25-29.

172. Oh CK, Filler SG, Cho SH: **Eukaryotic Translation Initiation Factor-6 Enhances Histamine and IL-2 Production in Mast Cells**. *The Journal of Immunology* 2001, **166**(5):3606.

173. Pain VM: **Initiation of protein synthesis in eukaryotic cells**. *Eur J Biochem* 1996, **236**(3):747-771.

174. Hizli AA, Chi Y, Swanger J, Carter JH, Liao Y, Welcker M, Ryazanov AG, Clurman BE: **Phosphorylation of eukaryotic elongation factor 2 (eEF2) by cyclin A-cyclin-dependent kinase 2 regulates its inhibition by eEF2 kinase**. *Molecular and cellular biology* 2013, **33**(3):596-604.

175. Schueren F, Lingner T, George R, Hofhuis J, Dickel C, Gartner J, Thoms S: **Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals**. *eLife* 2014, **3**:e03640.

176. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS: **Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling**. *Science* 2009, **324**(5924):218.

177. Varshavsky A: **The N-end rule**. *Cell* 1992, **69**(5):725-735.

178. Leach MD, Brown AJP: **Posttranslational Modifications of Proteins in the Pathobiology of Medically Relevant Fungi**. *Eukaryotic Cell* 2012, **11**(2):98.

179. Graves PR, Haystead TAJ: **Molecular Biologist's Guide to Proteomics**. 2002, **66**(1):39-63.

180. Scheele GA: **Two-dimensional gel analysis of soluble proteins. Charaterization of guinea pig exocrine pancreatic proteins**. 1975, **250**(14):5375-5385.

181. O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins**. 1975, **250**(10):4007-4021.

182. Klose J: **Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals**. *Humangenetik* 1975, **26**(3):231-243.

183. Edman P: **A method for the determination of amino acid sequence in peptides**. *Archives of biochemistry* 1949, **22**(3):475.

184. Aebersold RH, Teplow DB, Hood LE, Kent SB: **Electroblotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis**. 1986, **261**(9):4229-4238.

185. Zhu K, Min C, Xia H, Yang Y, Wang B, Chen K: **Characterisation of Indica Special Protein (ISP), a Marker Protein for the Differentiation of Oryza sativa Subspecies indica and japonica**. 2014, **15**(5):7332.

186. Feist P, Hummon AB: **Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples**. 2015, **16**(2):3537.

187. Switzar L, Nicolardi S, Rutten JW, Oberstein SAJL, Aartsma-Rus A, van der Burgt YEM: **In-Depth Characterization of Protein Disulfide Bonds by Online Liquid Chromatography-Electrochemistry-Mass Spectrometry**. *Journal of The American Society for Mass Spectrometry* 2016, **27**(1):50-58.

188. Lindahl M, Mata-Cabana A, Kieselbach T: **The Disulfide Proteome and Other Reactive Cysteine Proteomes: Analysis and Functional Significance**. 2011, **14**(12):2581-2642.

189. Yano H, Kuroda S, Buchanan BB: **Disulfide proteome in the analysis of protein function and structure**. 2002, **2**(9):1090-1096.

190. Boja ES, Fales HM: **Overalkylation of a Protein Digest with Iodoacetamide**. *Analytical Chemistry* 2001, **73**(15):3576-3582.

191. Rebecchi KR, Go EP, Xu L, Woodin CL, Mure M, Desaire H: **A general protease digestion procedure for optimal protein sequence coverage and PTM analysis of recombinant glycoproteins: Application to the characterization of hLOXL2 glycosylation**. *Analytical Chemistry* 2011, **83**(22):8484-8491.

192. Rey M, Mrázek H, Pompach P, Novák P, Pelosi L, Brandolin G, Forest E, Havlíček V, Man P: **Effective Removal of Nonionic Detergents in Protein Mass Spectrometry, Hydrogen/Deuterium Exchange, and Proteomics**. *Analytical Chemistry* 2010, **82**(12):5107-5116.

193. Wisniewski JR, Zougman A, Nagaraj N, Mann M: **Universal sample preparation method for proteome analysis**. *Nat Meth* 2009, **6**(5):359-362.

194. Santa C, Anjo SI, Manadas B: **Protein precipitation of diluted samples in SDS-containing buffer with acetone leads to higher protein recovery and reproducibility in comparison with TCA/acetone approach**. 2016, **16**(13):1847-1851.

195. Reynolds JA, Tanford C: **Binding of Dodecyl Sulfate to Proteins at High Binding Ratios. Possible Implications for the State of Proteins in Biological Membranes**. 1970, **66**(3):1002-1007.

196. Zangi R, Zhou R, Berne BJ: **Urea's Action on Hydrophobic Interactions**. *Journal of the American Chemical Society* 2009, **131**(4):1535-1541.

197. Rezus YLA, Bakker HJ: **Effect of urea on the structural dynamics of water**. 2006, **103**(49):18417-18420.

198. Stumpe MC, Grubmüller H: **Aqueous Urea Solutions: Structure, Energetics, and Urea Aggregation**. *The Journal of Physical Chemistry B* 2007, **111**(22):6220-6228.

199. Tsiatsiani L, Heck AJR: **Proteomics beyond trypsin**. 2015, **282**(14):2612-2626.

200. Rypniewski WR, Perrakis A, Vorgias CE, Wilson KS: **Evolutionary divergence and conservation of trypsin**. *Protein Engineering, Design and Selection* 1994, **7**(1):57-64.

201. Muhlia-Almazán A, Sánchez-Paz A, García-Carreño FLJJoCPB: **Invertebrate trypsins: a review**. 2008, **178**(6):655-672.

202. Rodriguez J, Gupta N, Smith RD, Pevzner PA: **Does Trypsin Cut Before Proline?** *Journal of Proteome Research* 2008, **7**(1):300-305.

203. Steen H, Mann M: **The ABC's (and XYZ's) of peptide sequencing**. *Nature reviews Molecular cell biology* 2004, **5**(9):699-711.

204. Saveliev S, Bratz M, Zubarev R, Szapacs M, Budamgunta H, Urh M: **Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis**. *Nature Methods* 2013, **10**:1134.

205. Wada Y, Kadoya M: **In-gel digestion with endoproteinase LysC**. *Journal of mass spectrometry : JMS* 2003, **38**(1):117-118.

206. Glatter T, Ludwig C, Ahrné E, Aebersold R, Heck AJR, Schmidt A: **Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion**. *Journal of Proteome Research* 2012, **11**(11):5145-5156.

207.   Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J: **The Perseus computational platform for comprehensive analysis of (prote)omics data**. *Nat Meth* 2016, **13**(9):731-740.

208.   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences* 2005, **102**(43):15545-15550.

209.   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome research* 2003, **13**(11):2498-2504.

210.   Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J: **ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks**. *Bioinformatics (Oxford, England)* 2009, **25**(8):1091-1093.

211.   Rifai N, Gillette MA, Carr SA: **Protein biomarker discovery and validation: the long and uncertain path to clinical utility**. *Nature Biotechnology* 2006, **24**:971.

212.   Chaparro M, Gisbert JP: **Clinical Usefulness of Proteomics in Inflammatory Bowel Disease: A Comprehensive Review**. 2018.

213.   Sands BE: **Biomarkers of Inflammation in Inflammatory Bowel Disease**. *Gastroenterology* 2015, **149**(5):1275-1285.e1272.

214.   Menees SB, Powell C, Kurlander J, Goel A, Chey WD: **A meta-analysis of the utility of C-reactive protein, erythrocyte sedimentation rate, fecal calprotectin, and fecal lactoferrin to exclude inflammatory bowel disease in adults with IBS**. *Am J Gastroenterol* 2015, **110**(3):444-454.

215.   Schoepfer AM, Pittet V, Vavricka SR, Straumann A, Beglinger C, Safroneeva E, Romero Y, Armstrong D, Schmidt C, Trummler M: **Fecal Calprotectin More Accurately Reflects Endoscopic Activity of Ulcerative Colitis than the Lichtiger Index, C-reactive Protein, Platelets, Hemoglobin, and Blood Leukocytes**. *Inflamm Bowel Dis* 2013, **19**(2):332-341.

216.   Mosli MH, Zou G, Garg SK, Feagan SG, MacDonald JK, Chande N, Sandborn WJ, Feagan BG: **C-Reactive Protein, Fecal Calprotectin, and Stool Lactoferrin for Detection of Endoscopic Activity in Symptomatic Inflammatory Bowel Disease Patients: A Systematic Review and Meta-Analysis**. *The American Journal Of Gastroenterology* 2015, **110**:802.

217.   Ho GT, Mowat C, Goddard CJR, Fennell JM, Shah NB, Prescott RJ, Satsangi J: **Predicting the outcome of severe ulcerative colitis: development of a novel risk score to aid early selection of patients for second-line medical therapy or surgery**. 2004, **19**(10):1079-1087.

218.   Hsieh S-Y, Shih T-C, Yeh C-Y, Lin C-J, Chou Y-Y, Lee Y-S: **Comparative proteomic studies on the pathogenesis of human ulcerative colitis**. *Proteomics* 2006, **6**(19):5322-5331.

219.   Starr AE, Deeke SA, Ning Z, Chiang C-K, Zhang X, Mottawea W, Singleton R, Benchimol EI, Wen M, Mack DR *et al*: **Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate Crohn's disease from UC**. *Gut* 2016.

220.   Chan PP, Wasinger VC, Leong RW: **Current application of proteomics in biomarker discovery for inflammatory bowel disease**. *World journal of gastrointestinal pathophysiology* 2016, **7**(1):27-37.

221.   Borneman J, Ye J, Ruegger PM, Presley LL, Li X, Braun J, LeBlanc J, Zhang Z, Jeske DR, Cui X *et al*: **Host–Microbe Relationships in Inflammatory Bowel Disease Detected by Bacterial**

and Metaproteomic Analysis of the Mucosal–Luminal Interface**. *Inflamm Bowel Dis* 2011, **18**(3):409-417.

222. Meuwis M-A, Fillet M, Geurts P, de Seny D, Lutteri L, Chapelle J-P, Bours V, Wehenkel L, Belaiche J, Malaise M *et al*: **Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling**. *Biochemical Pharmacology* 2007, **73**(9):1422-1433.

223. Nanni P, Parisi D, Roda G, Casale M, Belluzzi A, Roda E, Mayer L, Roda A: **Serum protein profiling in patients with inflammatory bowel diseases using selective solid-phase bulk extraction, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and chemometric data analysis**. *Rapid Commun Mass Spectrom* 2007, **21**(24):4142-4148.

224. Kanmura S, Uto H, Numata M, Hashimoto S, Moriuchi A, Fujita H, Oketani M, Ido A, Kodama M, Ohi H *et al*: **Human neutrophil peptides 1-3 are useful biomarkers in patients with active ulcerative colitis**. *Inflamm Bowel Dis* 2009, **15**(6):909-917.

225. Marko L, Szigeti N, Szabo Z, Boddi K, Takatsy A, Ludany A, Koszegi T, Molnar GA, Wittmann I: **Potential urinary biomarkers of disease activity in Crohn's disease**. *Scand J Gastroenterol* 2010, **45**(12):1440-1448.

226. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, Shah M, Halfvarson J, Tysk C, Henrissat B *et al*: **Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease**. *PloS one* 2012, **7**(11):e49138.

227. Fogt F, Jian B, Krieg RC, Wellmann A: **Proteomic analysis of mucosal preparations from patients with ulcerative colitis**. *Molecular medicine reports* 2008, **1**(1):51-54.

228. Shkoda A, Werner T, Daniel H, Gunckel M, Rogler G, Haller D: **Differential protein expression profile in the intestinal epithelium from patients with inflammatory bowel disease**. *J Proteome Res* 2007, **6**(3):1114-1125.

229. Barceló-Batllori S, André M, Servis C, Lévy N, Takikawa O, Michetti P, Reymond M, Felley-Bosco E: **Proteomic analysis of cytokine induced proteins in human intestinal epithelial cells: Implications for inflammatory bowel diseases**. 2002, **2**(5):551-560.

230. Bennike TB, Carlsen TG, Ellingsen T, Bonderup OK, Glerup H, Bøgsted M, Christiansen G, Birkelund S, Stensballe A, Andersen V: **Neutrophil Extracellular Traps in Ulcerative Colitis: A Proteome Analysis of Intestinal Biopsies**. *Inflamm Bowel Dis* 2015, **21**(9):2052-2067.

231. Bennike T, Birkelund S, Stensballe A, Andersen V: **Biomarkers in inflammatory bowel diseases: Current status and proteomics identification strategies**. *World Journal of Gastroenterology : WJG* 2014, **20**(12):3231-3244.

232. Savitski MM, Mathieson T, Zinn N, Sweetman G, Doce C, Becher I, Pachl F, Kuster B, Bantscheff M: **Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification**. *Journal of Proteome Research* 2013, **12**(8):3586-3598.

233. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification**. *Nat Biotech* 2008, **26**(12):1367-1372.

234. The UniProt Consortium: **UniProt: the universal protein knowledgebase**. *Nucleic Acids Research* 2017, **45**(D1):D158-D169.

235. Tyanova S, Cox J: **Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research**. *Methods in molecular biology (Clifton, NJ)* 2018, **1711**:133-148.

236. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods**. *Biostatistics (Oxford, England)* 2007, **8**(1):118-127.

237. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *Omics : a journal of integrative biology* 2012, **16**(5):284-287.

238. Supek F, Bosnjak M, Skunca N, Smuc T: **REVIGO summarizes and visualizes long lists of gene ontology terms**. *PloS one* 2011, **6**(7):e21800.

239. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nature genetics* 2000, **25**(1):25-29.

240. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 2000, **28**(1):27-30.

241. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences* 2005, **102**(43):15545.

242. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M *et al*: **InterPro in 2017-beyond protein family and domain annotations**. *Nucleic acids research* 2017, **45**(D1):D190-D199.

243. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B *et al*: **The Reactome Pathway Knowledgebase**. *Nucleic acids research* 2018, **46**(D1):D649-D655.

244. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I: **New and continuing developments at PROSITE**. *Nucleic Acids Research* 2013, **41**(D1):D344-D347.

245. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D *et al*: **WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research**. *Nucleic Acids Research* 2018, **46**(D1):D661-D667.

246. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al*: **STRING v10: protein-protein interaction networks, integrated over the tree of life**. *Nucleic Acids Res* 2015, **43**(Database issue):D447-452.

247. Davalos D, Akassoglou K: **Fibrinogen as a key regulator of inflammation in disease**. *Seminars in Immunopathology* 2012, **34**(1):43-62.

248. Prashar A, Schnettger L, Bernard EM, Gutierrez MG: **Rab GTPases in Immunity and Inflammation**. *Frontiers in Cellular and Infection Microbiology* 2017, **7**:435.

249. Ohira M, Oshitani N, Hosomi S, Watanabe K, Yamagami H, Tominaga K, Watanabe T, Fujiwara Y, Maeda K, Hirakawa K *et al*: **Dislocation of Rab13 and vasodilator-stimulated phosphoprotein in inactive colon epithelium in patients with Crohn's disease**. *International journal of molecular medicine* 2009, **24**(6):829-835.

250. Lin Y, Huo L, Liu Z, Li J, Liu Y, He Q, Wang X, Liang S: **Sodium Laurate, a Novel Protease- and Mass Spectrometry-Compatible Detergent for Mass Spectrometry-Based Membrane Proteomics**. *PloS one* 2013, **8**(3):e59779.

251. Pasing Y, Colnoe S, Hansen T: **Proteomics of hydrophobic samples: Fast, robust and low-cost workflows for clinical approaches**. *Proteomics* 2017, **17**(6):1500462-n/a.

252. Lin Y, Zhou J, Bi D, Chen P, Wang X, Liang S: **Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins**. *Analytical Biochemistry* 2008, **377**(2):259-266.

253. Moore SM, Hess SM, Jorgenson JW: **Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics**. *Journal of Proteome Research* 2016, **15**(4):1243-1252.

254. Mócz G, Bálint M: **Use of cationic detergents for polyacrylamide gel electrophoresis in multiphasic buffer systems**. *Analytical Biochemistry* 1984, **143**(2):283-292.

255. León IR, Schwämmle V, Jensen ON, Sprenger RR: **Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis**. *Molecular & Cellular Proteomics* 2013, **12**(10):2992-3005.

256. Cao SS: **Endoplasmic Reticulum Stress and Unfolded Protein Response in Inflammatory Bowel Disease**. *Inflamm Bowel Dis* 2015, **21**(3):636-644.

257. Bicknell AA, Babour A, Federovitch CM, Niwa M: **A novel role in cytokinesis reveals a housekeeping function for the unfolded protein response**. *The Journal of cell biology* 2007, **177**(6):1017-1027.

258. Smith JA: **Regulation of Cytokine Production by the Unfolded Protein Response; Implications for Infection and Autoimmunity**. *Frontiers in immunology* 2018, **9**:422-422.

259. Tieu K, Perier C, Caspersen C, Teismann P, Wu D-C, Yan S-D, Naini A, Vila M, Jackson-Lewis V, Ramasamy R *et al*: **D-β-Hydroxybutyrate rescues mitochondrial respiration and mitigates features of Parkinson disease**. *The Journal of Clinical Investigation*, **112**(6):892-901.

260. Plecko B, Stoeckler-Ipsiroglu S, Schober E, Harrer G, Mlynarik V, Gruber S, Moser E, Moeslinger D, Silgoner H, Ipsiroglu O: **Oral Beta-Hydroxybutyrate Supplementation in Two Patients with Hyperinsulinemic Hypoglycemia: Monitoring of Beta-Hydroxybutyrate Levels in Blood and Cerebrospinal Fluid, and in the Brain by In Vivo Magnetic Resonance Spectroscopy**. *Pediatr Res* 2002, **52**(2):301-306.

261. Coyle P, Philcox JC, Carey LC, Rofe AM: **Metallothionein: the multipurpose protein**. *Cellular and molecular life sciences : CMLS* 2002, **59**(4):627-647.

262. Foster M, Samman S: **Zinc and regulation of inflammatory cytokines: implications for cardiometabolic disease**. *Nutrients* 2012, **4**(7):676-694.

263. Haase H, Ober-Blobaum JL, Engelhardt G, Hebel S, Heit A, Heine H, Rink L: **Zinc signals are essential for lipopolysaccharide-induced signal transduction in monocytes**. *J Immunol* 2008, **181**(9):6491-6502.

264. Prasad AS: **Zinc is an Antioxidant and Anti-Inflammatory Agent: Its Role in Human Health**. *Frontiers in nutrition* 2014, **1**:14-14.

265. Aaseth J, Gerhardsson L, Skaug MA, Alexander J: **Chapter 1 - General Chemistry of Metal Toxicity and Basis for Metal Complexation**. In: *Chelation Therapy in the Treatment of Metal Intoxication.* Edited by Aaseth J, Crisponi G, Andersen O. Boston: Academic Press; 2016: 1-33.

266. Kruidenier L, Kuiper I, Lamers CB, Verspaget HW: **Intestinal oxidative damage in inflammatory bowel disease: semi-quantification, localization, and association with mucosal antioxidants**. *J Pathol* 2003, **201**(1):28-36.

267. Tran CD, Ball JM, Sundar S, Coyle P, Howarth GS: **The role of zinc and metallothionein in the dextran sulfate sodium-induced colitis mouse model**. *Dig Dis Sci* 2007, **52**(9):2113-2121.

268. Oz HS, Chen T, de Villiers WJ, McClain CJ: **Metallothionein overexpression does not protect against inflammatory bowel disease in a murine colitis model**. *Medical science monitor : international medical journal of experimental and clinical research* 2005, **11**(3):Br69-73.

269. Tran CD, Huynh H, van den Berg M, van der Pas M, Campbell MA, Philcox JC, Coyle P, Rofe AM, Butler RN: **Helicobacter-induced gastritis in mice not expressing metallothionein-I and II**. *Helicobacter* 2003, **8**(5):533-541.

270. Takano H, Inoue K, Yanagisawa R, Sato M, Shimada A, Morita T, Sawada M, Nakamura K, Sanbongi C, Yoshikawa T: **Protective role of metallothionein in acute lung injury induced by bacterial endotoxin**. *Thorax* 2004, **59**(12):1057-1062.

271. Youn J, Hwang SH, Ryoo ZY, Lynes MA, Paik DJ, Chung HS, Kim HY: **Metallothionein suppresses collagen-induced arthritis via induction of TGF-beta and down-regulation of proinflammatory mediators**. *Clinical and experimental immunology* 2002, **129**(2):232-239.

272. Kimura T, Itoh N, Takehara M, Oguro I, Ishizaki JI, Nakanishi T, Tanaka K: **Sensitivity of metallothionein-null mice to LPS/D-galactosamine-induced lethality**. *Biochem Biophys Res Commun* 2001, **280**(1):358-362.

273. Waelput W, Broekaert D, Vandekerckhove J, Brouckaert P, Tavernier J, Libert C: **A mediator role for metallothionein in tumor necrosis factor-induced lethal shock**. *J Exp Med* 2001, **194**(11):1617-1624.

274. Chu A, Foster M, Ward S, Zaman K, Hancock D, Petocz P, Samman S: **Zinc-induced upregulation of metallothionein (MT)-2A is predicted by gene expression of zinc transporters in healthy adults**. *Genes & nutrition* 2015, **10**(6):44-44.

275. Wax PM: **Current use of chelation in American health care**. *Journal of medical toxicology : official journal of the American College of Medical Toxicology* 2013, **9**(4):303-307.

276. Siva S, Rubin DT, Gulotta G, Wroblewski K, Pekow J: **Zinc Deficiency is Associated with Poor Clinical Outcomes in Patients with Inflammatory Bowel Disease**. *Inflamm Bowel Dis* 2017, **23**(1):152-157.

277. Moura FA, de Andrade KQ, dos Santos JCF, Araújo ORP, Goulart MOF: **Antioxidant therapy for treatment of inflammatory bowel disease: Does it work?** *Redox Biology* 2015, **6**:617-639.

278. Langenbach R, Loftin C, Lee C, Tiano H: **Cyclooxygenase knockout mice: Models for elucidating isoform-specific functions**. *Biochemical Pharmacology* 1999, **58**(8):1237-1246.

279. Gilroy DW, Colville-Nash PR, Willis D, Chivers J, Paul-Clark MJ, Willoughby DA: **Inducible cyclooxygenase may have anti-inflammatory properties**. *Nature Medicine* 1999, **5**:698.

280. Suzuki-Yamamoto T, Nishizawa M, Fukui M, Okuda-Ashitaka E, Nakajima T, Ito S, Watanabe K: **cDNA cloning, expression and characterization of human prostaglandin F synthase 1**. *FEBS Letters* 1999, **462**(3):335-340.

281. Komoto J, Yamada T, Watanabe K, Woodward DF, Takusagawa F: **Prostaglandin F2α Formation from Prostaglandin H2 by Prostaglandin F Synthase (PGFS): Crystal Structure of PGFS Containing Bimatoprost**. *Biochemistry* 2006, **45**(7):1987-1996.

282. Ricciotti E, FitzGerald GA: **Prostaglandins and inflammation**. *Arteriosclerosis, thrombosis, and vascular biology* 2011, **31**(5):986-1000.

283. Jegerschold C, Pawelzik SC, Purhonen P, Bhakat P, Gheorghe KR, Gyobu N, Mitsuoka K, Morgenstern R, Jakobsson PJ, Hebert H: **Structural basis for induced formation of the inflammatory mediator prostaglandin E2**. *Proc Natl Acad Sci U S A* 2008, **105**(32):11110-11115.

284. Cho H, Huang L, Hamza A, Gao D, Zhan C-G, Tai H-H: **Role of glutamine 148 of human 15-hydroxyprostaglandin dehydrogenase in catalytic oxidation of prostaglandin E2**. *Bioorganic & Medicinal Chemistry* 2006, **14**(19):6486-6491.

285. Vong L, Ferraz JGP, Panaccione R, Beck PL, Wallace JL: **A pro-resolution mediator, prostaglandin D(2), is specifically up-regulated in individuals in long-term remission from ulcerative colitis**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(26):12023-12027.

286. Sandborn WJ: **Step-up versus top-down therapy in the treatment of ulcerative colitis**. *Gastroenterology & hepatology* 2007, **3**(1):16-17.

Paper I

The Proteome of Ulcerative Colitis in Colon Biopsies from Adults - Optimized Sample Preparation and Comparison with Healthy Controls

Armin Schniers, Endre Anderssen, Christopher Graham Fenton, Rasmus Goll, Yvonne Pasing, Ruth Hracky Paulssen, Jon Florholmen, and Terkel Hansen*

Purpose: The purpose of the study was to optimize the sample preparation and to further use an improved sample preparation to identify proteome differences between inflamed ulcerative colitis tissue from untreated adults and healthy controls.

Experimental design: To optimize the sample preparation, we studied the effect of adding different detergents to a urea containing lysis buffer for a Lys-C/trypsin tandem digestion. With the optimized method, we prepared clinical samples from six ulcerative colitis patients and six healthy controls and analysed them by LC-MS/MS. We examined the acquired data to identify differences between the states.

Results: We improved the protein extraction and protein identification number by utilizing a urea and sodium deoxycholate containing buffer. Comparing ulcerative colitis and healthy tissue, we found 168 of 2366 identified proteins differently abundant. Inflammatory proteins are higher abundant in ulcerative colitis, proteins related to anion-transport and mucus production are lower abundant. A high proportion of S100 proteins is differently abundant, notably with both up-regulated and down-regulated proteins.

Conclusion and clinical relevance: The optimized sample preparation method will improve future proteomic studies on colon mucosa. The observed protein abundance changes and their enrichment in various groups improve our understanding of ulcerative colitis on protein level.

A. Schniers, T. Hansen
Natural Products and Medicinal Chemistry Research Group
Department of Pharmacy, Faculty of Health Sciences
UiT The Arctic University of Norway
Tromsø, Norway
E-mail: Terkel.Hansen@uit.no
E. Anderssen, C. G. Fenton, Y. Pasing, R. H. Paulssen
Department of Clinical Medicine, Faculty of Health Sciences
UiT The Arctic University of Norway
Tromsø, Norway
R. Goll, J. Florholmen
Department of Medical Gastroenterology
University Hospital of North Norway
Tromsø, Norway
Y. Pasing
Division of Internal Medicine
University Hospital of North Norway
Tromsø, Norway

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/prca.201700053

1. Introduction

1.1. Inflammatory Bowel Disease

The global incidence of inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), is increasing. In Europe 2.2 million predominantly younger people are suffering from the disease. The immune response is dysregulated in IBD such as in other autoimmune diseases.[1] The cause of the disease is unknown, but it is generally accepted that IBD represents a dysregulation of the intestinal immune tolerance to microbiological and/or nutritional antigens. This dysregulation results from a combination of environmental and genetic factors, in which dysfunctional gene products may initiate inflammation via signalling pathways that have been defined to varying degrees, including innate, adaptive and T-regulatory immunity systems (for a comprehensive review, please see[2]).

1.2. The IBD Immune Response

The immune response in IBD is far from fully characterized. In general, UC is described as an atypical Th2 and CD as a Th1 mediated disease.[3] Much knowledge about IBD is based on gene transcriptional data, whereas there is a lack of data for an extensive and comprehensible description of the proteomic immune response.[4–6] One of the first more complete analyses of the proteome in IBD was reported recently,[7] describing the proteome of CD and UC in children in comparison with each other and with healthy controls. Another study describes the UC proteome of uninflamed colon mucosa from adults under anti-inflammatory treatment.[8] However, previous studies do not provide a view into the proteome of inflamed tissue from untreated adult patients in comparison with healthy controls, which is essential to understand the underlying mechanisms.

The protein abundances are constantly changing with the state of a tissue, in addition to the posttranslational modifications, that can modify a protein's function and activity. Hence, proteomic data can reveal new insight to the pathophysiology of IBD.

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Clinical Applications
www.clinical.proteomics-journal.com

Clinical Relevance

Ulcerative colitis is, along with Crohn's disease, one major form of inflammatory bowel disease (IBD). IBD affects 0.5–1.0% of the population in North America and Europe. The colon and rectum in ulcerative colitis (UC) are in a state of chronic inflammation, which is accompanied by ulceration, abdominal pain, diarrhea, bloody stool, and further gastrointestinal and extraintestinal manifestations. The understanding of the UC pathophysiology is incomplete; the knowledge about proteome changes in the disease contributes to understanding the biochemical processes associated with UC. We identified 168 proteins as differently regulated in inflamed colon biopsies from adult UC patients when compared to samples from healthy controls, and we show which biological functions are mostly affected by these changes. The finding of decreased abundances in UC for the proteins S100A14 and S100A16, in contrast to several other S100 proteins with increased abundances, widens the view on the pathophysiological role of this protein family. The sample preparation method for colon biopsies, which we optimized in this study, gives reproducible results with high protein identification numbers, thus it will enhance the quality of future clinical studies.

1.3. Bottom-Up Proteomics

In-depth shotgun proteomic analyses of clinically relevant samples are currently dominated by bottom-up approaches.[9]

Bottom-up proteomics is based on the mass spectrometric analysis of a peptide solution. Regardless of the following steps, any proteomic sample preparation protocol starts with the extraction of proteins from the tissue. Mechanical force is usually necessary to break up the macroscopic structure of a tissue sample, thereby enabling detergents and chaotropes in the lysis buffer to solubilize the proteins. The solubilized proteins then undergo enzymatic digestion into peptides.

The choice of detergents and chaotropes in the lysis buffer determines the possibilities in the following sample preparation, because both detergents and chaotropes need to be removed prior to mass spectrometric analysis. Some detergent removal methods, such as FASP, must be conducted before enzymatic digestion.

1.4. Sample Preparation Method is Highly Determining for Proteome Depth and Reproducibility

Chaotropes and detergents increase the solubility of proteins, as they promote the disruption of non-covalent bonds. The disruption of these bonds facilitates the detachment of membrane constituents from membrane bound proteins and reduces interactions within and between proteins, which causes them to unfold. However, several chaotropes and detergents need to be diluted or removed prior to enzymatic digestion due to inhibition of trypsin, introducing one or more steps in the sample preparation that can influence the consistency in the results.[10] A feasible method to enable digestion into peptides in high urea and/or detergent con-

centrations without changing the tryptic digestion pattern is a pre-digestion with the endopeptidase Lys-C.[11] As Lys-C is active under harsh solubilisation conditions, the challenges with protein precipitation upon digestion can be reduced or avoided.

The aim of the method development in this study was to optimize the protein extraction and enzymatic digestion for colon mucosa biopsies. Our results achieved with SDC outperform other applied approaches when evaluating reproducibility and numbers of identified peptides and protein groups. Combining the tandem Lys-C/trypsin digestion with sodium deoxycholate is a simple and cost-effective way to enhance the sample preparation.

2. Experimental Section

2.1. Biopsy Collection

The recruited subjects were patients newly diagnosed with moderate to severe UC admitted at the University Hospital of North Norway, not receiving any treatment for IBD. The diagnosis was based on established clinical, endoscopic and histological criteria.[12] Recruited were patients with active disease as described in **Table 1**; one patient with total colitis, three patients with left-side colitis, and two patients with proctitis. The degree of illness was evaluated using the scoring system ulcerative colitis disease activity index (UCDAI, score 0–12), which is based on clinical signs and on endoscopic evaluation of the colon during colonoscopy (grade 0–3).[13]

Samples from subjects performing a cancer screening and with normal colonoscopy and normal colonic histological examination served as controls. All participants gave their informed written consent. The Regional Committee of Medical Ethics of North Norway and the Norwegian Social Science Data Services approved the study and the storage of biological material (REK NORD 2012/1349).

Colonic mucosal biopsies were obtained from the most inflamed area - the rectosigmoid - as typical for UC, and in the controls biopsies were obtained from the rectum. Biopsies (10–15 mg wet weight) were kept frozen at –70 °C for further processing as described below.

2.2. Protein Identification and Quantification

The methods are described in brief here; we provide further details as supporting information.

2.2.1. Sample Preparation for the Method Development

2.2.1.1. Sample Pooling for Method Development and Tissue Lysis. Shortly, 12 colon biopsies (six healthy, six UC) were pooled in water and homogenized in a MagNA Lyzer instrument (Roche) at 6500 rpm for 35 s. Protein concentrations were measured with Pierce BCA Protein Assay Kit following the vendor's protocol. Ten microliter aliquots of the homogenate, containing the tissue debris and 327 μg protein each, were added to

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**Proteomics
Clinical Applications**

www.clinical.proteomics-journal.com

Table 1. Demographics of 6 patients with moderate to severe ulcerative colitis (UC) at diagnosis of disease, and 6 healthy controls.

	Average age (SD)	Female/male (SD)	UCDAI score (SD)	Endoscopic score (SD)	TNF-alpha (SD)
UC patients	48 (14)	1/5	9.5 (2)	2.6 (0.5)	14 188 (12390)
Healthy controls	54 (10)	1/5	–	0	2933 (1084)

UCDAI: ulcerative colitis disease activity index. TNF: tumor necrosis factor (in copies/ μg RNA).

240 μl of various lysis buffers in ceramic bead tubes to produce the following conditions in 3 replicates each: SDC (5% sodium deoxycholate, 8 M urea, 100 mM TEAB), SDS (1% sodium dodecyl sulfate, 8 M urea, 100 mM TEAB), SL (5% sodium laurate, 5.7 M urea, 100 mM TEAB), Rapigest (0.25% Rapigest, 8 M urea, 100 mM TEAB), CTAB (1% cetyltrimethylammonium bromide, 8 M urea, 100 mM TEAB), NMDOA (5% N-methyldioctylamine, 6 M urea, 100 mM TEAB, adjustment to pH 6 with formic acid), Urea (8 M urea, 100 mM TEAB). The urea concentrations with SL and NMDOA were reduced in order to resolve solubility problems. The samples were treated in the MagNA Lyzer instrument at 6500 rpm for 35 s.

2.2.1.2. Reduction and Alkylation.   Disulfide bridges were reduced with 5 mM DTT for 30 min at 54 °C, followed by alkylation of the cysteines with 15 mM iodoacetamide for 30 min at room temperature in the dark. Excess iodoacetamide was quenched by adding DTT corresponding to 5 mM for 5 min at room temperature.[14]

2.2.1.3. Enzymatic Digestion.   The samples were, where applicable, diluted to 6 M urea prior to 8 h Lys-C digestion (Lys-C:protein ratio 1:100) at 37 °C. All samples were diluted to 1 M urea prior to tryptic digestion (trypsin:protein ratio 1:20) for 16 h at 37 °C. The buffers contained 1 mM CaCl$_2$ and 100 mM TEAB during the digestions.[11]

2.2.1.4. Detergent Removal.   CTAB, NMDOA and SDS were removed prior to enzymatic digestion by modified FASP methods.[15] SDC and SL were removed by acid precipitation, and Rapigest by acid induced cleavage, each after enzymatic digestion. A detailed description of the detergent removal methods is provided in the supporting information.

2.2.1.5. Desalting.   The samples were desalted with Omix C18 pipette tips according to the manufacturer's protocol, except for the use of TFA instead of heptafluorobutyric acid. The peptides were eluted with 0.1% FA in 50% acetonitrile.

Samples were dried in an Eppendorf concentrator before being re-dissolved in 10 μL 0.1% TFA.

2.2.2. Sample Preparation for the UC-Healthy Comparison Study

2.2.2.1. Protein Extraction.   Tissue samples from six UC patients and six healthy controls were transferred to ceramic bead tubes with 250 μL lysis buffer (8 M urea, 5% sodium deoxycholate, 0.1 M TEAB) and lysed in a MagNA Lyzer instrument for 35 seconds at 6500 rpm. After dilution as described in supporting information, the protein concentration was determined by BCA

assay according to the manufacturer's protocol. An aliquot of 60 μg protein was taken for the following steps.

2.2.2.2. Reduction and Alkylation.   Cysteines were reduced and alkylated as described in Section 2.2.1.

2.2.2.3. Enzymatic Digestion, Labelling and Sample Mixture.   Predigestion with Lys-C was done for 8 h at 37 °C in 6 M urea, 3.75% sodium deoxycholate, 1 mM CaCl$_2$, and 0.1 M TEAB with 0.6 μg Lys-C (Lys-C:protein ratio 1:100). Tryptic digestion followed after dilution to 1 M urea, 0.625% sodium deoxycholate, 1 mM CaCl$_2$, and 0.1 M TEAB with 3 μg trypsin (trypsin:protein ratio 1:20). Processed samples from six UC patients and six healthy controls were mixed in even peptide amounts to produce a standard. The standard and the samples were labelled with TMT as described in the supporting information.

2.2.2.4. SDC and Acetonitrile Removal.   Formic acid was added to a concentration of 5% and the precipitated sodium deoxycholate was removed by centrifugation.

The samples were dried in an Eppendorf concentrator to remove acetonitrile.

2.2.2.5. Desalting.   Samples were dissolved in 200 μL 0.1 M TEAB. 10.5 μL 50% FA and 22.2 μL 10% TFA were added prior to desalting with C18 pipette tips according to manufacturer's protocol. The samples were dried in an Eppendorf concentrator and redissolved in 8 μL 0.1% TFA.

2.2.3. Nanospray LC-MS/MS

The samples were analysed with a qExactive as described in the Supporting Information.

2.2.3.1. Inclusion Lists for the UC-Healthy Comparison Study. For the quantitative study, two additional LC-MS/MS analyses with inclusion lists were conducted for each sample as described in the Supporting Information.

2.2.4. Data Analysis

The resulting raw-files were analysed with MaxQuant version 1.5.5.1 and Perseus version 1.5.6.0 with a fasta-file containing the human proteins (downloaded 09.09.2016 from uniprot.org, including Swiss-Prot and TrEMBL data) and standard settings, except for peptide spectrum match, protein and site FDRs of 0.05 for the UC-healthy comparison.[16,17]
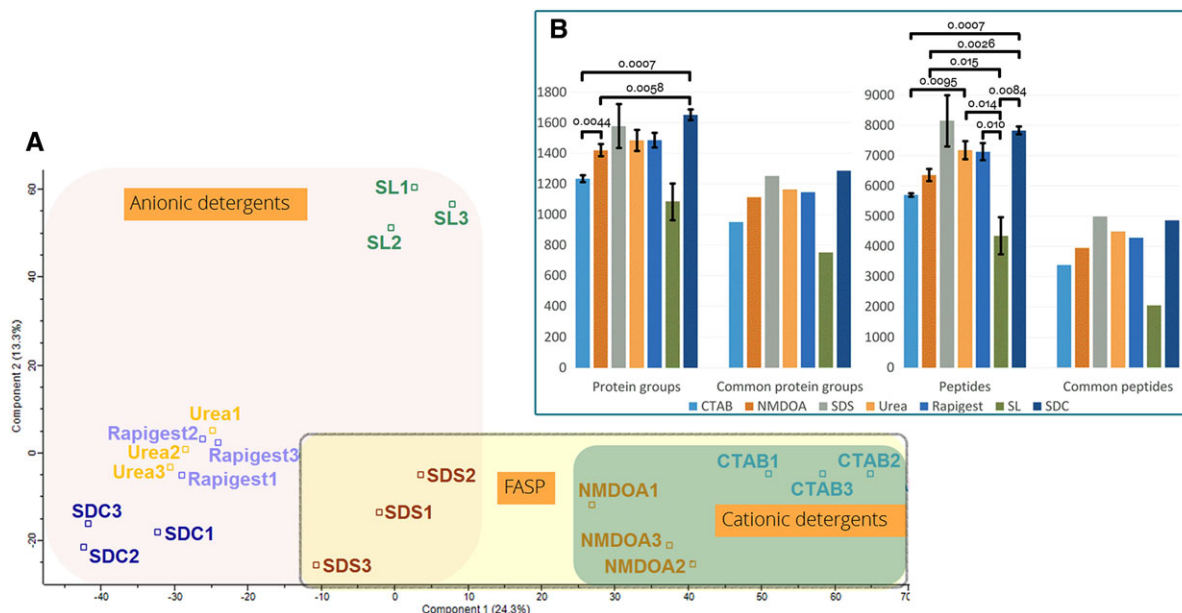
Figure 1. Comparison of sample preparation methods. (A) Principal component analysis (PCA) of protein group intensities resulting from the different sample preparation methods. CTAB = cetyl trimethylammonium bromide, NMDOA = N-methyldioctylamine, SDS = sodium dodecyl sulfate, SL = sodium laurate, SDC = sodium deoxycholate (B) Numbers of identified protein groups and peptides with the different detergents. Shown are the mean numbers of identifications from three technical replicates (error bars indicating the standard deviation), and the numbers of common identifications among all three replicates. Significant differences after correction for multiple testing are indicated and the respective p-values included. The applicable Benjamini-Hochberg critical values for an FDR of 0.05 are 0.0071 for proteins and 0.017 for peptides (Table 1, Supporting Information).

A log2(x+1) transformation of the protein intensities was done. A principal component analysis of the detergent comparison data gave **Figure 1**A. After transformation, the data for the comparison of UC and healthy state was quantile normalized. Batch effects were removed from the quantile normalized data using the combat method.[18] The data were then analyzed by principal component analysis (PCA), and by using the limma package, Cluster Profiler and revigo.

3. Results and Discussion

3.1. Method Development for Sample Preparation

To optimize the number of peptides and proteins identified, we report herein a study of different sample preparation methods of colon biopsies and evaluate their effectiveness based on numbers of identified peptides and proteins. We also assess the reproducibility between replicates, which is important when labelling on peptide level.

3.2. Which Detergents were Tested, and What is the Rationale?

In this publication, we compare seven different workflows, including two cationic (CTAB, NMDOA) and four anionic (SDS, Rapigest, SL, SDC) detergents, as well as one workflow without any detergent. We chose the detergent concentrations based on previously published methods. For the detergents with a fixed charge (SDS, CTAB) we chose 1% concentration, as previously reported for SDS.[19,20] We chose a 5% concentration, as

earlier applied for SDC,[21,22] for SDC, SL and NMDOA, which have a pH-dependent charge. We used a Rapigest concentration of 0.25%, since the lowest concentration recommended by the manufacturer is 0.1%. Increasing the concentration further is unfavorable due to the high cost of Rapigest.

Samples prepared with the same method cluster strongly in a PCA. In addition, similarities between methods result in grouping (Figure 1; see supporting information for correlation data, Figure 1, Supporting Information).

We observed no strong correlation between peptide yield and protein identification numbers (see Figure 2, Supporting Information).

3.3. SDC Performs Best

SDC gave the overall highest average of protein identification numbers (Figure 1). SDC is significantly better than CTAB and NMDOA regarding both peptide and protein identification numbers and better than SL for peptide identification numbers (Benjamini-Hochberg FDR < 0.05, Figure 1).[23] The high proportion of missed cleavages observed with SL (Figure 4, Supporting Information) indicates that the enzymatic digestion is strongly impaired when both SL and urea are present. The protein and peptide identification numbers with SDC are not significantly higher than for urea and Rapigest when correcting for multiple testing, but the higher numbers of commonly identified proteins and peptides (Figure 1) make SDC preferable.

The identification numbers of SDC and SDS are not significantly different either; however, the variances of identification numbers (Figure 1) and of correlation coefficients are higher

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Clinical Applications
www.clinical.proteomics-journal.com

for SDS than for SDC (spearman rank correlation mean ± SD: SDCpeptides 0.889 ± 0.004; SDSpeptides 0,836 ± 0.04; SDCproteins 0.907 ± 0,006; SDSproteins 0.883 ± 0.035; see Figure 3, Supporting Information). For most applications, including our clinical study, a low variance in the sample preparation is a necessity, as a high variance will decrease the meaningfulness of a dataset strongly. Furthermore, the risk of misinterpretations affecting single samples would be detrimental for possible clinical applications. Recurring small margins of error are a lesser problem.

A high reproducibility in the generation of peptides is necessary for reliable and precise quantifications using labelling reagents at peptide level, such as TMT. Labelling after enzymatic cleavage to peptides demands that samples are processed separately before mixing the samples. This can lead to non-reproducible changes of the sample composition and in turn to an imprecise determination of protein abundance ratios between samples. In clinical samples, such shifts in the measured protein abundance ratios may be mistaken for actual differences between the investigated tissues.

We obtained the highest number of common proteins, calculated as proteins identified in all three replicates, with SDC (Figure 1). The reproducible detection of the same protein groups over multiple LC-MS/MS experiments limits the number of missing values in later analyses. SDC has previously been shown to also perform very well with highly lipophilic tissues.[20]

An effortful additional sample processing such as FASP is not required for the removal of SDC, which saves time and diminishes the risk of contaminations.

SDC is among the best under all evaluated criteria, while other tested detergents show major weaknesses in at least one of these. We conclude that SDC complements the tandem Lys-C/trypsin digestion as a simple and cost-effective measure to enhance peptide and protein identifications with high reproducibility.

3.4. Preferential Identification of Transmembrane Proteins with Cationic Detergents

A comparably high proportion of transmembrane proteins in the identified proteins was observed for samples prepared with cationic detergents, especially with NMDOA (proportion of transmembrane proteins ± SD: CTAB 15.3% ± 0.8%; NMDOA 16.1% ± 0.8%; SDS 11.8% ± 1.5%; urea 13.2% ± 0.4%; Rapigest 13.8% ± 0.3%; SL 11.0% ± 0.3%; SDC 14.0% ± 0.4%; Figure 4 and Table 3, Supporting Information).

These results indicated that cationic detergents could be a promising option when targeting transmembrane proteins, but the applied methods demand further optimization.

3.5. Comparative Study of UC-Affected and Healthy Human Colon Mucosa

3.5.1. UC-Affected Differs Strongly from Healthy Tissue

Principal component analysis of the batch corrected data[24] shows clear separation in principal component one between
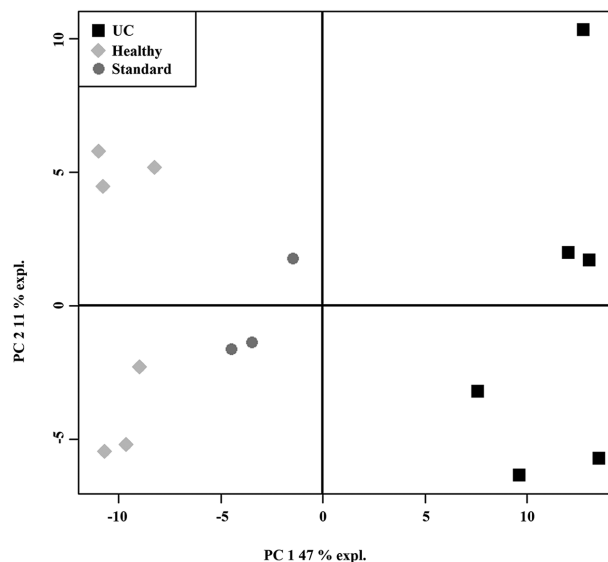
Figure 2. Principal component analysis. The colon biopsies from patients with ulcerative colitis separate strongly from healthy tissue samples. The reference is a mixture of both UC and healthy tissue samples and groups in between.

six patients newly diagnosed with UC and healthy controls (**Figure 2**). The reference samples, which are a mixture of UC and control material, lie between the two groups as expected. 2366 proteins with zero counts in fewer than five samples were included in the PCA analysis. Component one explains 47% of variance due to the strength of the phenotype. A normal mucosa contains primarily epithelial and stromal cells, and this is compared to a severely inflamed tissue with strong infiltration of many types of immune cells, as reflected in the gene ontology analysis (**Figure 3**). This creates a single strong biological signal.

3.5.2. Which Proteins and Functions are Affected?

Differentially expressed proteins were determined by applying the limma package to the batch corrected counts data.[25] Proteins were called as differentially expressed if they had a p-value smaller than 0.05 after correction for multiple testing and a fold change (log base 2) larger than |0.7|. This identified 68 proteins that were up-regulated in UC patients and 100 that were down-regulated (Table 2, Supporting Information).

To help interpret the biomedical significance of the differentially expressed proteins, the lists were analyzed for overrepresented gene ontology terms using Cluster Profiler.[26] Cluster profiler allows the identification of biological processes that are overrepresented among the lists of differentially expressed genes. The hierarchical nature of the gene ontology causes the detection of a large number of similar categories often with similar gene memberships. We therefore used Revigo to identify non redundant gene ontology terms to focus the interpretation on (Figure 3).[27]

Up-regulated proteins are dominated by immune system related processes. This is unsurprising given the significant infiltration of immune cells into the gut lining during active UC. This

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

Proteomics
Clinical Applications

www.clinical.proteomics-journal.com

Figure 3. Differential regulation of gene ontology categories in UC compared to healthy tissue.

Table 2. The 20 most regulated proteins in ulcerative colitis.

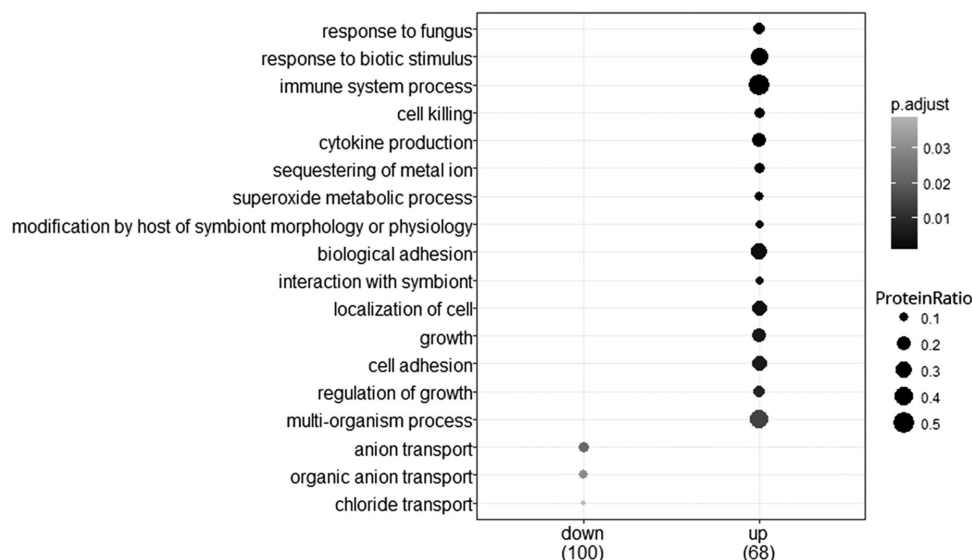| Protein name	Gene name	Fold change [intensity UC/healthy]	p.adj.
Carbonic anhydrase 1	CA1	0.194	1.12E-06
Fatty acid-binding protein, liver	FABP1	0.196	4.59E-07
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	0.218	5.59E-03
Carbonic anhydrase 2	CA2	0.222	2.03E-10
Poly(ADP-ribose) glycohydrolase	PARG	0.225	1.60E-05
3-keto-steroid reductase isoform 2	HSD17B7	0.257	5.16E-08
Trefoil factor 3	TFF3	0.271	2.24E-07
Creatine kinase B-type	CKB	0.277	2.71E-07
Hydroxymethylglutaryl-CoA synthase, mitochondrial	HMGCS2	0.282	2.11E-06
Normal mucosa of esophagus-specific gene 1 protein	C15orf48	0.305	4.88E-06
Selenium-binding protein 1	SELENBP1	0.305	4.80E-08
Cytochrome c oxidase subunit 2	COX2	0.315	8.77E-06
Neural cell adhesion molecule 2	NCAM2	0.322	1.44E-06
Calcium-activated chloride channel regulator 4	CLCA4	0.324	1.41E-04
Intelectin-1	ITLN1	0.328	1.56E-04
Protein S100-A12	S100A12	3.286	2.64E-06
Protein S100-A8	S100A8	3.875	2.95E-06
Protein S100-A9	S100A9	3.903	2.53E-06
Neutrophil defensin 1	DEFA1	3.916	3.21E-06
Myeloblastin	PRTN3	5.082	1.40E-05

Fifteen proteins are lower abundant in UC compared to healthy tissue (ratio < 1), while five higher abundant inflammatory proteins (ratio > 1) are among the most regulated proteins. Three of these are S100 proteins.

infiltration of immune cells reflects in the gene ontology terms related to response to biotic stimulus and response to fungus. The cell-cell communication, such as between immune cells and epithelial cells, is involved in maintaining the inflamed state in UC. We see this communication in gene ontology categories like cytokine production and cell killing, the latter reflects probably the influx of cytotoxic T cells. On the down-regulated side the picture is less clear, but the gene ontology terms reflect ion trans-

port both of organic and inorganic ions such as SELENBP1 and CLCA4. Other transport related proteins are also down-regulated, such as CDH17, and epithelial cell adhesion molecules such as EPCAM, CEACAM7 and CLDN7. Downregulation of these normal epithelial functions accommodates immune cell infiltration and may also reflect the dysfunction of the epithelium involved in the UC etiology. Mucus proteins, e.g. MUC2 and TFF3, are also lower abundant in UC.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**Proteomics
Clinical Applications**

www.clinical.proteomics-journal.com

3.5.3. S100 Proteins

Our study found seven of in total 11 quantified S100 proteins differently abundant in UC than in healthy tissue. Three of these are among the five most up-regulated proteins (**Table 2**).

The S100 family consists of 24 proteins with EF-hand calcium-binding properties. These proteins are further categorized into three subgroups, which exert only intracellular, both intracellular and extracellular, and mainly extracellular effects, respectively. The proteins of the S100 family are involved in a wide range of functions. These functions include for instance regulatory effects on immune response, cell migration, metabolism, and tissue repair and development,[28] as well as antimicrobial effects.[29]

Our findings confirm the previously reported increased levels in UC for S100A8 and S100A9, which form the heterodimer calprotectin, S100A12, S100P,[30] and S100A11.[7]

The calprotectin complex of S100A8 and S100A9, in combination with a second stimulus such as bacterial lipopolysaccharide, is an endogenous ligand of Toll-like receptor 4 (TLR-4) and induces the NF-κB-dependent expression of pro-inflammatory proteins such as TNFα. S100A8 seems to be the active part in the interaction with TLR-4,[31] while S100A9 stabilizes S100A8. The latter was shown in MRP-14 (S100A9) null mice, in which S100A8 was undetectable, even though the mRNA for S100A8 was present.[32] Fold changes for S100A8 and S100A9 provided in Supporting Information support these findings (Figure S7).

The Receptor for Advanced Glycation Endproducts (RAGE) is a common, but not exclusive, interaction partner of S100A11, S100A12 and S100P.[33,34]

Conversely, our data shows that S100A14 and S100A16 are significantly lower abundant in UC tissue. Less is known about these proteins, and most research focused on these proteins' roles in cancer. S100A14 seems to have the capability of both increasing and decreasing tumor invasiveness, depending on whether the respective tissue expresses wild type or mutant p53. S100A16 is up-regulated in different cancers[28] and an adipogenesis-promoting factor with negative impact on insulin sensitivity.[35]

These findings indicate a substantial role of S100 proteins in the pathophysiology of UC, which should be elucidated in future studies.

3.5.4. Comparison with Previous Findings

The findings in our study are largely consistent with previous studies with different study designs and analytical approaches. The study designs comprise comparisons of macroscopically normal tissue from treated UC patients with healthy controls,[8] of inflamed tissue from pediatric IBD patients with pediatric controls,[7] and of inflamed tissue with uninflamed tissue from the same adult UC patients.[36] Applied analytical approaches are SDS-PAGE followed by immunoblotting[37] and peptide mass fingerprinting,[36] respectively, label-free LC-MS/MS,[8] and SILAC.[7] Despite these differences, significantly changes in abundances point in the same direction for 37 of 38 proteins, which were commonly identified as significantly changed in both our dataset and at least one other study. The only exception is thiosulfate transferase (TST); this protein was found

to be significantly down-regulated in the inflamed tissue both in the pediatric study by Starr et al. and by us, while it was significantly up-regulated in the comparison of inflamed and uninflamed tissue from the same UC patients. The differences in the magnitude of fold-changes, especially between label-free quantification and other methods, are most likely an artifact resulting from the different analytical approaches.

Our results validate the findings from previous studies and vice versa. The similarity in proteome changes despite very different study designs is remarkable.

3.5.5. Perspectives

The study identifies differences between healthy and UC-affected tissue. This gives a deep view into the pathophysiological background and is a first step toward the identification of diagnosis biomarkers. However, using the full potential of biomarker identification demands a larger sample basis and the investigation of more groups and subgroups, e.g. the comparison of UC and Crohn's disease. This will be the focus in upcoming studies. Once clinically relevant biomarkers are identified from biopsies, the subsequent aim must be to quantify and validate these with targeted approaches in easier accessible sample material, such as feces and blood. Calprotectin, a complex of the proteins S100A8 and S100A9, is an example for a fecal biomarker used in IBD management,[38] indicating that feces can be a promising sample material for protein biomarkers.

4. Conclusions

This study reveals the so far deepest view into the ulcerative colitis proteome in colon biopsies from untreated adult patients in the initial phase of the disease. The biopsy preparation was optimized for this purpose; SDC augmentation of a urea containing lysis buffer enhances both the peptide and the protein identification. We found 68 up- and 100 down-regulated proteins ($p < 0.05$, fold change (log base 2) $> |0.7|$) in ulcerative colitis compared with healthy controls. Immune response proteins are up-regulated in UC, while proteins involved in anion transport, epithelial cell adhesion and mucus production are down-regulated.

Abbreviations

BCA, bicinchoninic acid; CD, Crohn's disease; FA, formic acid; FASP, filter-aided sample preparation; IBD, Inflammatory bowel disease; Lys-C, endoproteinase Lys-C; NMDOA, N-methyldioctylamine; PCA, principal component analysis; SDC, sodium deoxycholate; SL, sodium laurate; TEAB, Triethylammonium bicarbonate buffer; Th, helper T cells; TLR-4, Toll-like receptor 4; TMT, Tandem Mass Tag; TNF, tumor necrosis factor; UC, ulcerative colitis; UCDAI, ulcerative colitis disease activity index

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Proteomics
Clinical Applications
www.clinical.proteomics-journal.com

Conflict of Interest

The authors have declared no conflict of interest.

Keywords

colon biopsies, inflammatory bowel disease, S100 proteins, sodium deoxycholate, ulcerative colitis

[1] J. L. Round, S. K. Mazmanian, *Nat. Rev. Immunol.* **2009**, *9*, 313.
[2] M. F. Neurath, *Nat. Rev. Immunol.* **2014**, *14*, 329.
[3] S. Danese, C. Fiocchi, *N. England J. Med.* **2011**, *365*, 1713.
[4] Y. Yau, R. W. Leong, M. Zeng, V. C. Wasinger, *J. Gastroenterol. Hepatol.* **2013**, *28*, 1076.
[5] J.-M. Park, N. Y. Han, Y.-M. Han, M. K. Chung, H. K. Lee, K. H. Ko, E.-H. Kim, K. B. Hahm, *World J. Gastroenterol.* **2014**, *20*, 13466.
[6] T. Bennike, S. Birkelund, A. Stensballe, V. Andersen, *World J. Gastroenterol.* **2014**, *20*, 3231.
[7] A. E. Starr, S. A. Deeke, Z. Ning, C.-K. Chiang, X. Zhang, W. Mottawea, R. Singleton, E. I. Benchimol, M. Wen, D. R. Mack, A. Stintzi, D. Figeys, *Gut* **2017**, *66*, 1573.
[8] T. B. Bennike, T. G. Carlsen, T. Ellingsen, O. K. Bonderup, H. Glerup, M. Bøgsted, G. Christiansen, S. Birkelund, A. Stensballe, V. Andersen, *Inflammat. Bowel Dis.* **2015**, *21*, 2052.
[9] M. Mann, N. L. Kelleher, *Proc. Natl Acad. Sci.* **2008**, *105*, 18132.
[10] T. Viswanatha, M. J. Pallansch, I. E. Liener, *J. Biological Chem.* **1955**, *212*, 301.
[11] T. Glatter, C. Ludwig, E. Ahrné, R. Aebersold, A. J. R. Heck, A. Schmidt, *J. Proteome Res.* **2012**, *11*, 5145.
[12] B. E. Sands, C. Cuffari, J. Katz, S. Kugathasan, J. Onken, C. Vitek, W. Orenstein, *Inflammatory Bowel Disease* **2004**, *10*, 677.
[13] L. R. Sutherland, F. Martin, S. Greer, M. Robinson, N. Greenberger, F. Saibil, T. Martin, J. Sparr, E. Prokipchuk, L. Borgen, *Gastroenterology* **1987**, *92*, 1894.
[14] K. R. Rebecchi, E. P. Go, L. Xu, C. L. Woodin, M. Mure, H. Desaire, *Analyt. Chem.* **2011**, *83*, 8484.
[15] J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, *Nat. Method.* **2009**, *6*, 359.
[16] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367.
[17] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, J. Cox, *Nat. Methods* **2016**, *13*, 731.
[18] W. E. Johnson, C. Li, A. Rabinovic, *Biostatistics (Oxford, England)* **2007**, *8*, 118.
[19] N. Zhang, R. Chen, N. Young, D. Wishart, P. Winter, J. H. Weiner, L. Li, *Proteomics* **2007**, *7*, 484.
[20] Y. Pasing, S. Colnoe, T. Hansen, *Proteomics* **2017**, *17*, 1500462.
[21] I. R. León, V. Schwämmle, O. N. Jensen, R. R. Sprenger, *Mol. Cell. Proteomics* **2013**, *12*, 2992.
[22] Y. Lin, J. Zhou, D. Bi, P. Chen, X. Wang, S. Liang, *Analyt. Biochem.* **2008**, *377*, 259.
[23] Y. Benjamini, Y. Hochberg, *J. Royal Statistical Soc.* **1995**, *57*, 289.
[24] J. Misra, W. Schmitt, D. Hwang, L. L. Hsiao, S. Gullans, G. Stephanopoulos, G. Stephanopoulos, *Genom. Res.* **2002**, *12*, 1112.
[25] G. K. Smyth, *Statist. Applicat. Genetics Mol. Biol.* **2004**, *3*, 1544.
[26] G. Yu, L. G. Wang, Y. Han, Q. Y. He, *Omics* **2012**, *16*, 284.
[27] F. Supek, M. Bosnjak, N. Skunca, T. Smuc, *PloS One* **2011**, *6*, e21800.
[28] R. Donato, B. R. Cannon, G. Sorci, F. Riuzzi, K. Hsu, D. J. Weber, C. L. Geczy, *Curr. Mol. Med.* **2013**, *13*, 24.
[29] J. P. Zackular, W. J. Chazin, E. P. Skaar, *J. Biological Chem.* **2015**, *290*, 18991.
[30] A. C. Manolakis, A. N. Kapsoritakis, E. K. Tiaka, S. P. Potamianos, *Digestive Dis. Sci.* **2011**, *56*, 1601.
[31] T. Vogl, K. Tenbrock, S. Ludwig, N. Leukert, C. Ehrhardt, M. A. D. van Zoelen, W. Nacken, D. Foell, T. van der Poll, C. Sorg, J. Roth, *Nat. Methods* **2007**, *13*, 1042.
[32] J. A. R. Hobbs, R. May, K. Tanousis, E. McNeill, M. Mathies, C. Gebhardt, R. Henderson, M. J. Robinson, N. Hogg, *Mol. Cell. Biol.* **2003**, *23*, 2564.
[33] E. Leclerc, G. Fritz, S. W. Vetter, C. W. Heizmann, *Biochim. et Biophys. Acta (BBA)* **2009**, *1793*, 993.
[34] L. J. Sparvero, D. Asafu-Adjei, R. Kang, D. Tang, N. Amin, J. Im, R. Rutledge, B. Lin, A. A. Amoscato, H. J. Zeh, M. T. Lotze, *J. Translat. Med.* **2009**, *7*, 17.
[35] Y. Liu, R. Zhang, J. Xin, Y. Sun, J. Li, D. Wei, A. Z. Zhao, *Endocrinology* **2011**, *152*, 903.
[36] N. A. Poulsen, V. Andersen, J. C. Møller, H. S. Møller, F. Jessen, S. Purup, L. B. Larsen, *BMC Gastroenterol.* **2012**, *12*, 76.
[37] R. F. Giovanni, L. Renzo, C. Giuseppe, F. Adriana, M. Y. Sambuy, *Digestive Dis. Sci.* *43*, 2086.
[38] N. E. Walsham, R. A. Sherwood, *Clin. Experiment. Gastroenterol.* **2016**, *9*, 21.

Supporting information

Detergent removal

FASP

TFA-NaOH solution was prepared as follows: 460 µl trifluoroacetic acid were added bit by bit with cooling periods in between to 350 µl water. 6 M NaOH was added slowly and interrupted by cooling until pH > 7 was reached and the solution was buffered with 200 µl 1 M TEAB pH 8.5 solution.

CTAB

After alkylation, the samples were diluted to 200 µl with 150 µl of 5 M urea in 0.1 M TEAB. The filter devices were equilibrated with 100 µl 8 M urea in 0.1 M TEAB and centrifuged for 20 min at 16 000 g. The diluted samples were added on the filter and centrifuged (30 min, 16 000 g). The samples were washed with 100 µl 5 M urea in 0.1 M TEAB and centrifuged (20 min, 16 000 g). They were incubated with 50 µl TFA-NaOH solution for 10 min and centrifuged 5 min after addition of 50 µl isopropanol. 200 µl TFA-NaOH solution were immediately added, followed by 30 min centrifugation. The samples were incubated for 10 min with 50 µl TFA-NaOH solution, and centrifuged 5 min after addition of 50 µl isopropanol. Following this, washing steps with 200 µl of TFA-NaOH solution, 8 M urea in 0.1 M TEAB, 0.1 M TEAB, and 6 M urea in 0.1 M TEAB respectively were performed with 30 min centrifugation each. 45 µl 8 M urea in 0.1 M TEAB were added to the filters.

NMDOA

After alkylation, the samples were diluted to 200 µl with 150 µl 6 M urea in water. The filter devices were equilibrated with 100 µl 8 M urea in 0.1 M TEAB and centrifuged for 20 min at 16 000 g. The diluted samples were added on the filters and centrifuged (20 min, 16 000 g). The samples were washed with 150 µl 6 M urea in water (20 min centrifugation), with 200 µl 8 M urea in 0.1 M TEAB (20 min centrifugation), three times with 200 µl TFA-NaOH solution (20 min, 20 min, and 40 min centrifugation, respectively), with 200 µl 8 M urea in 0.1 M TEAB (30 min centrifugation), and two times with 150 µl and once with 300 µl 6 M urea in 0.1 M TEAB (20 min, 20 min and 30 min centrifugation, respectively). 45 µl 8 M urea in 0.1 M TEAB were added to the filters.

SDS

After alkylation, the samples were diluted to 200 µl with 150 µl of 8 M urea in 100 mM Tris. The filter devices were equilibrated with 100 µl 8 M urea, 0.1 M TEAB and centrifuged for 20 min at 16 000 g. The diluted samples were added on the filter and centrifuged (20 min, 16 000 g). The samples were washed once with 200 µl and three times with 100 µl 8 M urea in 100 mM Tris (20 min, 20 min, 20 min, and 30 min centrifugation, respectively), and

three times with 100 µl 100 mM TEAB buffer (20 min centrifugation each). 45 µl 8 M urea in 0.1 M TEAB were added to the filters.

Non-FASP

Acid precipitation

After enzymatic cleavage, FA in water (50% v/v) was added to SDC and SL samples to a final FA concentration of 2.5%. Detergent sedimentation was enforced by centrifugation (15 min, 16 000 g) and the peptide solutions were transferred to new Eppendorf tubes.

Acid cleavage

10 % TFA was added to a final concentration of 2% TFA. The samples were incubated at 37°C for 30 min, centrifuged for 15 min and the peptide solutions were transferred to new Eppendorf tubes.

Additional statistical data for identification numbers

Table S1: Average protein and peptide identification numbers from three replicates and their standard deviations (SD).

Detergent	Average Protein IDs	SD protein IDs	Average peptide IDs	SD peptide IDs
CTAB	1234	22	5699	84
NMDOA	1420	37	6358	229
SDS	1578	123	8156	877
Urea	1484	60	7177	329
Rapigest	1486	44	7128	305
SL	1085	101	4344	643
SDC	1652	33	7830	158

Table S2: p-values for the comparisons of identification numbers. Significant changes after correcting for multiple testing are indicated and the applicable Benjamini-Hochberg values can be determined.

proteins					peptides		
compared detergents		p-value	rank	B-H critical value 0.05	compared detergents		p-value
CTAB	SDC	0.00070274	1	0.002380952	CTAB	SDC	0.00068614
CTAB	NMDOA	0.00435248	2	0.004761905	NMDOA	SDC	0.00256563
NMDOA	SDC	0.00582587	3	0.007142857	SL	SDC	0.00843249
SL	Rapigest	0.01154641	4	0.00952381	CTAB	urea	0.00947029
SL	SDC	0.02180624	5	0.011904762	SL	Rapigest	0.01023476
CTAB	urea	0.02196687	6	0.014285714	SL	Urea	0.01406103
CTAB	Rapigest	0.0241305	7	0.016666667	NMDOA	SL	0.01493767
urea	SDC	0.02713754	8	0.019047619	CTAB	Rapigest	0.02101646
NMDOA	SL	0.04140153	9	0.021428571	urea	SDC	0.02465533
CTAB	SDS	0.06730272	10	0.023809524	NMDOA	urea	0.02766078
SL	Urea	0.07014393	11	0.026190476	CTAB	NMDOA	0.03150387
Rapigest	SDC	0.08264946	12	0.028571429	CTAB	SDS	0.04356462
SL	SDS	0.08605078	13	0.030952381	NMDOA	Rapigest	0.04648088
CTAB	SL	0.18401627	14	0.033333333	SL	SDS	0.04776679
NMDOA	Rapigest	0.26136542	15	0.035714286	CTAB	SL	0.06656174
NMDOA	SDS	0.2736621	16	0.038095238	Rapigest	SDC	0.0792265
NMDOA	urea	0.29595327	17	0.04047619	NMDOA	SDS	0.10647926
urea	SDS	0.29836417	18	0.042857143	Rapigest	SDS	0.22660956
Rapigest	SDS	0.49548441	19	0.045238095	urea	SDS	0.27314364
SDS	SDC	0.49716464	20	0.047619048	SDS	SDC	0.63340622
urea	Rapigest	0.98069951	21	0.05	urea	Rapigest	0.88732555
Significant for Benjamini-Hochberg FDR 0.05							

Correlations between the detergents
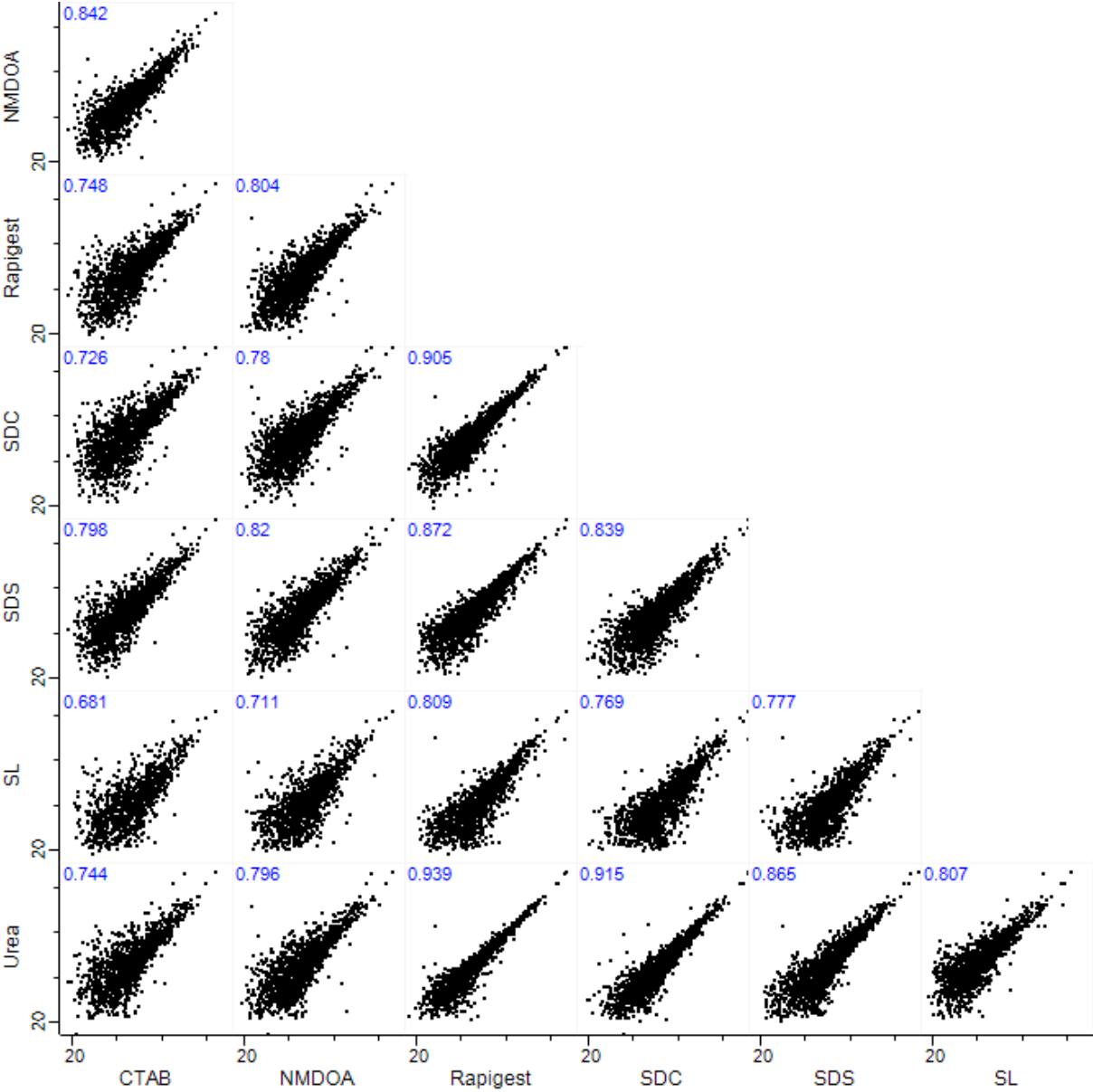
Figure S1: Correlation between protein intensities [log(2)] detected in three replicates with the respective lysis buffers. Blue numbers indicate spearman rank correlations.

Peptide yield and protein identification numbers
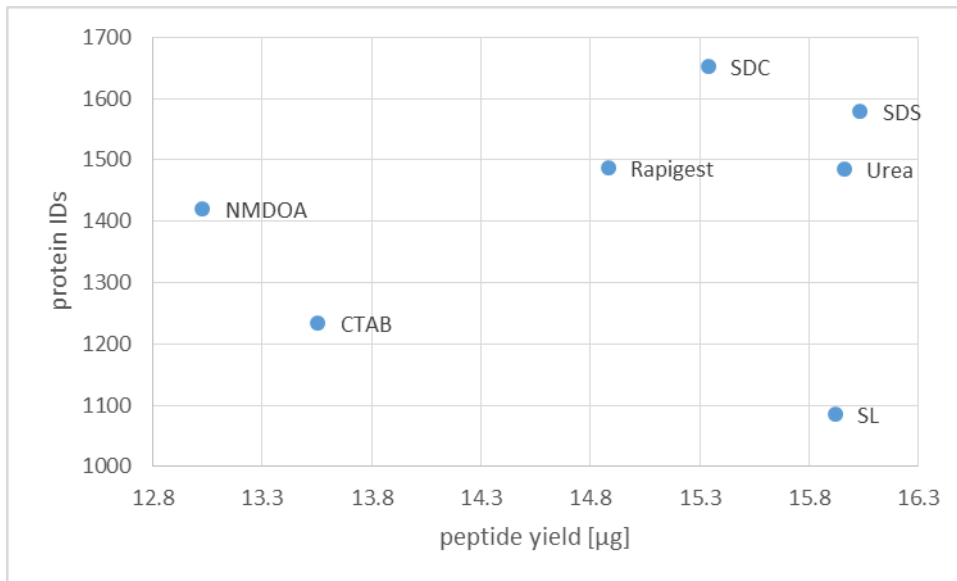
Figure S2: Peptide yield from 60 µg protein (as determined by BCA) and the corresponding number of protein identifications. The cationic detergents show a comparably low peptide yield and low protein ID numbers. However, overall there is no strong correlation (r^2 = 0.03). When excluding SL because the high missed cleavage numbers likely contribute to the low ID numbers r^2 raises to 0.49.
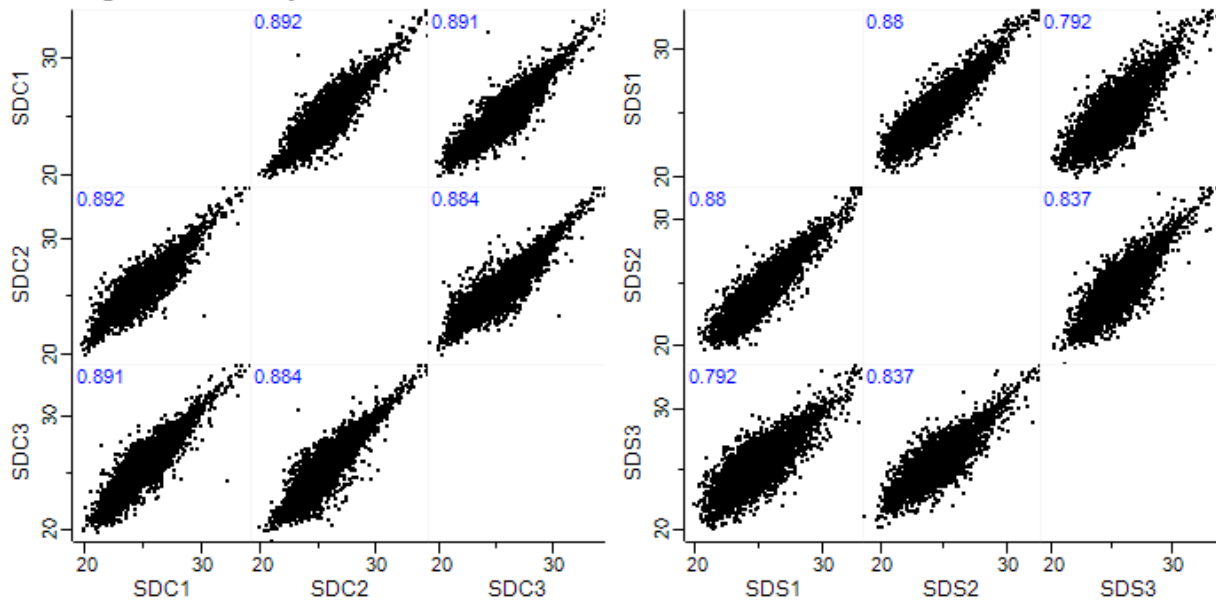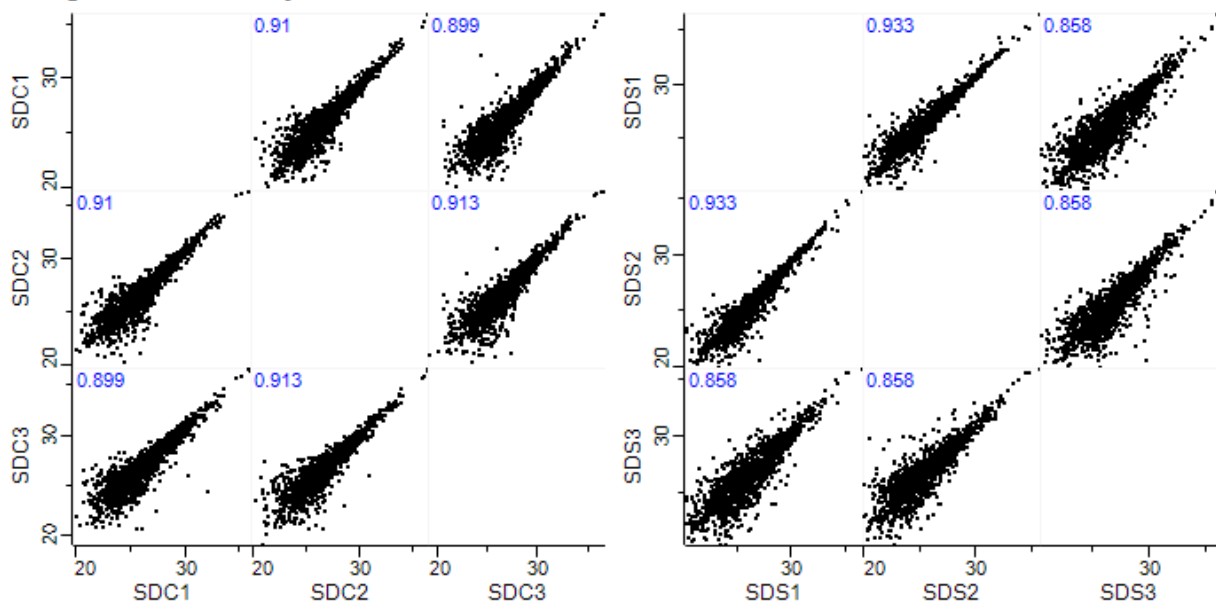
Figure S3: Correlations between peptide (A) and protein (B) intensities [log(2)] for replicates prepared with SDC and SDS, respectively. Blue numbers indicate spearman rank correlations.

Missed cleavages
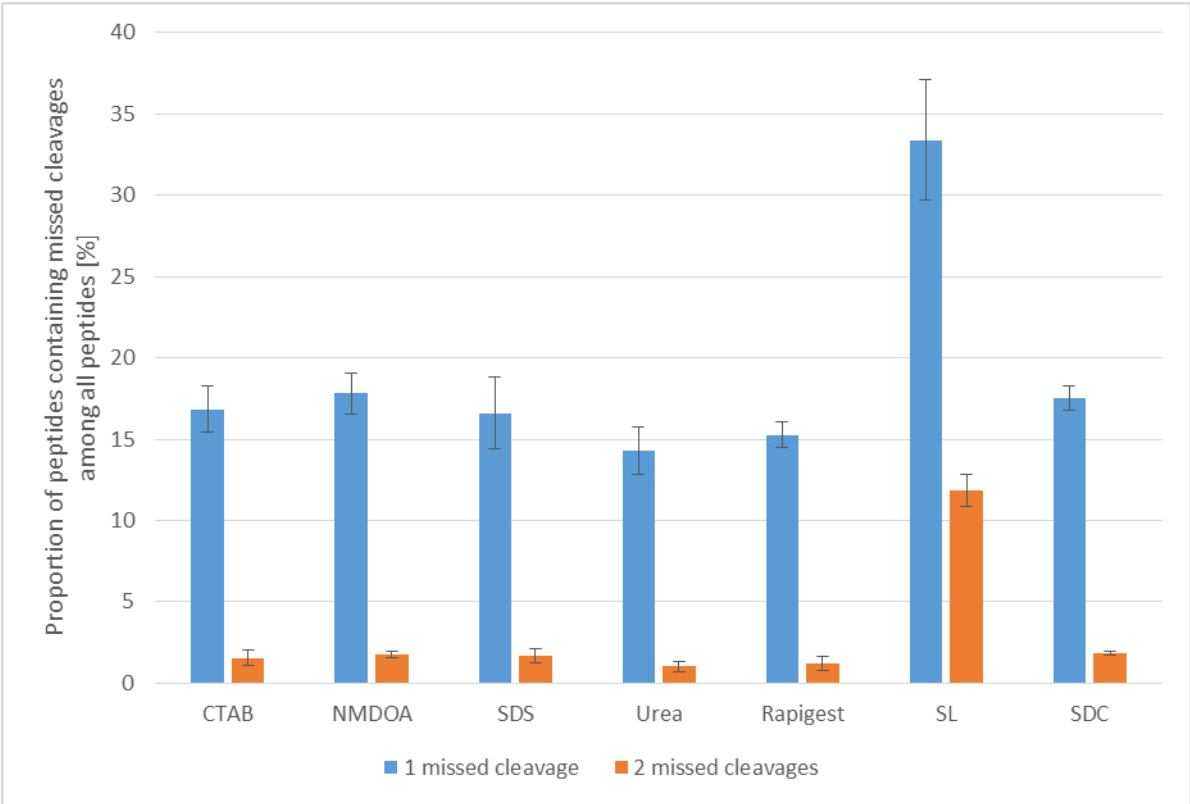
Figure S4: Relative missed cleavages

Preferential identification of transmembrane proteins with cationic detergents
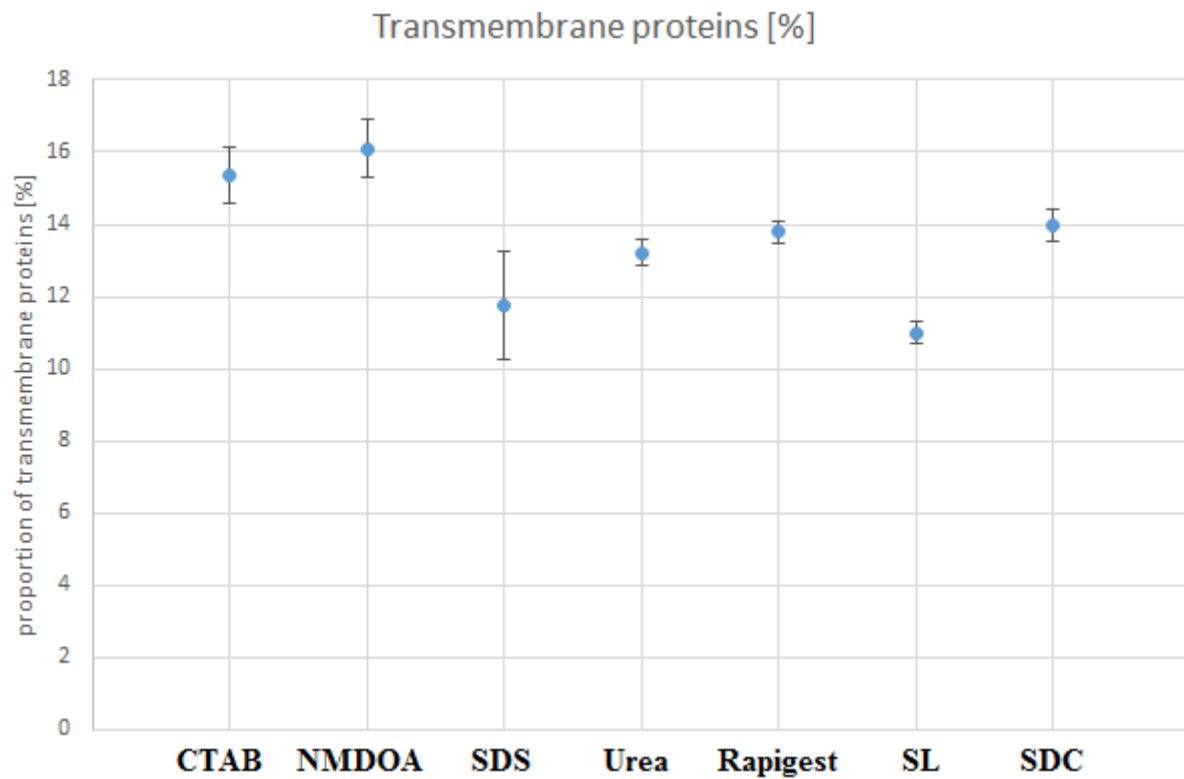
Figure S5: Average percentage of transmembrane proteins among all proteins identified with the different detergents. We determined the transmembrane proteins in Perseus by adding the "Keywords"-annotation (annotation database from http://annotations.perseus-framework.org) to the identified proteins and filtering for proteins with the annotation "Transmembrane".Error bars indicating the standard deviation:

Table S3: Absolute numbers of identified transmembrane proteins:

identified transmembrane proteins

Detergent	Experiment 1	Experiment 2	Experiment 3	average
CTAB	172	196	200	**189**
NMDOA	206	246	234	**229**
SDS	168	154	239	**187**
Urea	185	189	215	**196**
Rapi	217	202	196	**205**
SL	127	132	100	**120**
SDC	214	237	241	**231**

Measurement of protein concentration for clinical samples in buffer containing TEAB

The aliquots for the protein measurements were diluted 1:20 with water to yield a 5 mM TEAB concentration. Further solutions for the measurement were prepared by dilution with 5 mM TEAB. These solutions were measured against BSA standards in 5 mM TEAB.

Labelling methodology

For the 12 samples and the standard, aliquots of 25 μg peptides were labelled according to the manufacturer's protocol and the following scheme with 0.4 mg of the corresponding TMT label. Aliquots of 3.3 μg peptides from two ulcerative colitis samples, two control samples, and the standard were mixed for each LC-MS/MS experiment. We alternated the labels used for UC and healthy samples between the LC-MS/MS experiments in order to reduce a possible bias resulting from the labelling reagents.

Table S4: Sample labelling

	LC-MS/MS Experiment 1	LC-MS/MS Experiment 2	LC-MS/MS Experiment 3
• TMT6-126 Label Reagent	standard	standard	standard
• TMT6-127 Label Reagent	healthy1		healthy5
• TMT6-128 Label Reagent	UC1	healthy3	UC5
• TMT6-129 Label Reagent	healthy2	UC3	
• TMT6-130 Label Reagent	UC2	healthy4	healthy6
• TMT6-131 Label Reagent		UC4	UC6

Nanospray LC-MS/MS

The peptide concentration was determined from the absorption at $\lambda = 205$ nm with a NanoDrop ND-1000 Spectrophotometer.

1 μg peptides in 0.1% formic acid was injected on a EASY-LC system connected to a qExactive (Thermo Fisher Scientific, Bremen, Germany). Peptides were separated on a C18 column (2 μm, 100 Å, 50 μm, 50 cm) with a 4-40% gradient of acetonitrile in 0.1% FA over 60 min at a flow rate of 200 nl/min. The eluting peptides were analysed by the qExactive instrument operated in data dependent mode.

Inclusion lists for the UC-healthy comparison study

Inclusion lists for a second LC-MS/MS analysis of each sample were prepared with Proteome Discoverer software. These lists comprise the 5000 best peptide identifications among those, which did not reach high

confidence level, and which were derived from proteins, that were identified with less than high confidence. The retention time window was +/- 1 min. We retained the other settings for LC-MS/MS analysis unchanged, except for increased Maximum IT of 200 ms for the full MS .

We prepared inclusion lists for a third LC-MS/MS analysis from the combined data. These contained peptides, which were found with high confidence in at least one of the previous analyses, but were identified with at highest medium confidence in the corresponding sample. Peptides from proteins with a Sum PEP score >80 were excluded from this inclusion list. The retention time window was +/- 2 min. The experimental settings were the same as for the analyses with the first inclusion lists.

Batch correction by combat method
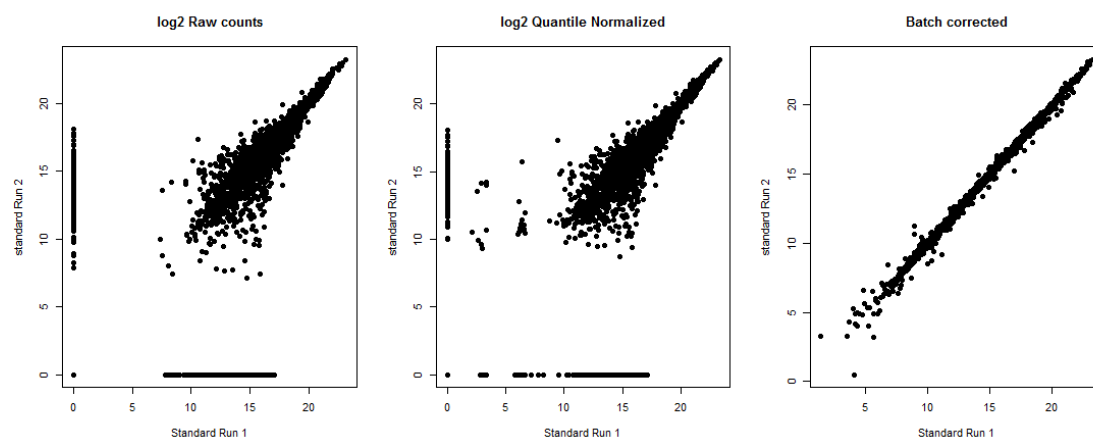
Figure S6: Batch correction of proteomics data.

Raw data repeats across batches show an average correlation of r2 = 0.34. Quantile normalization of the data does not significantly improve this. By batch correction using the combat method[1] repeatability is improved (r2 = 0.99).

Comparison of S100A8 and S100A9 abundances

log2(abundance in sample) - [average log2(abundance]]	healthy1	healthy2	healthy3	healthy4	healthy5	healthy6	UC1	UC2	UC3	UC4	UC5	UC6
Protein S100-A8	-0.987513	-1.04406	-1.15073	-0.684654	-1.03379	-0.961916	0.677631	0.614681	1.5319	0.678534	1.44281	0.917095
Protein S100-A9	-1.04642	-1.1719	-1.1056	-0.802761	-0.924828	-0.841908	0.699507	0.725145	1.45739	0.8116	1.45188	0.747898

■ Protein S100-A8 ■ Protein S100-A9

Figure S7: The figure shows a strong interdependence of the protein abundances detected for S100A8 and S100A9. The data is presented as log(2) difference from the average of all samples combined. Thereby, the figure also visualizes the up-regulation of both proteins in UC tissue.

Manufacturers

Trypsin: Sequencing Grade Modified Trypsin, Promega Corporation, Madison, USA

Lys –C: Wako Chemicals GmbH, Neuss, Germany

OMIX C18 100 µl pipette tips: Agilent Technologies, Santa Clara, USA

RapiGest™ SF: Waters, Manchester, UK

FASP filter devices: VWR Centrifugal Filter, modified PES 30K, 500 µl; VWR International

Pierce BCA Protein Assay Kit: Thermo Scientific, Rockford, USA

MagNA Lyzer: Roche Diagnostics AG, Rotkreuz, Switzerland

TMTsixplex reagent: Thermo Fisher Scientific, Rockford, USA

qExactive: Thermo Fisher Scientific, Bremen, Germany

[1] Johnson, W. E., Li, C., Rabinovic, A., Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England) 2007, 8, 118-127.

Paper II

RESEARCH

Ulcerative colitis: functional analysis of the in-depth proteome

Armin Schniers[1], Rasmus Goll[2,3], Yvonne Pasing[4], Sveinung Wergeland Sørbye[5], Jon Florholmen[2,3] and Terkel Hansen[1*]

Abstract

Background: Ulcerative colitis (UC) is one major form of inflammatory bowel disease. The cause and the pathophysiology of the disease are not fully understood and we therefor aim in this study to identify important pathophysiological features in UC from proteomics data.

Methods: Colon mucosa biopsies from inflamed tissue of untreated UC patients at diagnosis and from healthy controls were obtained during colonoscopy. Quantitative protein data was acquired by bottom-up proteomics and furthermore processed with MaxQuant. The quantitative proteome data was analyzed with Perseus and enrichment data was analyzed by ClueGO for Cytoscape.

Results: The generated proteome dataset is to-date the deepest from colon mucosa biopsies with 8562 identified proteins whereof 6818 were quantified in > 70% of the samples. We report abundance differences between UC and healthy controls and the respective p values for all quantified proteins in the supporting information. From this data set enrichment analysis revealed decreased protein abundances in UC for metallothioneins, PPAR-inducible proteins, fibrillar collagens and proteins involved in bile acid transport as well as metabolic functions of nutrients, energy, steroids, xenobiotics and carbonate. On the other hand increased abundances were enriched in immune response and protein processing in the endoplasmic reticulum, e.g. unfolded protein response and signal peptidase complex proteins.

Conclusions: This explorative study describes the most affected functions in UC tissue. Our results complemented previous findings substantially. Decreased abundances of signal peptidase complex proteins in UC are a new discovery.

Keywords: Inflammatory bowel disease, Ulcerative colitis, Calprotectin, Signal peptidase complex

Background

Ulcerative colitis (UC) is a form of inflammatory bowel disease (IBD). The chronic inflammation of the colon in UC starts at the rectum and can progress continuously to proximal colon parts. UC affects 0.51% of the European and 0.25% of the North American population, with increasing prevalence [1]. The disease has a complex pathophysiology and the exact disease causes remain unclear. A genetic component, environmental factors, defects of the epithelial barrier, and dysregulated immune responses are involved [2].

Pathophysiological changes likely lie or are reflected in the abundance and state of proteins, the major functional units in every tissue. The proteomic analysis of colon biopsies affected by UC gives data on these changes at the main site of the disease. Advances in sample preparation, instrumentation, and analysis software allow for increasing proteome coverage and improved quantification. Higher proteome coverage, growing and new pathway databases, and improved software packages benefit comprehensive enrichment analyses of tissues.

---

*Correspondence: Terkel.Hansen@uit.no
[1] Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway
Full list of author information is available at the end of the article

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 2 of 13

Proteomic studies with different designs regarding samples and analytical method were previously conducted on colon mucosa affected by UC. Inflamed mucosa from untreated patients at UC debut [3] as well as macroscopically normal mucosa from treated UC patients [4] have been compared to biopsies from healthy controls in previous studies. A comparison of inflamed mucosa from untreated UC patients with mucosa from healthy controls as well as from Crohn's disease patients was conducted on pediatric patients [5]. Another study compared inflamed and non-inflamed tissue from the same patients [6]. Despite the differences in the study designs, the studies present remarkably similar proteome changes [3]. They revealed novel insights into the UC pathophysiology, e.g. the role of neutrophil extracellular traps [4] and S100 proteins [3]. They also identified biomarkers for differentiating between UC and Crohn's disease [5] and between pancolitis and partial colitis [7]. Furthermore, multiple studies investigated specific proteins of interest and their functions in UC, such as inducible nitric oxide synthase (NOS2) [8] and histaminase (AOC1) [9].

However, despite the challenges in extracting biological information from omics data and the shortcomings of focusing on proteins with the highest fold-changes [10], enrichment analyses were not the main scope of previous proteomics studies on UC. Omics studies typically produce large amounts of data. Especially in experiments with strong phenotypes these datasets can contain so many significant findings that a sensible interpretation without additional statistical analyses is barely possible. Interpretations based on only the highest fold-changes in such a dataset disregard large amounts of data and fail to recognize functional changes as a result of multiple moderately different abundances. Moreover, the biological interpretation is prone to be arbitrary, and it has further disadvantages. Enrichment analyses can overcome these difficulties [10] and the enrichment factors give an additional indication for the importance of the respective functions for the disease pathophysiology.

Beside proteomics, transcriptomic approaches have been applied to investigate UC [11]. These studies show differential mRNA expressions in UC tissue. The pathophysiological implications of these finding however are less clear because the correlation of mRNA and protein abundances is poor. The cause for that low correlation is that both the translation and the degradation of proteins are subject to several regulation mechanisms that are independent of mRNA levels [12].

We conducted a comprehensive enrichment analysis on a proteomics dataset of 8562 identified proteins from colon mucosa biopsies.

Materials and methods

Patients included and biopsies collection

Mucosal biopsies were collected from newly diagnosed treatment-naïve UC patients in Norway. The UC diagnosis was established upon clinical, endoscopic and histological criteria defined by the European Crohn and Colitis Organization (ECCO) guidelines [13]. Furthermore, the degree of inflammation was evaluated during colonoscopy using the scoring system of ulcerative colitis disease activity index (UCDAI) [14]. Moreover, TNF mRNA levels were measured by real-time PCR to assess the level of UC activity [15]. Subjects admitted for a cancer screening and with normal colonoscopy and histological findings served as healthy controls. None of the recruited subjects suffered from irritable bowel disease and they were not on nonsteroidal anti-inflammatory drug medication prior to the colonoscopy. The biopsies from the UC patients were obtained from the most inflamed area in rectum or sigmoid colon. Biopsies from the control group were obtained from the rectum. From each study participant, 3 adjacent biopsies were obtained from the inflamed mucosa. One biopsy was immediately immersed in RNAlater (Qiagen, Germany). The second biopsy was frozen immediately in a dry cryotube tube at −70 °C until further analysis. The third biopsy was obtained for ordinary histological examinations (haematoxylin and eosin staining). 17 patients with debut of UC and 15 healthy controls were recruited as shown in Table 1.

Table 1 Baseline characteristics in patients included at debut of ulcerative colitis and in healthy controls (see "Patients included and biopsies collection" and "Baseline characteristics" sections for further details)

Study group	Number of subjects	Average age (SD)	Sex female/male	Median TNF-α (IQR)	Average UCDAI score (SD)	Average Geboes index (SD)	Extend (Montreal classification [65])
UC patients	17	39.4 (16.7)	4/13	14,350 (16,725)	8.7 (2.2)	7.9 (3.8)	2 Proctitis 9 left sided UC 6 Extensive UC
Healthy control	15	51.9 (14.3)	5/10	4500 (2400)	–	–	–

UCDAI score and Geboes index averages calculated from available data for 15 and 16 patients, respectively

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 3 of 13

Baseline characteristics

The baseline characteristics are shown in Table 1 (see Additional file 1: Table 1). All subjects included showed typical histological findings according to UC and normal, non-inflamed mucosa, respectively.

Sample preparation

Samples were homogenized with 250 µl cooled lysis buffer [8 M urea, 5% sodium deoxycholate (SDC), 100 mM triethylammonium bicarbonate buffer pH 8.5 (TEAB)] in MagNA Lyser Green Bead tubes (Roche Diagnostics AG, Rotkreuz, Switzerland) with a MagNA Lyser Instrument (Roche Diagnostics AG, Rotkreuz, Switzerland) for 35 s at 6500 rpm. Lysates were frozen at −70 °C until further sample preparation. All samples and standards were diluted/produced to contain 5 mM TEAB for the BCA assay (Pierce™ BCA Protein Assay Kit, Thermo Fisher Scientific). The assay was performed according to the manufacturer's protocol. Lysate aliquots of 60 µg protein were transferred to Protein LoBind tubes (Eppendorf AG, Hamburg, Germany). Disulfide bridges were reduced with 1,4-dithiothreitol (DTT) at a concentration of 5 mM by incubation at 54 °C for 30 min. Cysteins were alkylated with 15 mM iodoacetamide (IAA) and incubation for 30 min at room temperature in the dark. DTT solution corresponding to a final concentration of 5 mM was added to remove excess IAA. Lys-C predigestion was performed under gentle agitation for 8 h at 37 °C with 0.6 µg Lys-C (enzyme-to-protein ratio 1:100, w/w, Wako Chemicals GmbH, Neuss, Germany) in a buffer containing 1 mM calcium chloride, 6 M urea, and 100 mM TEAB. Calcium chloride solution, 3 µg trypsin (enzyme-to-protein ratio 1:20, w/w, Sequencing Grade Modified Trypsin, Promega Corporation, Madison, USA), water, and 1 M TEAB were added to a final concentration of 1 mM calcium chloride, 1 M urea, and 100 mM TEAB. Tryptic digestion was performed under gentle agitation for 16 h at 37 °C.

Equal peptide amounts from each sample were combined to produce a standard mixture, which was subsequently labelled with TMT126. Aliquots corresponding to 25 µg peptides were labelled with the remaining TMTsixplex isotopes according to the manufacturer's protocol. 0.4 mg TMT reagent were used for 25 µg peptides from samples with starting protein concentrations ≥ 2200 µg/ml. Higher amounts of TMT reagent were used for lower starting concentrations (see Additional file 1: Table 2).

The differently labelled peptides were mixed in equal amounts to in total 100 µg peptides. The digests were acidified with 50% formic acid (FA) to a final concentration of 2.5% FA and pH ≤ 2. The samples were centrifuged

at 16,000 g for 15 min and the supernatants were carefully transferred to fresh tubes. Acetonitrile was removed by evaporation in a vacuum concentrator.

Trifluoroacetic acid (TFA) was added to a concentration of 0.1%. The samples were fractionated with Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher, Rockford, USA) according to the manufacturer's protocol. Samples exceeding 300 µl were stepwise loaded with 2 min centrifugation after each loading. The fractions were dried in a vacuum concentrator and redissolved in 10 µl 0.1% TFA for subsequent LC–MS/MS analyses.

LC–MS/MS

The nano-LC–MS/MS analysis was performed on a Q Exactive mass spectrometer coupled to an EASY-nLC 1000 system (Thermo Fisher Scientific, Bremen, Germany).

Shortly, 2.0 µg peptides, as measured with a Nanodrop 2000 (Thermo Fisher Scientific, Bremen, Germany) at a wavelength of 205 nm and extinction coefficient 31 mg/ml [16], per sample were injected. They were concentrated on a reversed-phase trap column and subsequently separated on a reversed-phase main column. A binary solvent gradient of 0.1% FA (solvent A) and 0.1% FA in acetonitrile (solvent B) was used. The ACN proportion was increased from 0 to 5% over 19 min, further to 30% at 180 min and to 100% at 200 min. A detailed description of the LC–MS/MS conditions can be found in the supporting information (see Additional file 1).

Data evaluation

The raw data was analyzed with MaxQuant against a UniProt fasta file of all human proteins (see Additional file 1) [17, 18]. Further analyses were performed with Perseus [19] and Cytoscape/ClueGO [20, 21].

The data processing and visualization in Perseus is partially based on a recently published protocol by the Perseus developers [22]. The intensities were log2-transformed in Perseus, and the entries were filtered for the labels "potential contaminant", "reverse" and "only identified by site". Standard intensities were subtracted from the corresponding sample intensities and the resulting intensities *Z*-score transformed (matrix access: columns), which returns the normalized values (see Additional file 1).

Statistics and generation of figures and lists

Statistical analyses were performed in Perseus. The samples were assigned to the groups UC patients and healthy controls, respectively, and the groups containing the separate values for each sample were compared to each other with a two-sample test. Proteins were

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 4 of 13

annotated with gene ontology terms from the databases GOBP (biological process), GOMF (molecular function), GOCC (cellular component) [23], KEGG [24], GSEA [25], UniProt keywords [18], InterPro [26], Reactome [27], and PROSITE [28]. The lists and the article figures were prepared in Perseus and the Cytoscape app ClueGO (see Additional file 1). The WikiPathways database was used for the ClueGO analysis [29]. Pathways were mapped with HumanCyc (see Additional file 1) [30].

Results and discussion

Exploratory analysis

The proteomes of UC and healthy tissue differed strongly, as shown in Fig. 1. We present the 20 proteins with the highest abundance increases and decreases, respectively, in Table 2. The Pearson correlations of normalized intensities (see Additional file 1) were higher when samples of the same group were compared (i.e. UC to UC, or healthy control to healthy control) than when UC samples were compared to healthy controls (Fig. 1a). Also the hierarchical clustering of the sample correlations resulted in
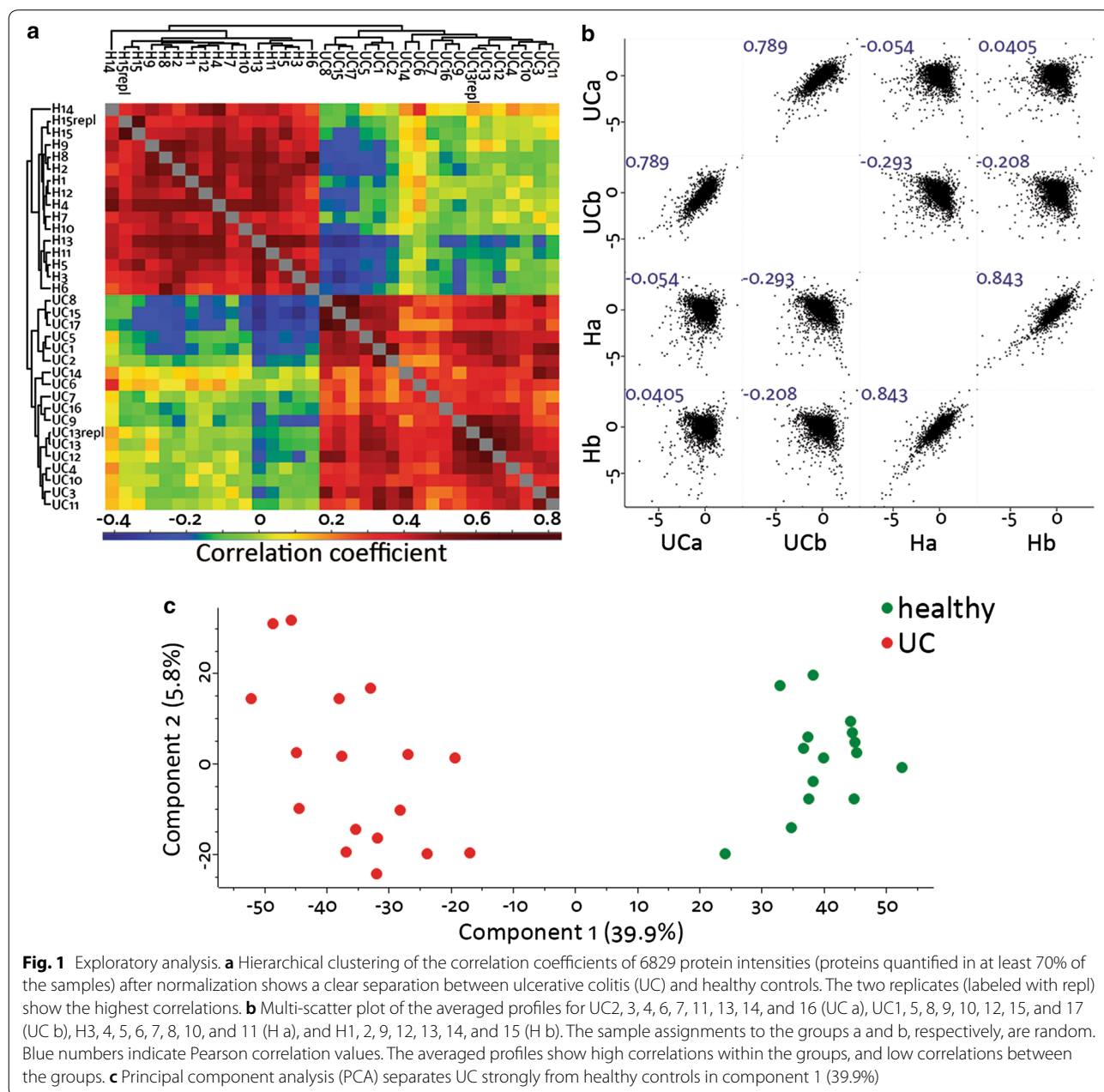
Fig. 1 Exploratory analysis. **a** Hierarchical clustering of the correlation coefficients of 6829 protein intensities (proteins quantified in at least 70% of the samples) after normalization shows a clear separation between ulcerative colitis (UC) and healthy controls. The two replicates (labeled with repl) show the highest correlations. **b** Multi-scatter plot of the averaged profiles for UC2, 3, 4, 6, 7, 11, 13, 14, and 16 (UC a), UC1, 5, 8, 9, 10, 12, 15, and 17 (UC b), H3, 4, 5, 6, 7, 8, 10, and 11 (H a), and H1, 2, 9, 12, 13, 14, and 15 (H b). The sample assignments to the groups a and b, respectively, are random. Blue numbers indicate Pearson correlation values. The averaged profiles show high correlations within the groups, and low correlations between the groups. **c** Principal component analysis (PCA) separates UC strongly from healthy controls in component 1 (39.9%)

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 5 of 13

Table 2 20 proteins with highest abundance increase and decrease, respectively, in UC patients as compared to healthy controls

Ratio UC/H	Protein name	Gene names	Majority protein ID	— Log p value
Proteins of increased abundance in UC compared to healthy controls				
10.12	Kinesin-like protein	KIF26B	B7WPD9	10.55
6.72	Protein S100-A8	S100A8	P05109	8.41
6.70	Cathelicidin antimicrobial peptide	CAMP	J3KNB4	9.71
6.17	Protein S100-A9	S100A9	P06702	8.07
4.95	Protein S100-A12	S100A12	P80511	8.36
4.73	Neutrophil defensin 1	DEFA1	P59665	7.02
4.47	Lactotransferrin	LTF;HEL110	E7EQB2	9.42
4.39	Myeloblastin	PRTN3	P24158	7.49
4.18	Neutrophil gelatinase-associated lipocalin	NGAL;LCN2	B2ZDQ1	13.57
3.54	Neutrophil elastase	ELANE;ELA2	P08246	7.48
3.51	Azurocidin	AZU1	P20160	8.36
3.49	Cysteine-rich secretory protein 3	CRISP3	J3KPA1	6.44
3.22	Myeloperoxidase	MPO	P05164-2	9.05
3.20	Bactericidal permeability-increasing protein	BPI	A2NX48	4.31
3.18	60S ribosomal protein L14	RPL14	P50914	1.62
3.17	Bactericidal permeability-increasing protein	BPI	P17213	8.26
3.16	Lysozyme	LYZ	B2R4C5	8.50
3.12	Marginal zone B- and B1-cell-specific protein	MZB1	Q8WU39	14.01
3.07	Ig heavy chain V-I region EU	IGHV1-69-2	P01742	3.98
3.01	Ficolin-1	FCN1	O00602	6.65
Proteins of decreased abundance in UC compared to healthy controls				
0.12	Hydroxymethylglutaryl-CoA synthase, mitochondrial	HMGCS2	A0A140VJL2	14.77
0.17	Fatty acid-binding protein, liver	FABP1	Q6FGL7	14.41
0.19	3-keto-steroid reductase	HSD17B7;hCG_2024989	P56937-2	14.32
0.24	Sulfate transporter	SLC26A2	P50443	18.52
0.26	Tumor necrosis factor alpha-induced protein 8-like protein 3	TNFAIP8L3	Q5GJ75	5.42
0.29	Carbonic anhydrase 2	HEL-76;CA2	V9HW21	12.09
0.30	Selenium-binding protein 1	HEL-S-134P;SELENBP1	V9HWG1	15.36
0.33	cAMP-dependent protein kinase inhibitor beta	PKIB	Q5T0Z6	7.57
0.34	Chloride anion exchanger	SLC26A3;DKFZp686P10213	P40879	10.33
0.35	UDP-glucuronosyltransferase 2A3	UGT2A3	Q6UWM9	14.95
0.36	Chymotrypsin-C	CTRC	Q99895	5.37
0.36	Hyaluronan and proteoglycan link protein 3	HAPLN3	A0A024RC58	8.55
0.37	Calcium-activated chloride channel regulator 1	CLCA1	A8K7I4	6.68
0.37	Tissue alpha-L-fucosidase	FUCA1	P04066	10.42
0.37	Phospholipase A(2)	PLA2G1B	F8W062	2.71
0.38	UDP-glucuronosyltransferase 2B17	UGT2B17;UGT2B15	O75795	3.10
0.38	Aquaporin-8	AQP8	Q53GF6	7.86
0.38	Mimecan	OGN	Q7Z532	9.66
0.39	Cadherin-17	CDH17	Q12864	17.16
0.39	Alcohol dehydrogenase 1C	ADH1C	P00326	10.05

Included are only significantly different proteins as determined in the two-sample test comparing the groups UC patients and healthy controls (see Additional file 1).
Only the first majority protein ID for each identification is given, for further IDs see Additional file 2

separate grouping of UC and healthy samples (Fig. 1a). In addition averaged profiles showed high Pearson correlations (> 0.7) for comparisons within the groups UC and healthy, respectively, and low correlations (< 0.1) for comparisons of UC to healthy (Fig. 1b). Several multi scatter plots were investigated with random samples

from UC patients or healthy controls in each group, and all showed the same trend with high correlations internally in the UC and healthy sample groups. Upon unsupervised investigation by principal component analysis (PCA) (Fig. 1c) a strong separation of UC and healthy in component 1 which explains 39.9% of the variance was observed. To investigate possible biases we controlled for grouping by gender, and such grouping did not occur.

Comparison with previous proteomics studies on UC

The in total 8562 identified proteins provided the so far deepest view into the proteome of human colon mucosa and of UC. We report the abundance changes including p values for all quantified proteins (see Additional file 2: Excel file). 6818 of these proteins were quantified in at least 70% of the samples and were subject to enrichment analysis. Our data confirmed and extends previous findings, but also includes novel findings, such as the increased abundance of signal peptidase complex

proteins. We searched PubMed for the terms "ulcerative colitis proteomics" and reviewed the titles and abstracts to select the relevant studies. Additionally we included studies previously known to the authors, totaling to 7 studies for the comparison [3–9]. The present study was largely in agreement with those previous studies about the direction of abundance changes of differently abundant proteins. However, the magnitudes of the changes differed.

Functional analysis

Hierarchical clustering of the 596 most differently abundant proteins (8.7% of all proteins quantified in >70% of samples; s0 = 2, FDR = 0.01) revealed three major clusters (Fig. 2). The proteins which were higher abundant in UC divided into two clusters, a small cluster of 20 proteins with a more pronounced abundance change in a subgroup of samples, and a large cluster of 247 proteins. The lower abundant proteins formed one cluster of 313
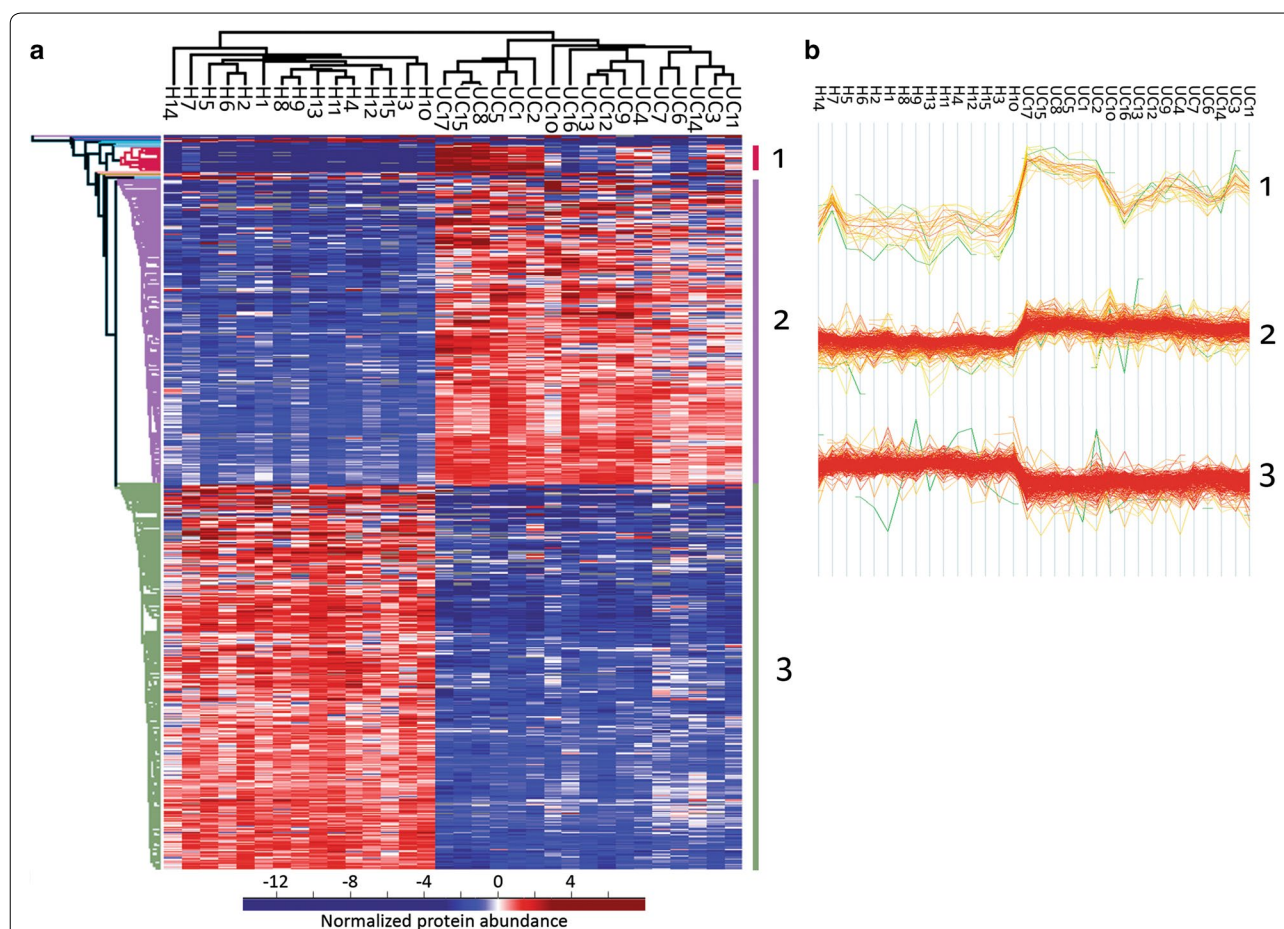
Fig. 2 Clustering of the 596 most differently abundant proteins between ulcerative colitis (UC) and healthy controls. **a** Hierarchical clustering reveals three main protein clusters. Low abundances are indicated in blue, high abundances in red. **b** Protein abundance profiles of the three main clusters. 1: minor cluster of 20 proteins (including calprotectin) with increased abundance in UC, with a subgroup of UC samples showing a more pronounced increase. 2: major cluster of 247 proteins with increased abundance in UC. 3: Cluster of 313 proteins with a decreased abundance in UC

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 7 of 13

proteins. In total 16 proteins were outside these three clusters.

The extraction of biological information from omics datasets is a major challenge. Enrichment analyses can overcome some of the difficulties [10]. Enrichment analyses annotate categorical terms (e.g. functions, localizations, pathways) to proteins and subsequently determine which terms are overrepresented among the differently abundant proteins compared to the entire dataset of quantified proteins. Knowledge about the involved pathways and functions can help to understand the disease, understand complications, and develop new treatment options.

Abundance changes found in tissue proteomics could result from differential regulation of the protein abundances in cells, but also from changes of the cell population (e.g. from the migration of immune cells into inflamed tissue in UC [11]) or the extracellular matrix. It seems plausible that each of these factors affects all measured abundances in such an experiment to varying degrees.

Lists of the enriched gene ontology terms among proteins which were found to be differently abundant in UC as compared to healthy tissue, and a list of the proteins which contributed to the functions discussed in the following are provided as Additional file 2: Excel file.

This article can neither cover all enriched terms we found, nor can it discuss the impact of each differently abundant protein on a function or the magnitude of its abundance change in detail. We therefore strongly encourage the reader to consult the above mentioned tables for further information on functions and proteins of interest.

Lower abundant in UC
321 proteins showed decreased abundances in UC as compared to healthy control samples. 313 of these grouped into one cluster (Fig. 2, cluster 3). Functional analysis revealed that metabolic pathways were overrepresented among the proteins with decreased abundance in UC. The enrichment of proteins with decreased abundances in our dataset affects short-chain and long-chain fatty acid, ketone body and amino acid (particularly tryptophan) metabolism as well as the tricarboxylic acid (TCA) cycle (see Additional file 1: SuppPathway1−4) and the electron transport chain (ETC; Fig. 3a).

The observed abundance decreases of butyrate utilizing proteins are in line with previous studies which show a decreased butyrate metabolism by UC tissue. Butyrate is a short-chain fatty acid (SCFA) and the main energy source of healthy colon mucosa [31]. It was previously hypothesized that ulcerative colitis is an energy-deficiency disease resulting from the failure to utilize
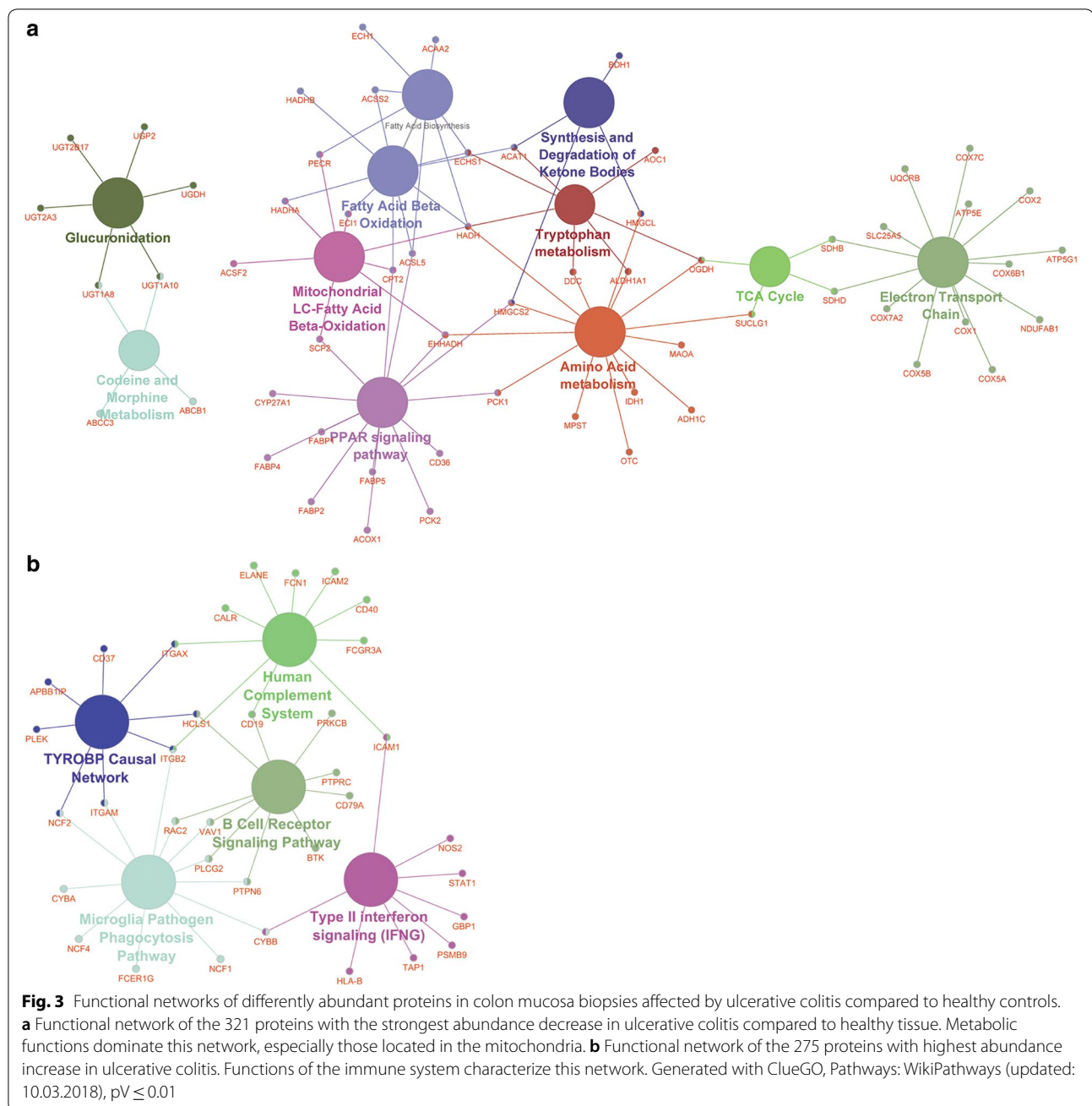
butyrate [31]. Colonocytes convert a part of the butyrate to ketone bodies [32]. Our data showed that proteins related to this specific utilization of butyrate were lower abundant in UC. Several studies have investigated whether butyrate has a positive effect on inflamed colon tissue, however the results of these studies are conflicting [33].

The decreased protein abundances also affected the transport and metabolism of the long-chain fatty acids (LCFA), i.e. the long carbon chain analogs of SCFAs. LCFAs have an ambiguous role in ulcerative colitis. Linoleic acid increases the chances of developing ulcerative colitis [34]. Omega-3 fatty acids, such as docosahexaenoic acid (DHA), may reduce the chance of developing the disease [35]. Studies on the efficacy of omega-3 fatty acids in the treatment of ulcerative colitis show controversial results. Systematic reviews however conclude that large high-quality studies fail to show positive effects, and that there is no sufficient evidence for the efficacy of omega-3 fatty acids in the treatment of UC [36, 37].

Mitochondrial enzymes involved in the metabolism of SCFA and LCFA as well as the tricarboxylic acid (TCA) and the electron transport chain (ETC) were found in our dataset to be low abundant in UC. These findings reflect a mitochondrial dysfunction which possibly contributes to the proposed energy deficiency [38].

Proteins related to nuclear receptors were of decreased abundance in UC. Peroxisome proliferator-activated receptor (PPAR)-inducible proteins were lower abundant in UC (Fig. 3). The changes in the LCFA transport and the PPAR pathway are interdependent. Activation of the intracellular PPAR receptors by LCFA demands the import of the fatty acids into the cell and the transport within the cell. One protein involved in the transport into the cell is CD36 which is low abundant in UC tissue. Fatty acid binding proteins (FABP) facilitate the transport within the cell [39]. Four FABPs (FABP1, FABP2, FABP4, and FABP5) are among the proteins with the lowest relative abundance in UC tissue compared to healthy tissue. A decreased transport of LCFA might in turn be partially causative for the decreased abundances of PPAR-inducible proteins, such as ACOX1, HMGCS2, FABP1, and PCK1 [40]. In accordance with a previous study [41], we found the vitamin D3 receptor NR1i1 to be lower abundant in UC tissue. However, we observed no general abundance changes of proteins regulated by this receptor. Previous studies found decreased abundances of hepatocyte nuclear factor 4 alpha (HNF4A) in UC and suggested a role in the disease pathophysiology [42]. Our data supports these findings.

All detected metallothioneins—MT1H, MT2A, and MT1F—were found to be lower abundant in UC tissue compared to healthy controls. Metallothioneins are

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 8 of 13

Fig. 3 Functional networks of differently abundant proteins in colon mucosa biopsies affected by ulcerative colitis compared to healthy controls. **a** Functional network of the 321 proteins with the strongest abundance decrease in ulcerative colitis compared to healthy tissue. Metabolic functions dominate this network, especially those located in the mitochondria. **b** Functional network of the 275 proteins with highest abundance increase in ulcerative colitis. Functions of the immune system characterize this network. Generated with ClueGO, Pathways: WikiPathways (updated: 10.03.2018), pV $\leq$ 0.01

small, cysteine-rich proteins, which bind heavy metal cations. Our findings confirmed several previous studies, which concluded that the decreased metallothionein abundances may result in an inefficient antioxidant response in the mucosa and in turn contributes to the IBD pathophysiology [43]. However, other studies present conflicting results and conclusions [44].

Histamine is the major mediator of mast cells and contributes to the immuno-inflammatory reaction in IBD. Previous studies showed increased histamine secretion and levels in UC affected tissues [45] and decreased histaminase (AOC1) activity [9]. The present study revealed that both major histamine degrading enzymes—Histaminase (AOC1) and histamine *N*-methyltransferase (HNMT)—were lower abundant in UC tissue compared to healthy tissue (not based on enrichment analysis). This finding constitutes a plausible causative explanation for some of the previously reported findings on histamine in UC.

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 9 of 13

Chemical compounds undergo phase I and phase II biotransformation reactions which make them more hydrophilic [46]. This allows their excretion with urine because the polar compounds are not reabsorbed in the kidneys [47]. Phase I reactions introduce or change functional groups by oxidation, reduction, or hydrolysis. In phase II reactions, enzymes attach hydrophilic endogenous compounds [46]. Our data showed that the abundances of a large proportion of both phase I and phase II enzymes are decreased in UC. A decrease of xenobiotic metabolism gene products in UC was shown earlier on mRNA level [48]. The affected phase II enzymes are involved in sulfonation, glucuronidation, and glutathione conjugation. Proteins related to both phase 1 (monoamine oxidase A, MAOA) and phase 2 metabolism (sulfotransferases) of catecholamines were found in our dataset to be less abundant in UC. This could contribute to previously found increases of norepinephrine in rectal UC mucosa [49].

Synthetic steroids are an important treatment option for ulcerative colitis. Our data showed impairments in the steroid metabolism in UC. This suggests a role of endogenous steroids in the pathophysiology of the disease. The hydroxysteroid dehydrogenases HSD11B2, HSD17B11, HSD17B2, and HSD17B7 were among the most down-regulated proteins in UC (HSD17B8 is downregulated to a lesser degree). The abundances of HSD17B4, HSD17B10, and HSD17B12 were not different in UC from healthy tissue. Decreased abundances of HSD11B2 [50] and HSD17B7 [3] have been reported previously. Our data furthermore showed that proteins related to the transport of bile acids are enriched among the lower abundant proteins, which supports findings of a previous study [41]. Pathway mapping shows abundance decreases of proteins involved in the neutral pathway of bile acids biosynthesis (Additional file 1: SuppPathway5). However, the implications of this finding are not clear, because bile acids are primarily synthesized in hepatocytes, whereas there is no evidence of bile acid synthesis in enterocytes [51].

The carbonic anhydrases CA1, CA2, CA4, and CA12 were found to be lower abundant in UC. CA3 was the only quantified but not significantly lower abundant carbonic anhydrase. A decreased abundance in UC was reported earlier for CA1 [52].

Fibrillar collagens are a subgroup of collagens and provide three-dimensional frameworks for tissues and organs [53]. The fibrillar collagens were enriched among the proteins that were of decreased abundance in UC. Proteins that are involved in collagen degradation, e.g. metalloproteinases, on the other hand were enriched among the proteins with increased abundance in UC (see below). This finding is in accordance with a previous study on the role of collagen degradation in IBD [54].

Higher abundant proteins in UC

As expected, many proteins with increased abundance in UC are involved in inflammatory and immune processes (Fig. 3b). Proteins related to neutrophils (e.g. NADPH oxidase complex proteins which generate superoxide, metalloproteinases) and B cells (e.g. V and C regions of immunoglobulins) were highly enriched. This could result from a tissue infiltration by these immune cells.

The higher abundant proteins divided into two main clusters under hierarchical clustering (Fig. 2, clusters 1 and 2). The UC patients with high cluster 1 protein abundances (UC1, UC2, UC5, UC8, UC15, UC17) grouped together (Fig. 2a, b). Cluster 1 contained proteins occurring in neutrophils, among these Protein S100A8 and Protein S100A9 which together form the calprotectin complex (see Additional file 1: Table 3). Several proteins (calprotectin, S100A12, lactotransferrin) of the minor cluster have been used to differentiate between IBD and non-IBD [55]. Interestingly, the average Geboes index in the patient cluster with high cluster 1 protein abundances is 11.0 (SD = 3.1), while it is 6.5 (SD = 3.2) for the other patients.

Our data showed that inducible nitric oxide synthase (NOS2) is higher abundant in UC tissue (not based on enrichment analysis). NOS2 produces nitric oxide (NO) which has various physiological roles as a messenger molecule. It functions most prominently as a vasodilator [56], but it is also involved in the immune system [57]. NO is directly toxic to pathogens, induces or suppresses apoptosis, and regulates the immune reaction [58]. Excess NO is cytotoxic and induces cell death. NOS2 is inducible by inflammation, infection and other stimuli. Increased NOS2 abundances in UC were shown earlier by immunostaining [8] and our results confirm these findings.

Proteins with functions in protein processing in the endoplasmic reticulum (ER) are overrepresented among the proteins which were more abundant in UC. Also unfolded protein response (UPR) and signal peptidase complex (SPC) proteins are related to protein processing, located in the ER, and enriched among proteins with increased abundances in UC. UPR has previously been identified as one factor in the pathophysiology of ulcerative colitis [59]. The enrichment of SPC proteins among proteins with increased abundances in UC is a novel finding from our data.

However, ER proteins were not generally higher abundant. For instance, many metabolic enzymes are located in the ER and enriched among the proteins with decreased abundances (e.g. UDP-glucuronosyltransferases, see "Lower abundant in UC" section).

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 10 of 13

The SPC is not inherently inflammatory. Nonetheless its proteins were among those with the highest abundance increases in UC. The SPC is located in the membrane of the endoplasmic reticulum and cleaves the signal peptide cotranslationally from nascent proteins [60]. The SPC subunits SPCS1, SPCS2, and SPCS3 and the SPC catalytic subunit SEC11C were more abundant in UC. They furthermore showed highly similar abundance profiles (Additional file 1: Figure 3). The SPC catalytic subunit SEC11 however does not follow this pattern and was similarly abundant in UC and healthy tissue.

Accumulation of unfolded and misfolded proteins causes stress to the endoplasmic reticulum. The UPR aims to remove these proteins. To achieve this, the cell stops the translation of further proteins, degrades misfolded proteins, and produces chaperones, which correct the protein folding. Our data showed strong enrichment of proteins related to UPR, response to ER stress, and the ER chaperone complex among the higher abundant proteins in UC. Specifically, several chaperones that can be induced by ATF6 alpha (HSP90B1, calreticulin [61], and HSPA5 [62]), which plays a central role in UPR, were enriched. Our findings support previous assumptions that the UPR may play a role in the pathophysiology of UC.

Foreseeable applications
The functional changes we observed in the present study allow for hypothesis generation for treatment approaches.

Alterations of metabolic pathways in ulcerative colitis imply that nutritional interventions directed at the respective metabolites could be effective in the treatment of the disease. For instance, decreased metallothionein abundances indicate that the detoxification of heavy metals and the utilization of polyvalent metal ions could be impaired. Considering the decreased detoxification capacities, reduced intake or chelation of toxic heavy metals might have beneficial effects. On the other hand, the supplementation of essential polyvalent metal ions could be advantageous. Furthermore, the decreased metabolic and transport capabilities related to butyrate contribute probably to the energy deprivation of colonocytes. Previous studies investigated the treatment of UC with butyrate and other SCFA. But the pathophysiological changes we observed related to the utilization of those compounds imply that supplementation with a later metabolite such as beta-hydroxybutyrate might be preferable. Beta-hydroxybutyrate is a ketone body that becomes systemically available after oral administration and has been used in the treatment of other pathophysiological conditions [63, 64].

The enrichment of proteins related to signaling pathways such as PPAR designates the respective pathways as potential drug targets.

Abundance changes of single proteins may have similar implications, especially if they are known to be drugable. Proteins with high abundance differences between UC patients and healthy controls and high corresponding −Log p values are of potential interest as biomarker candidates. Proteins with these characteristics are for instance SLC26A2, HMGCS2, and CD38 (Additional file 1: Figure 4; Additional file 2).

However, further testing, clinical studies, and validation will be necessary before any clinical application.

Conclusions
Our study presents new evidence and complements previous findings about changes of biological functions in UC. The major novel findings of this study were the increased abundances of SPC proteins and the presence of two distinguished clusters of higher abundant proteins, with the calprotectin complex proteins S100A8 and S100A9 in the minor cluster 1. This cluster of 20 proteins diverged from the abundance changes of the majority of proteins with increased abundances in UC that we observed in cluster 2.

In conclusion, abundances in UC tissue were increased compared to healthy controls for proteins related to the immune system and to protein processing in the ER, e.g. UPR and SPC proteins. Immune cell infiltration into the inflamed tissue contributes probably to the increased immune system protein abundances. The high abundance of UPR proteins indicates an ER stress response. Moreover, NOS2 which produces NO was higher abundant in UC tissue.

Lower abundant in UC were predominantly metabolic proteins. These metabolic proteins comprise mitochondrial enzymes for the metabolization of SCFA and LCFA as well as proteins of the TCA cycle and the ETC. These changes are probably contributing to the previously reported mitochondrial dysfunction and energy deficiency of colonocytes in UC. The abundances of LCFA transport proteins were decreased which may be partially causative for the decreased abundances of PPAR-inducible proteins. Decreased abundances of histamine degrading enzymes probably contribute to the increased histamine levels in UC tissue which were found in previous studies. Furthermore, abundances of proteins involved in phase I and phase II biotransformation were decreased, as well as the abundances of several metallothioneins, hydroxysteroid dehydrogenases, and carbonic anhydrases.

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 11 of 13

We provide a list of all 8562 identified proteins including the fold changes UC/H and p values as supporting information (see Additional file 2: Excel file).

Additional files

Additional file 1. Detailed method descriptions, additional figures and tables, mapping of metabolic pathways.

Additional file 2. Complete list of identified proteins including UC/Healthy ratios and p-values, lists of gene ontology term enrichments, enriched functions discussed in the article and their proteins.

Abbreviations

BCA: bicinchoninic acid; DHA: docosahexaenoic acid; ER: endoplasmic reticulum; ETC: electron transport chain; FA: formic acid; FABP: fatty acid binding protein; IBD: inflammatory bowel disease; LCFA: long-chain fatty acid; Lys-C: endoproteinase Lys-C; NO: nitric oxide; PCA: principal component analysis; PPAR: peroxisome proliferator-activated receptor; SCFA: short-chain fatty acid; SDC: sodium deoxycholate; SPC: signal peptidase complex; TCA cycle: tricarboxylic acid cycle; TEAB: triethylammonium bicarbonate buffer; TMT: tandem mass tag; UC: ulcerative colitis; UPR: unfolded protein response.

Authors' contributions

AS, YP and TH planned the experiments. AS and TH analyzed and interpreted the proteome data with input from YP. SWS performed the histological examination to determine the Geboes scores. JF and RG recruited patients, provided colon mucosa biopsies, and reported the baseline characteristics. All authors contributed with writing to their parts as well as approved the final version of the manuscript.

Author details

[1] Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway. [2] Department of Medical Gastroenterology, University Hospital of North Norway, Tromsø, Norway. [3] Gastroenterology and Nutrition Research Group, Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway. [4] Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway. [5] Department of Clinical Pathology, University Hospital of North Norway, Tromsø, Norway.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The Regional Committee of Medical Ethics of North Norway and the Norwegian Social Science Data Services approved the study and the storage of biological material under the number REK NORD 2012/1349. In addition, all enrolled subjects have signed an informed and written consent form.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, Benchimol EI, Panaccione R, Ghosh S, Barkema HW, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time: based on systematic review. Gastroenterology. 2012;142(1):46.e42–54.e42.
2. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389(10080):1756–70.
3. Schniers A, Anderssen E, Fenton Christopher G, Goll R, Pasing Y, Paulssen Ruth H, Florholmen J, Hansen T. 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.
4. Bennike TB, Carlsen TG, Ellingsen T, Bonderup OK, Glerup H, Bøgsted M, Christiansen G, Birkelund S, Stensballe A, Andersen V. Neutrophil extracellular traps in ulcerative colitis: a proteome analysis of intestinal biopsies. Inflamm Bowel Dis. 2015;21(9):2052–67.
5. Starr AE, Deeke SA, Ning Z, Chiang C-K, Zhang X, Mottawea W, Singleton R, Benchimol EI, Wen M, Mack DR, et al. Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate Crohn's disease from UC. Gut. 2016;66:1573–83.
6. Poulsen NA, Andersen V, Møller JC, Møller HS, Jessen F, Purup S, Larsen LB. Comparative analysis of inflamed and non-inflamed colon biopsies reveals strong proteomic inflammation profile in patients with ulcerative colitis. BMC Gastroenterol. 2012;12(1):76.
7. Deeke SA, Starr AE, Ning Z, Ahmadi S, Zhang X, Mayne J, Chiang C-K, Singleton R, Benchimol EI, Mack DR, et al. Mucosal–luminal interface proteomics reveals biomarkers of pediatric inflammatory bowel disease-associated colitis. Am J Gastroenterol. 2018;113(5):713–24.
8. Leonard N, Bishop AE, Polak JM, Talbot IC. Expression of nitric oxide synthase in inflammatory bowel disease is not affected by corticosteroid treatment. J Clin Pathol. 1998;51(10):750–3.
9. Mennigen R, Kusche J, Streffer C, Krakamp B. Diamine oxidase activities in the large bowel mucosa of ulcerative colitis patients. Agents Actions. 1990;30(1–2):264–6.
10. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.
11. Taman H, Fenton CG, Hensel IV, Anderssen E, Florholmen J, Paulssen RH. Transcriptomic landscape of treatment—Naïve ulcerative colitis. J Crohn's Colitis. 2018;12(3):327–36.
12. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett. 2009;583(24):3966–73.
13. Stange EF, Travis SP, Vermeire S, Reinisch W, Geboes K, Barakauskiene A, Feakins R, Flejou JF, Herfarth H, Hommes DW, et al. European evidence-based Consensus on the diagnosis and management of ulcerative colitis: definitions and diagnosis. J Crohns Colitis. 2008;2(1):1–23.
14. Sutherland LR, Martin F, Greer S, Robinson M, Greenberger N, Saibil F, Martin T, Sparr J, Prokipchuk E, Borgen L. 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. Gastroenterology. 1987;92(6):1894–8.
15. Olsen T, Goll R, Cui G, Husebekk A, Vonen B, Birketvedt G, Florholmen J. Tissue levels of tumor necrosis factor-alpha correlates with grade of inflammation in untreated ulcerative colitis. Scand J Gastroenterol. 2007;42(11):1312–20.
16. Scopes RK. Measurement of protein by spectrophotometry at 205 nm. Anal Biochem. 1974;59(1):277–82.
17. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–72.

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 12 of 13

18. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.

19. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731–40.

20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

21. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-into decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics (Oxford, England). 2009;25(8):1091–3.

22. Tyanova S, Cox J. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. Methods Mol Biol (Clifton, NJ). 2018;1711:133–48.

23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25(1):25–9.

24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005;102(43):15545.

26. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017;45(D1):D190–9.

27. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, et al. The Reactome pathway knowledge-base. Nucleic Acids Res. 2018;46(D1):D649–55.

28. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE. Nucleic Acids Res. 2013;41(D1):D344–7.

29. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. 2018;46(D1):D661–7.

30. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PDJGB. Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. 2004;6(1):R2.

31. Roediger WEW. The colonic epithelium in ulcerative colitis: an energy-deficiency disease? Lancet. 1980;316(8197):712–5.

32. Roediger WE. Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. Gut. 1980;21(9):793–8.

33. Vinolo MAR, Rodrigues HG, Nachbar RT, Curi R. Regulation of inflammation by short chain fatty acids. Nutrients. 2011;3(10):858–76.

34. Tjonneland A, Overvad K, Bergmann MM, Nagel G, Linseisen J, Hallmans G, Palmqvist R, Sjodin H, Hagglund G, Berglund G, et al. Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the aetiology of ulcerative colitis: a nested case-control study within a European prospective cohort study. Gut. 2009;58(12):1606–11.

35. John S, Luben R, Shrestha SS, Welch A, Khaw K-T, Hart AR. Dietary n-3 polyunsaturated fatty acids and the aetiology of ulcerative colitis: a UK prospective cohort study. Eur J Gastroenterol Hepatol. 2010;22(5):602–6.

36. Ungaro F, Rubbino F, Danese S, D'Alessio S. Actors and factors in the resolution of intestinal inflammation: lipid mediators as a new approach to therapy in inflammatory bowel diseases. Front Immunol. 2017;8:1331.

37. Turner D, Shah PS, Steinhart AH, Zlotkin S, Griffiths AM. Maintenance of remission in inflammatory bowel disease using omega-3 fatty acids (fish oil): a systematic review and meta-analyses. Inflamm Bowel Dis. 2011;17(1):336–45.

38. Novak EA, Mollen KP. Mitochondrial dysfunction in inflammatory bowel disease. Front Cell Dev Biol. 2015;3:62.

39. Georgiadi A, Kersten S. Mechanisms of gene regulation by fatty acids. Adv Nutr. 2012;3(2):127–34.

40. Rakhshandehroo M, Hooiveld G, Müller M, Kersten S. Comparative analysis of gene regulation by the transcription factor PPARα between mouse and human. PLoS ONE. 2009;4(8):e6796.

41. Jahnel J, Fickert P, Hauer AC, Hogenauer C, Avian A, Trauner M. Inflammatory bowel disease alters intestinal bile acid transporter expression. Drug Metab Dispos Biol Fate Chem. 2014;42(9):1423–31.

42. Ahn SH, Shah YM, Inoue J, Morimura K, Kim I, Yim S, Lambert G, Kurotani R, Nagashima K, Gonzalez FJ, et al. Hepatocyte nuclear factor 4alpha in the intestinal epithelial cells protects against inflammatory bowel disease. Inflamm Bowel Dis. 2008;14(7):908–20.

43. Kruidenier L, Kuiper I, van Duijn W, Mieremet-Ooms MAC, van Hogezand RA, Lamers CBHW, Verspaget HW. Imbalanced secondary mucosal antioxidant response in inflammatory bowel disease. J Pathol. 2003;201:17–27.

44. Rahman MT, Karim MM. Metallothionein: a potential link in the regulation of zinc in nutritional immunity. Biol Trace Elem Res. 2018;182(1):1–13.

45. Raithel M, Matek M, Baenkler HW, Jorde W, Hahn EG. Mucosal histamine content and histamine secretion in Crohn's disease, ulcerative colitis and allergic enteropathy. Int Arch Allergy Immunol. 1995;108(2):127–33.

46. Omiecinski CJ, Vanden Heuvel JP, Perdew GH, Peters JM. Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. Toxicol Sci. 2011;120(1):S49–75.

47. Bardal SK, Waechter JE, Martin DS. Chapter 2—Pharmacokinetics. In: Bardal SK, Waechter JE, Martin DS, editors. Applied pharmacology. Philadelphia: Content Repository Only; 2011. p. 17–34.

48. Langmann T, Moehle C, Mauerer R, Scharl M, Liebisch G, Zahn A, Stremmel W, Schmitz G. Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. Gastroenterology. 2004;127(1):26–40.

49. Penttila O, Kyosola K, Klinge E, Ahonen A, Tallqvist G. Studies of rectal mucosal catecholamines in ulcerative colitis. Ann Clin Res. 1975;7(1):32–6.

50. Takahashi K-I, Fukushima K, Sasano H, Sasaki I, Matsuno S, Krozowski ZS, Nagura H. Type II 11β-hydroxysteroid dehydrogenase expression in human colonic epithelial cells of inflammatory bowel disease. Dig Dis Sci. 1999;44(12):2516–22.

51. Dawson PA, Karpen SJ. Intestinal transport and metabolism of bile acids. J Lipid Res. 2015;56(6):1085–99.

52. Giovanni RF, Renzo L, Giuseppe C, Adriana F, Sambuy MY. Carbonic anhydrase I reduction in colonic mucosa of patients with active ulcerative colitis. Dig Dis Sci. 1998;43(9):2086–92.

53. Bella J, Hulmes DJS. Fibrillar collagens. In: Parry DAD, Squire JM, editors. Fibrous proteins: structures and mechanisms. Cham: Springer; 2017. p. 457–90.

54. Koelink PJ, Overbeek SA, Braber S, Morgan ME, Henricks PAJ, Roda MA, Verspaget HW, Wolfkamp SC, te Velde AA, Jones CW, et al. Collagen degradation and neutrophilic infiltration: a vicious circle in inflammatory bowel disease. Gut. 2014;63(4):578–87.

55. Iskandar HN, Ciorba MA. Biomarkers in inflammatory bowel disease: current practices and recent advances. Transl Res. 2012;159(4):313–25.

56. Chen K, Pittman RN, Popel AS. Nitric oxide in the vasculature: where does it come from and where does it go? A quantitative perspective. Antioxiod Redox Signal. 2008;10(7):1185–98.

57. Bogdan C. Nitric oxide and the immune response. Nat Immunol. 2001;2:907.

58. Coleman JW. Nitric oxide in immunity and inflammation. Int Immunopharmacol. 2001;1(8):1397–406.

59. Cao SS. Epithelial ER stress in Crohn's disease and ulcerative colitis. Inflamm Bowel Dis. 2016;22(4):984–93.

60. Sakaguchi M. Eukaryotic protein secretion. Curr Opin Biotechnol. 1997;8(5):595–601.

61. Fox RM, Andrew DJ. Transcriptional regulation of secretory capacity by bZip transcription factors. Front Biol. 2015;10(1):28–51.

Schniers *et al. Clin Proteom*   (2019) 16:4

Page 13 of 13

62. Thuerauf DJ, Morrison L, Glembotski CC. Opposing roles for ATF6α and ATF6β in endoplasmic reticulum stress response gene induction. J Biol Chem. 2004;279(20):21078–84.

63. Tieu K, Perier C, Caspersen C, Teismann P, Wu D-C, Yan S-D, Naini A, Vila M, Jackson-Lewis V, Ramasamy R, et al. D-β-Hydroxybutyrate rescues mitochondrial respiration and mitigates features of Parkinson disease. J Clin Investig. 2003;112(6):892–901.

64. Plecko B, Stoeckler-Ipsiroglu S, Schober E, Harrer G, Mlynarik V, Gruber S, Moser E, Moeslinger D, Silgoner H, Ipsiroglu O. Oral beta-hydroxybutyrate supplementation in two patients with hyperinsulinemic hypoglycemia: monitoring of beta-hydroxybutyrate levels in blood and cerebrospinal fluid, and in the brain by in vivo magnetic resonance spectroscopy. Pediatr Res. 2002;52(2):301–6.

65. Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. Gut. 2006;55(6):749–53.

Supplementary Data Content 1

Supplementary texts

LC-MS/MS

Materials

LC-MS/MS was performed with an EASY-nLC 1000 coupled to a Q Exactive (both Thermo Fisher Scientific, Bremen, Germany).
The LC was equipped with an Acclaim PepMap® 100 precolumn, C18, particle size 3 μm, pore size 100 Å, inner diameter 75 μm, length 2 cm, nanoViper,(Thermo Fisher Scientific, Bremen, Germany). The separation column was an EASY-Spray, PepMAP® RSLC, C18, particle size 2 μm, pore size 100 Å, inner diameter 75 μm, length 50 cm (Thermo Fisher Scientific, Bremen, Germany).
The solvents for separation were water with 0.1% FA, and acetonitrile with 0.1% FA, respectively.

LC settings

The peptides were concentrated on a reversed-phase trap column with 0.1% formic acid (FA) at a flow rate of 20 µl/min. The peptides were separated on a reversed-phase main column with a binary solvent gradient. The column temperature was set to 60 °C and the flow rate to 200 nl/min. The ACN proportion was increased from 2% to 5% over 19 min, further to 30% at 180 min and to 100% at 200 min. The column was regenerated with 100% ACN for additional 10 min.

MS/MS settings

The Q Exactive mass spectrometer was run in positive mode with the following settings: Chromatographic peak width 15 s, default charge state 2, full MS survey scans from 400 to 2,000 m/z, resolution 70,000, AGC target value 3e6, maximum injection time 100 ms for MS scans. Subject the 10 most intense peaks to MS/MS with the following settings: resolution 17,500, dynamic exclusion 10 s, underfill ratio 1%, charge states +2, +3 and +4, exclude isotopes, normalized collision energy 31, isolation window 2 m/z, AGC target value 1e5 and maximum injection time 50 ms, fixed first mass 120.

MaxQuant settings

The mass spectrometry data was searched in MaxQuant version 1.6.1.0 against a fasta file from https://www.uniprot.org/ (download date: 24.11.2017). This file included all human proteins (canonical and isoforms). The TMT reporter ion distribution was configured according to the manufacturer's Certificate of Analysis. The quantification method was set to "Reporter ion MS2/TMTsixplex". File names and fraction numbers were set according to the experimental setting. The following further parameters were applied:
Enzyme: Trypsin/P (specific), fixed modifications: Carbamidomethyl (C), variable modifications: Oxidation (M) and Acetyl (Protein N-term), max. 2 missed cleavages, PSM FDR: 0.01, Protein FDR: 0.01 Site FDR: 0.01, Use Normalized Ratios For Occupancy: TRUE, Min. peptide Length: 7, Min. score for unmodified peptides: 0, Min. score for modified peptides: 40, Min. delta score for unmodified peptides: 0, Min. delta score for modified peptides: 6, Min. unique peptides: 0, Min. razor peptides: 1, Min. peptides: 1, Use only unmodified peptides and: TRUE, Modifications included in protein quantification: Oxidation (M) and Acetyl (Protein N-term), Peptides used for protein quantification: Razor, Discard unmodified counterpart peptides: TRUE, Label min. ratio count: 2, Use delta score:

FALSE, iBAQ: FALSE, iBAQ log fit: FALSE, Match between runs: TRUE, Matching time window [min]: 0.7, Alignment time window [min]: 20, Find dependent peptides: FALSE, Decoy mode: revert, Include contaminants: TRUE, Advanced ratios: TRUE, Second peptides: FALSE, Calculate peak properties: FALSE, Main search max. combinations: 200, Advanced site intensities: TRUE, Max. peptide mass [Da]: 4600, Min. peptide length for unspecific search: 8, Max. peptide length for unspecific search: 25, Razor protein FDR: TRUE, Disable MD5: FALSE, Max mods in site table: 3, Match unidentified features: FALSE, MS/MS tol. (FTMS): 20 ppm, Top MS/MS peaks per Da interval. (FTMS): 12, Da interval. (FTMS): 100, MS/MS deisotoping (FTMS): TRUE, MS/MS deisotoping tolerance (FTMS): 7, MS/MS deisotoping tolerance unit (FTMS): ppm, MS/MS higher charges (FTMS): TRUE, MS/MS water loss (FTMS): TRUE, MS/MS ammonia loss (FTMS): TRUE, MS/MS dependent losses (FTMS): TRUE, MS/MS recalibration (FTMS): FALSE.

Perseus

Data normalization

The proteinGroups file generated by MaxQuant was loaded into Perseus version 1.6.1.1. The "Reporter Intensity Corrected" columns were loaded as main columns. These intensities were log(2) transformed. The respective standard intensities were then subtracted. The intensity columns were renamed according to the respective sample name and categorical annotations were added (UC and H, respectively). Rows were filtered to remove proteins labelled as "Only identified by site", "Reverse" and "Potential contaminant". The intensities were then Z-score normalized (matrix access: column).

The Z-score normalized values were used for all analyses, except the determination of the UC/H ratio.

Creation of lists and figures

Figures

For figure 1A the rows were filtered to have at least 70% values and column correlations were calculated for the normalized protein intensities with standard settings. The results were visualized by hierarchical clustering. The replicates were only included for this figure and were removed for all other analyses.

For figure 1B, the Z-score transformed dataset was filtered for 70% valid values. The intensities of each protein for UC2, 3, 4, 6, 7, 11, 13, 14, and 16 (UC a), UC1, 5, 8, 9, 10, 12, 15, and 17 (UC b), H3, 4, 5, 6, 7, 8, 10, and 11 (H a), and H1, 2, 9, 12, 13, 14, and 15 (H b), respectively, were averaged. This was done by assigning the respective groups to the samples, followed by the function "Annot. rows → Average groups, Average type: mean". The averaged intensities were visualized in a multi scatter plot.

For figure 1C, the dataset without replicates was filtered for 100% valid values before the PCA was generated.

Figure 2 was generated by filtering the normalized dataset without replicates for 70% valid values. A two-sample test was performed to determine significant differences between the UC samples and healthy controls with standard settings, except for s0 = 2 and FDR = 0.01. The dataset was then filtered for significantly changed proteins. The figure was prepared by hierarchical clustering of the resulting dataset, and setting the number of clusters to 17.

For figure 3A and B, the lower abundant proteins and higher abundant proteins, respectively, which were identified in the previous two-sample test were selected. Their gene identifiers were loaded into the Cytoscape app ClueGO and analyzed against WikiPathways (updated 07.02.2018), showing only pathways with pV ≤ 0.01. GO term grouping was activated for the higher abundant proteins, but not for the lower abundant proteins. The ClueGO Layout was selected and manually refined. All other parameters were at the standard setting.

Metabolic pathways (suppPathway1-5) were mapped at https://humancyc.org/overviewsWeb/celOv.shtml# (date: 01.01.2019) with Gene IDs and logaritmized UC/H ratios of all quantified proteins.

List of enriched GO terms

To generate the enrichment lists provided with the supporting information, annotations were added to the unfiltered matrix resulting from the two sample test (see preparation of Figure 2). Lists of more and less, respectively, abundant proteins were generated, based on the two-sample test. These lists were separately matched back to the matrix with the added annotations with the function "matching rows by name" and the option of an indicator in the new matrix. Fisher exact tests with the categorical column indicating the respective matrix of enriched proteins and standard settings gave the enrichment lists.

List of proteins in enriched terms

For the ratio UC/H, the normalized intensities **before Z-scoring** (i.e. after log(2) transformation and subtraction of the respective standard intensity) were used. The mean of the sample intensities (matrix acces: column) was subtracted. The difference between the normalized log(2) intensities of UC and H was calculated in a two-sample test. The resulting log(2) differences were transformed back by a 2^x operation which gives the ratio of the normalized protein intensities.
The UC/H ratio column was then matched with the function "Matching rows by name" to the column "majority protein" into a matrix of the 6818 proteins which were quantified in >70% of the samples, including two-sample test results (after Z-scoring) and GO annotations.
A volcano plot was created from this list. The rider "categories" allows the selection of all proteins in a given category by clicking on the category name in the list. The respective proteins were subsequently exported into a separate matrix (Rider: "Points" → "Export selection (reduce matrix)"), which was then exported into a tab delimited text file (Right-click → "Plain matrix export").

List of all proteins

The Z-score difference and the –log p-value were determined in a two-sample test from the normalized values after Z-scoring from all 8562 identified proteins. The corresponding Ratio UC/H was calculated and implemented as described in "List of proteins in enriched terms". The indicators for whether a difference is significant (as generated for figure 2) were matched into this matrix with the function "Matching rows by name/Matching column: Majority Proteins IDs".

Supplementary tables

Grade	Average	SD	Range
Architectural changes (0-3)	1.06	0.66	0-3
Chronic inflammatory infiltrate (0-3)	1.31	0.46	1-2
Eosinophils in lamina propria (0-3)	0.88	0.60	0-2
Neutrophils in lamina propria (0-3)	1.06	0.83	0-2
Neutrophils in epithelium (0-3)	1.69	0.68	1-3
Crypt destruction (0-3)	1.31	1.04	0-3
Erosions and ulcerations (0-4)	0.63	1.36	0-4
Total (0-22)	7.94	3.77	3-16

Suppl. table1: Averages, standard deviations, and ranges of Geboes index grades for the UC patients.

Starting protein concentration [µg/ml] in sample as determined in BCA assay	TMTsixplex for 25 µg peptides
≥ 2200	0.4 mg
1530 to 2200	0.5 mg
1173 to 1530	0.6 mg
952 to 1173	0.7 mg
800 to 952	0.8 mg

Suppl. table2: Used TMTsixplex amounts in dependence of starting concentration of samples. The TMTsixplex amount was adjusted for low concentrated samples, because poor labeling was observed otherwise during the method development.

| Ratio UC/H	-log p-value UC_K	Protein names	Gene names	Majority protein IDs
10.1177	10.5488	Kinesin-like protein	KIF26B	B7WPD9
6.7208	8.40544	Protein S100-A8	S100A8	P05109
6.69903	9.70578	Cathelicidin antimicrobial peptide	CAMP	J3KNB4
6.16675	8.07071	Protein S100-A9	S100A9	P06702
4.94808	8.35806	Protein S100-A12;Calcitermin	S100A12	P80511
4.72862	7.01854	Neutrophil defensin 1	DEFA1	P59665
4.47297	9.41786	Lactotransferrin	LTF;HEL110	E7EQB2
4.39275	7.48579	Myeloblastin	PRTN3	P24158
4.17913	13.5737	Neutrophil gelatinase-associated lipocalin	NGAL;LCN2	B2ZDQ1
3.5434	7.48418	Neutrophil elastase	ELANE;ELA2	P08246
3.50704	8.35501	Azurocidin	AZU1	P20160
3.49049	6.43954	Cysteine-rich secretory protein 3	CRISP3	J3KPA1
3.21954	9.05008	Myeloperoxidase	MPO	P05164-2
3.2025	4.30977		BPI	A2NX48
3.16036	8.49697	Lysozyme	LYZ	B2R4C5
3.01386	6.64713	Ficolin-1	FCN1	O00602
2.85388	6.53514	Resistin	RETN	Q9HD89
2.72326	6.35354			B7Z507
2.53399	6.33617	Matrix metalloproteinase-9	MMP9	P14780
2.43248	5.56569	Neutrophil collagenase	MMP8	P22894

Suppl. table3: Proteins from the minor cluster of the proteins which are more abundant in UC (cluster 1, see article figure 2). Given for each protein is only the first Majority protein ID states (see Excel file, Supplementary Data Content 2, which lists all 8562 identified protein, for further IDs and data on these proteins (labeled there with "Cluster 1 (minor cluster increased abundance)")).
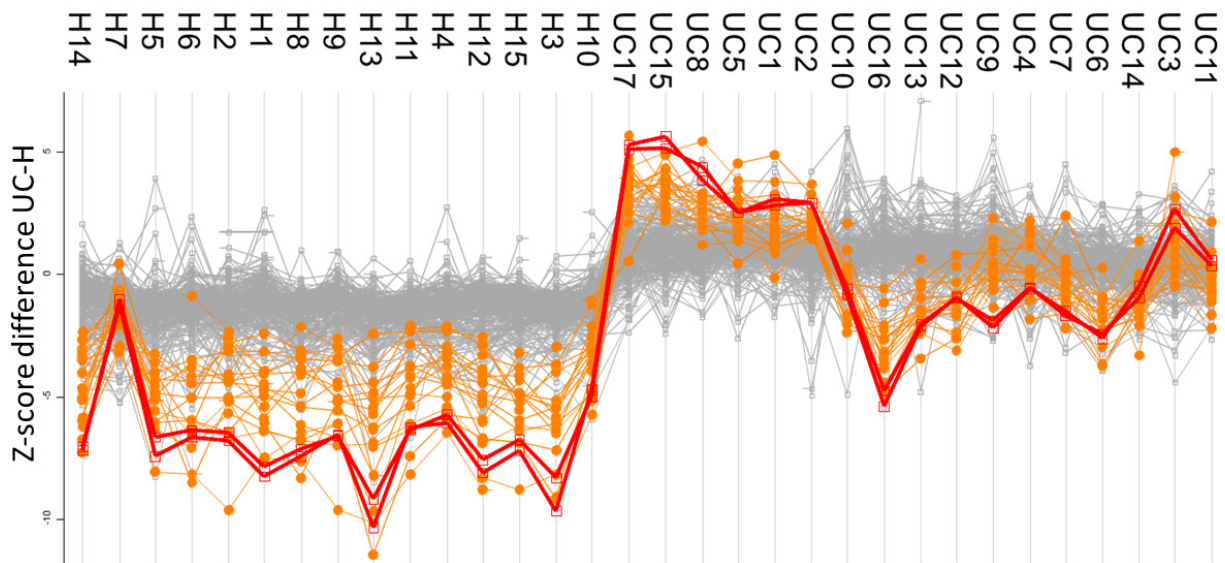
Suppl. figure1: Profile plot of the minor cluster (orange and red) of higher abundant proteins in UC in comparison with the major cluster of more abundant proteins (grey). The profile plots of S100A8 and S100A9, which form calprotectin, are shown in red.
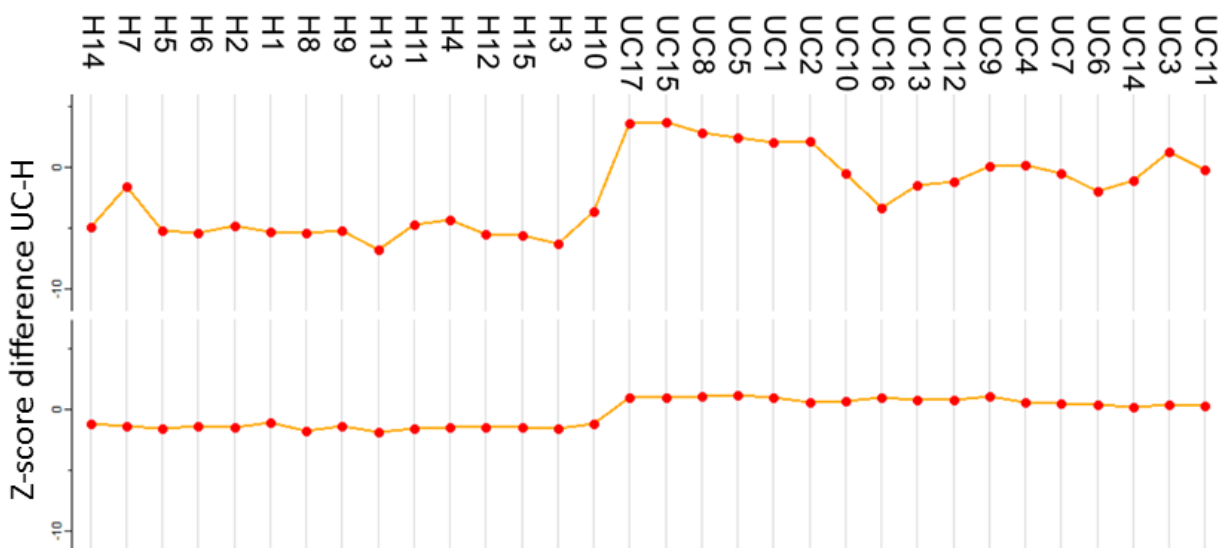

Suppl. figure2: Profile plot of the averages of the normalized protein intensities from the minor (top) and major (bottom) cluster of upregulated proteins in UC.
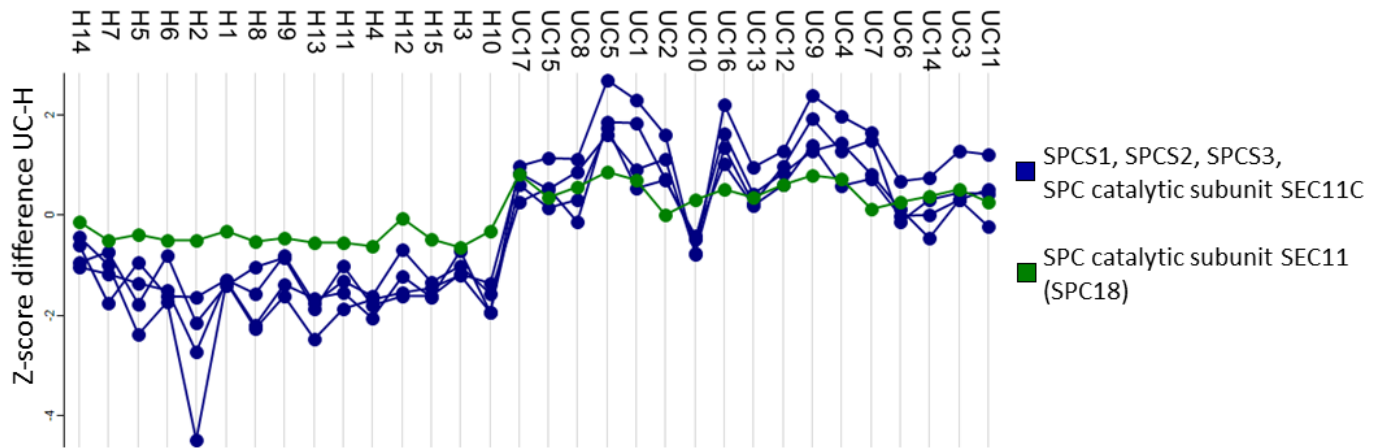
Ratio UC/H	p-value UC_H	Gene name
1.63153	9.60672	SPCS1
1.66381	11.2304	SPCS3
1.70966	10.7076	SPCS2
2.252	10.7787	SEC11C
1.26938	12.2656	SPC18

Suppl. figure3: Profile plot and statistics of signal peptidase complex protein abundances. SPCS1, SPCS2, SPCS3, and SEC11C present similar abundance profiles. SPC18 does not follow this pattern and its abundance increase in UC compared to healthy tissue is lower.
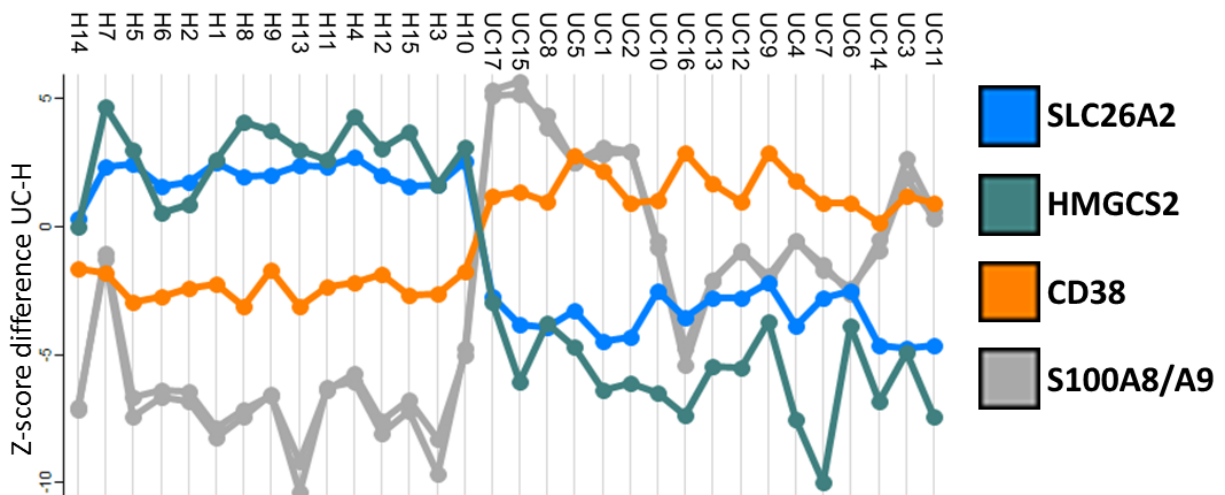
Suppl. figure4: Profile plot of normalized abundances of three proteins which differ strongly between UC and healthy tissue (SLC26A2, sulfate transporter; HMGCS2, Hydroxymethylglutaryl-CoA synthase; CD38, ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1). These were selected based on p-value and the difference between the group averages. Calprotectin (S100A8 and S100A9) is shown for reference. It shows a large difference between the UC and healthy average intensities, but it presents with low significance. Overlaps occur in which single healthy tissue samples show higher calprotectin abundances than single UC samples (e.g. H7 and UC16).

Mapping of metabolic pathways

Fatty Acid and Lipid Degradation

a phytoceramide → a (2R)-2-hydroxy-phytoceramide

a 2-hydroxy fatty acid

a (2R)-2-hydroxy even numbered straight chain 2,3,4-saturated fatty acid

a (R)-2-hydroxy even numbered straight chain 2,3,4-saturated fatty acyl CoA

an even numbered straight chain 2,3,4-saturated fatty aldehyde

formyl-CoA

formate

an odd numbered straight chain 2,3,4-saturated fatty acid

an odd numbered straight chain 2,3,4-saturated fatty acyl CoA

a 3-methyl-branched 2,3,4-saturated fatty acid

a 3-methyl-branched 2,3,4-saturated fatty acyl-CoA

a 2-hydroxy-3-methyl-branched 2,3,4-saturated fatty acyl-CoA

formyl-CoA

formate

a 2-methyl branched 2,3,4-saturated fatty aldehyde

a 2-methyl branched 2,3,4-saturated fatty acid

a 2-methyl branched 2,3,4-saturated fatty acyl-CoA

coenzyme A a 2,3,4-saturated fatty acid

a 1-phosphatidyl-1D-myo-inositol-4,5-bisphosphate

a 1,2-diacyl-sn-glycerol

a 1,2-diacyl-sn-glycerol 3-phosphate

a 1-acyl-sn-glycero-3-phosphocholine

a fatty acid

a phosphatidylcholine

a phosphatidylcholine

a 2-acyl-sn-glycero-3-phosphocholine

a 1,2-diacyl-sn-glycerol

a phosphatidylcholine

a phosphatidylcholine

a long-chain acyl-CoA

an O-long-chain-acyl-L-carnitine

an O-acyl-L-carnitine

an O-acyl-L-carnitine

O-palmitoyl-L-carnitine

palmitoyl-CoA

a trans-2-enoyl-CoA

a (3S)-3-hydroxyacyl-CoA

a (3R)-3-hydroxyacyl-CoA

a 3-oxoacyl-CoA

a 2,3,4-saturated fatty acyl CoA

a 2,3,4-saturated fatty acyl CoA

a 2,3,4-saturated fatty acid

a (3R)-3-hydroxyacyl-CoA

a (3S)-3-hydroxyacyl-CoA

a (3Z)-alk-3-enoyl-CoA

a trans-2-enoyl-CoA

a 2,3,4-saturated fatty acyl CoA

a (3S)-3-hydroxyacyl CoA

a 3-oxoacyl-CoA

a 2,3,4-saturated fatty acyl CoA

acetyl-CoA

+1.47 +1.20 +0.94 +0.67 +0.40 +0.13 −0.13 −0.40 −0.67 −0.94 −1.20 −1.47
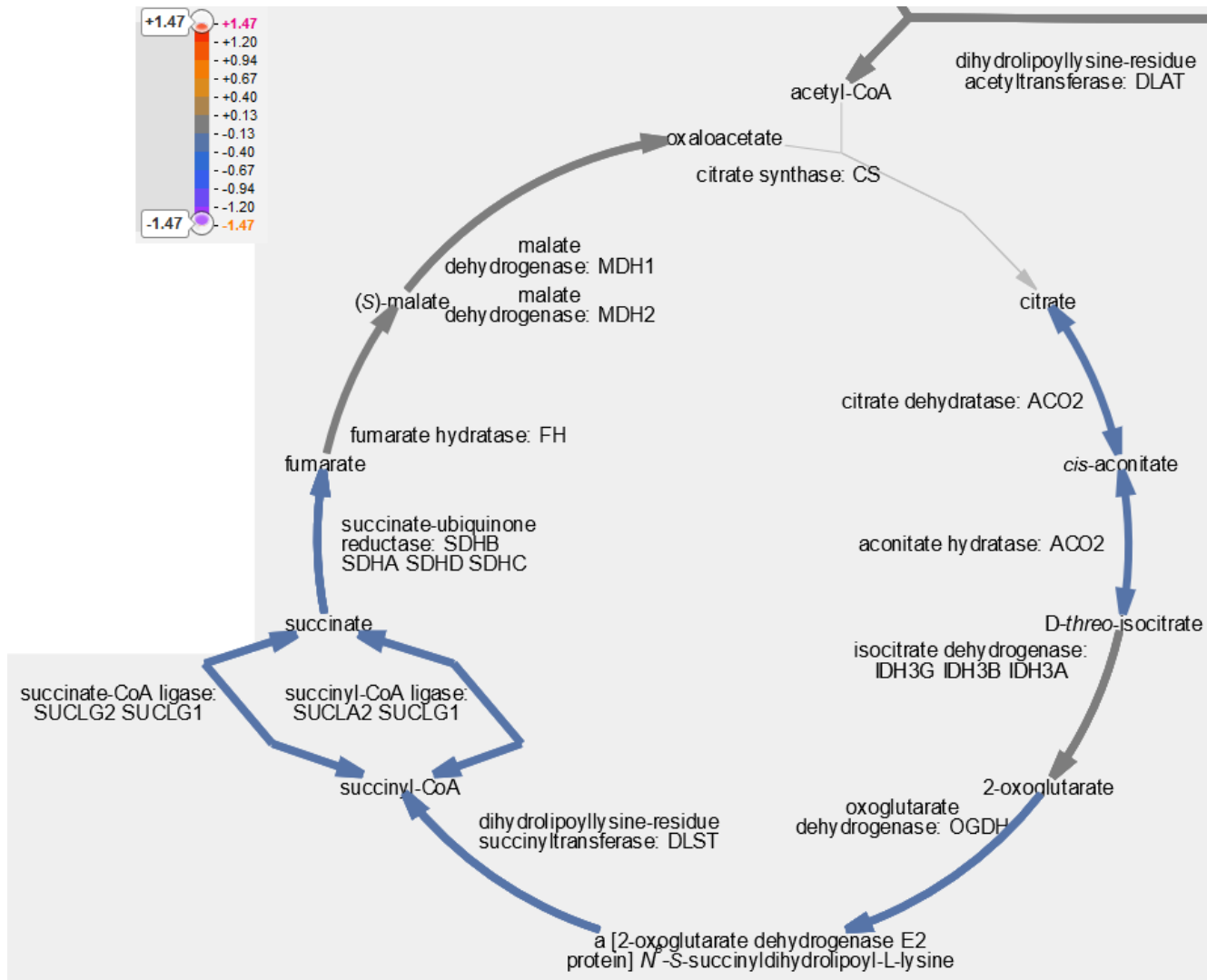
+1.47

−1.47

SuppPathway1: Protein abundance changes in selected fatty acid and lipid degradation pathways. Arrow colors indicate the Log ratios UC/healthy according to the color scale. Generated with https://humancyc.org.[1]
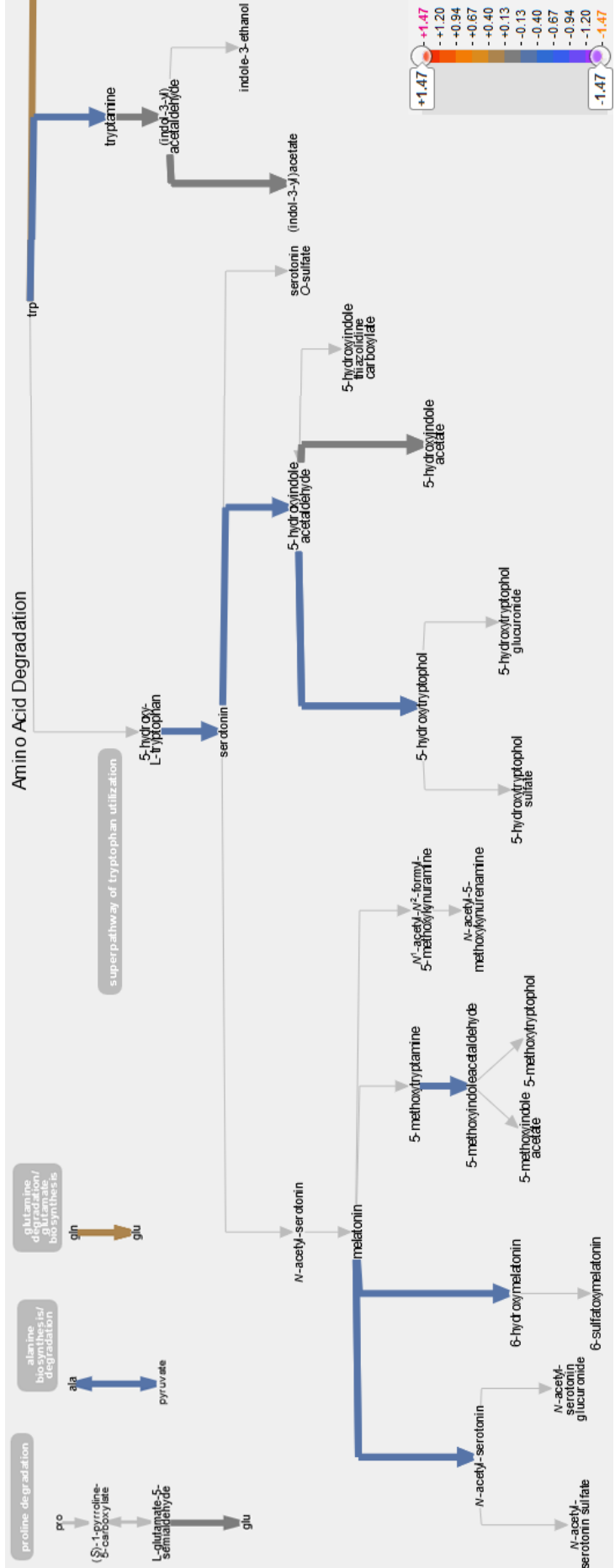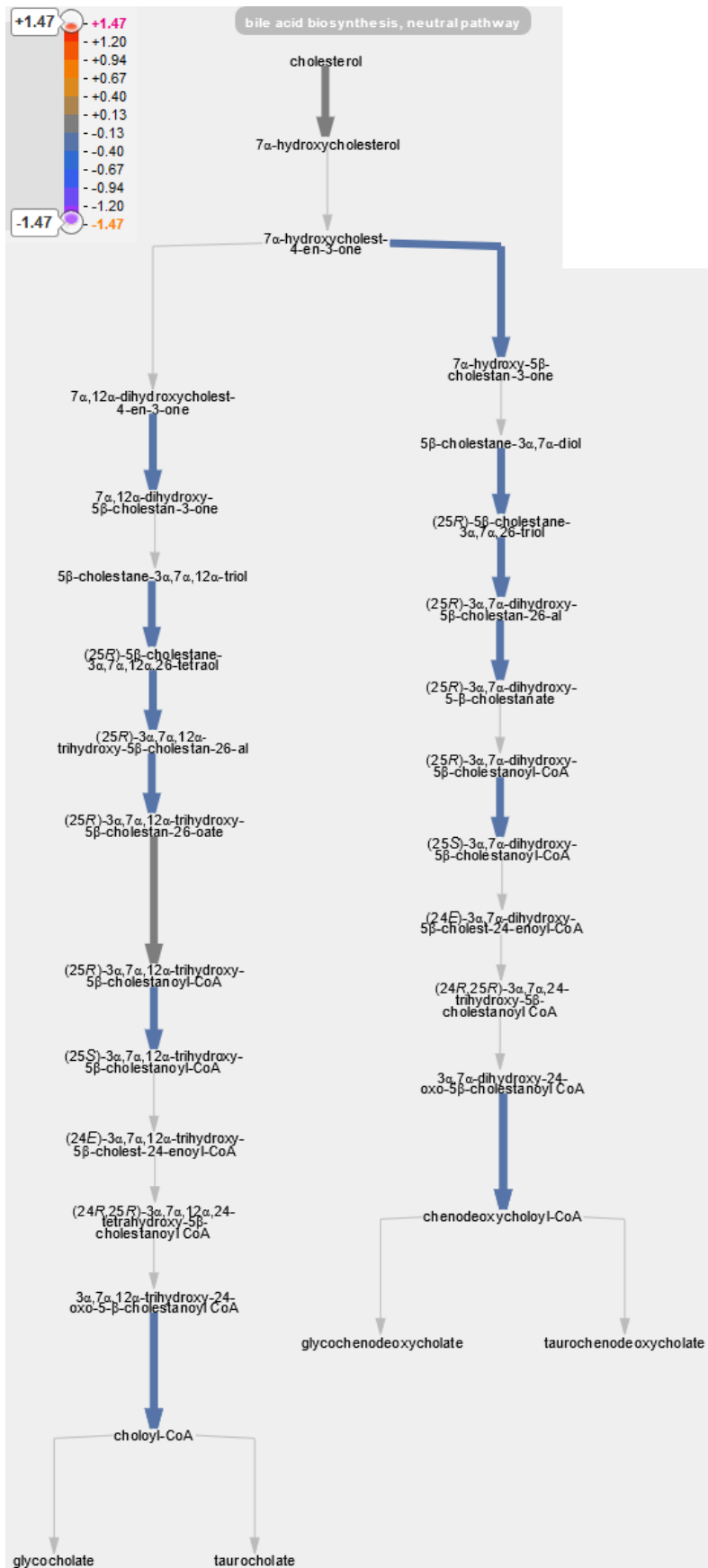
SuppPathway2: Protein abundance changes in selected amino acid degradation pathways (part 1), with focus on tryptophan (trp). Arrow colors indicate the Log ratios UC/healthy according to the color scale. Generated with https://humancyc.org.[1]
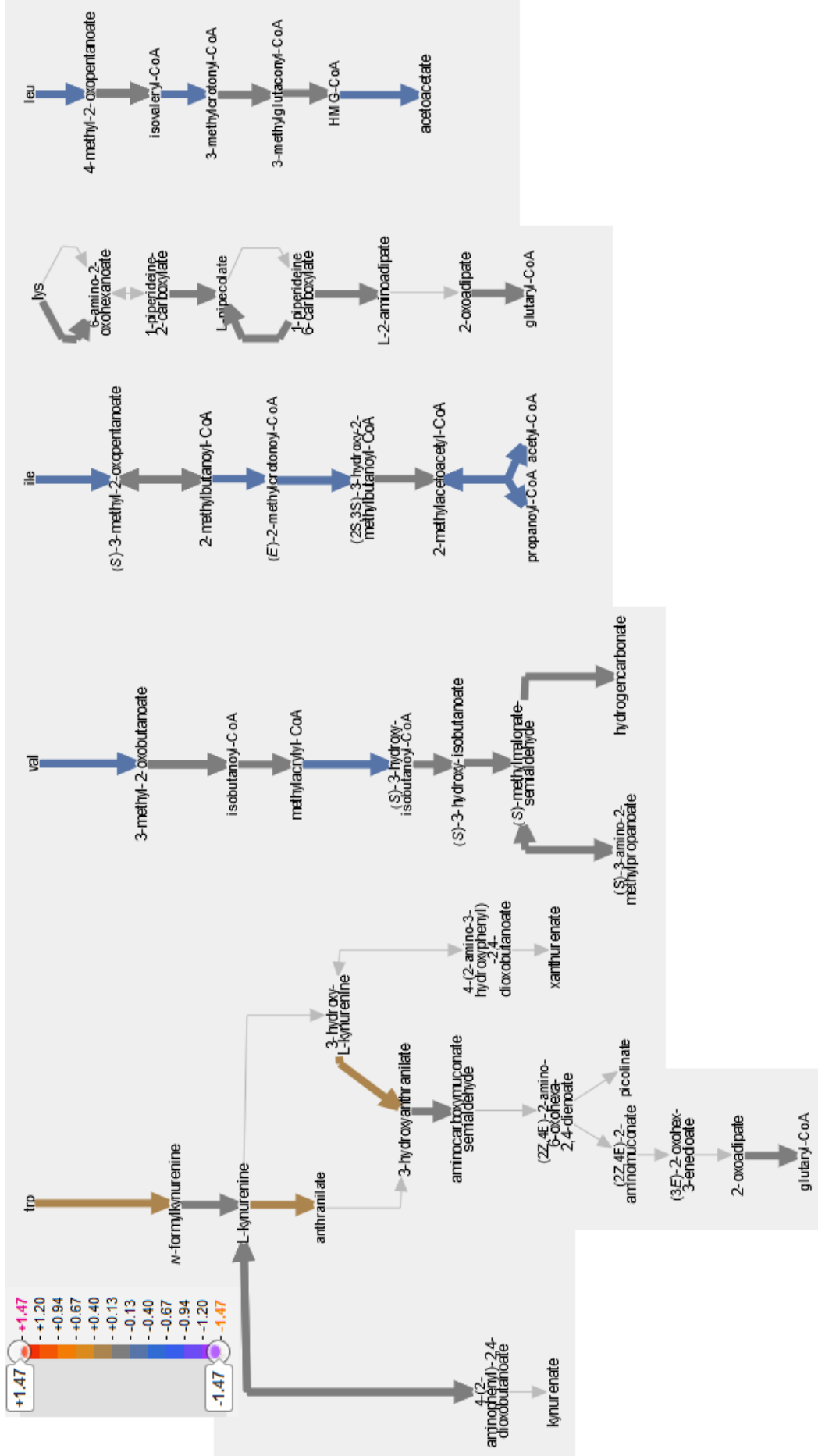
SuppPathway3: Protein abundance changes in selected amino acid degradation pathways (part 2). Arrow colors indicate the Log ratios UC/healthy according to the color scale. Generated with https://humancyc.org.[1]

SuppPathway4: Protein abundance changes in tricarboxylic acid (TCA) cycle. Arrow colors indicate the Log ratios UC/healthy according to the color scale. Generated with https://humancyc.org.[1]

SuppPathway5: Protein abundance changes in bile acid synthesis, neutral pathway. Arrow colors indicate the Log ratios UC/healthy according to the color scale. Generated with https://humancyc.org.[1]

1.   Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PDJGB:
      **Computational prediction of human metabolic pathways from the complete
      human genome**. 2004, **6**(1):R2.

Paper III

The proteome of ulcerative colitis in remission – functional differences in comparison with healthy controls and the active disease

Armin Schniers[1], Rasmus Goll[2,3], Sveinung Wergeland Sørbye[4], Jon Florholmen[2,3], Terkel Hansen*[1]

[1]Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[2]Department of Medical Gastroenterology, University Hospital of North Norway, Tromsø, Norway

[3]Gastroenterology and Nutrition Research Group, Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[4]Department of Clinical Pathology, University Hospital of North Norway, Tromsø, Norway

*Corresponding author: Associate Professor Terkel Hansen, Department of Pharmacy, UiT The Arctic University of Norway, NO-9037 Tromsø, Terkel.Hansen@uit.no, phone: +47 77 64 61 60

Abbreviations: BCA, bicinchoninic acid; CA, Carbonic anhydrase; DTT, 1,4-dithiothreitol; ETC, electron transport chain; FA, formic acid; FDR, false discovery rate; GO, gene ontology; HDL, high-density lipoproteins; IAA, iodoacetamide; IBS, Irritable bowel syndrome; LDL, low-density lipoproteins; NO, nitric oxide; PPAR, peroxisome proliferator-activated receptor; RAGE, Receptor for Activated Glycation End Products; SDC, sodium deoxycholate; SPC, signal peptidase complex; TCA cycle, tricarboxylic acid cycle; TEAB, Triethylammonium bicarbonate buffer; TFA, trifluoroacetic acid; TMT, tandem mass tag; UC, ulcerative colitis; UCDAI, ulcerative colitis disease activity index; UPR, unfolded protein response; VLDL, very-low-density lipoproteins

1 Abstract

Background: The primary treatment objective in ulcerative colitis (UC) is to reach remission. The colon mucosa in UC remission appears healed, but clinical observations and the susceptibility to relapses indicate that it is different from actual healthy tissue. The remission proteome is poorly described. The study aim is to characterize the proteome of patients in remission, compared with both healthy controls and treatment naïve patients with active UC. Emphasis will be placed on functional analyses due to the complexity of the data.

Methods: Colon mucosa biopsies from patients in remission from UC, from inflamed tissue of untreated UC patients at diagnosis, and from healthy controls were obtained during colonoscopy. Bottom-up proteomics with TMT labeling was applied to acquire quantitative protein data. Proteins were assigned to groups depending on how their abundances in remission relate to those in active UC and healthy controls. These groups were subsequently subjected to functional enrichment analysis.

Results: Our proteomic data set is to-date the deepest on colon mucosa biopsies from patients in remission. Our analyses show that most protein abundances were within the range observed in healthy controls and active UC, hence the remission proteome can be viewed as a combination of features from those two states. Only a small group of proteins was higher abundant in remission than in both active UC and in healthy controls. Metallothioneins and immunoglobulins retained in remission the abundances as found in active UC. Other immune system related proteins and fibrillar collagen returned to abundances as in healthy controls. Abundances of nutrient and energy metabolism recovered only partially, as their abundances were between those of active UC and healthy controls.

Conclusion: We present the first comprehensive characterization of the remission proteome in IBD. The present study reveals how abundances of 6775 proteins in remission relate to the abundances in active UC and healthy controls. Several biological functions were associated with distinct protein abundance patterns. These findings have implications for our understanding of symptoms and relapses in UC remission, and will make it possible to investigate how the patients are moving from active state of UC towards remission. Moreover, they may contribute to our understanding of the effects of current treatments and to the development of improved treatments to sustain the remission state in ulcerative colitis.

Key words: colon biopsies, ulcerative colitis remission, inflammatory bowel disease, metallothioneins, immunoglobulins

2   Introduction

Remission is the primary treatment objective in ulcerative colitis (UC) which apart from removal of the colon is a non-curable chronic disease. Ulcerative colitis affects 0.51% of the European and 0.25% of the North American population.(1)

No uniform definition of remission in UC exists. Remission in UC can be defined based on symptom relief, but it is increasingly recognized that endoscopic and histological healing is an important target for treatments. The symptom relief is important for the short-term quality of life, while reaching histological and endoscopic remission is an important measure to reduce complications.(2)

Clinical observations indicate that colon mucosa in endoscopic UC remission is different from healthy tissue. For instance, the prevalence of irritable bowel syndrome (IBS)-like symptoms in endoscopic UC remission is higher than in healthy controls.(3) The chronic nature of the disease with the susceptibility to relapses supports this assumption further.

Proteins are the major functional molecular units in organisms. Hence several proteomics studies investigated colon mucosa biopsies from UC patients in comparison with healthy controls(4, 5) and Crohn's disease subjects.(6) They further compared inflamed and non-inflamed mucosa from the same UC patients(7), and searched for biomarkers.(6, 8, 9) One proteomics study that included patients in remission was performed on serum, and quantified 91 inflammatory proteins and aimed to differentiate between UC and healthy controls, Crohn's disease and healthy controls, and UC and Crohn's disease.(10) However, no comprehensive investigation was conducted to characterize the proteomic features of UC remission in colon biopsies.

One transcriptomic study investigated the UC remission and compared it to healthy controls and active UC. The protein expression was subsequently determined by immunostaining for a selection of four proteins.(11) The quantification of the actual protein abundance as it was performed in this study is necessary, because mRNA levels correlate only poorly with the abundances of the respective proteins.(12)

Proteome changes in remission may explain the symptoms and relapses observed in patients in UC remission. In relapses the non-inflamed mucosa in remission acquires inflammation. The underlying causes for this may lie in the proteome and are of interest with regard to the UC pathophysiology and possibly its etiology.

We conducted the first investigation of the in-depth proteome in UC remission. Functional analyses reveal how protein functions change in comparison to both active UC and healthy controls, i.e. for most proteins whether they in remission return to abundances as in healthy controls, keep similar

abundances as in active UC, or lie in between. Basis for these analyses is the proteomics data of colon biopsies with the currently highest numbers of protein identifications.

3   Materials and Methods

3.1   Enrolled patients and biopsies collection

The mucosal biopsies were collected in Norway from patients in deep remission from UC, newly diagnosed treatment-naïve UC patients, and healthy controls. The deep remission state was defined as endoscopically healed mucosa (Mayo score = 0) and normalized TNF-α values. Clinical and endoscopic criteria were applied to establish the UC diagnosis as defined by the European Crohn and Colitis Organization (ECCO) guidelines.(13) The ulcerative colitis disease activity index (UCDAI) was used to evaluate the degree of inflammation.(14) TNF-α mRNA levels were determined by real-time PCR.(15) The healthy controls are subjects admitted for cancer screening with normal endoscopic and histological findings. The recruited patients did not suffer from irritable bowel disease and were not on NSAIDs medication before the colonoscopy. Biopsies from UC patients in deep remission and healthy controls were obtained from the rectum. The biopsies from patients with active UC patients were obtained from the most inflamed area in rectum or sigmoid colon. The biopsies were immediately frozen in dry cryotubes and stored at -70 °C until sample preparation. 14 patients in deep remission, 17 patients with active UC at debut, and 15 healthy controls were enrolled as presented in table 1.

3.2   Patient and control characteristics

The patient and control characteristics are presented in Table 1. The UC active and healthy control subjects included showed typical histological findings for UC and normal, non-inflamed mucosa, respectively.

Table 1: Baseline characteristics of 14 patients in deep UC remission,17 patients with active ulcerative colitis at debut, and 15 healthy controls. The UCDAI score and Geboes index(16) averages and SD for patients with active UC were calculated from available values for 15 and 16 patients, respectively.

Study Group	Number of Subjects	Average Age (SD)	Sex Female/Male	Average years since UC debut (SD)	Median TNF-α (IQR)	Average UCDAI Score (SD)	Average Geboes index (SD)
UC remission	14	46.8 (16.0)	5/9	7.4 (4.5)	4800 (3000)	0.4 (0.95)	1.15 (1.7)
UC debut active	17	39.4 (16.7)	4/13	0	14350 (16725)	8.7 (2.2)	7.9 (3.8)
Healthy Control	15	51.9 (14.3)	5/10	-	4500 (2400)	-	-

8 of the patients in UC remission had a total score of 0, 2 patients had a total score of 2, 1 patient had a total score of 3 and the last patient had a total score of 6, mainly resulting from score 4 in erosion. The latter patient was the only patient scoring above 1 in any of the sub-scores.

3.3   Tissue lysis

The biopsies were added to 250 µl cold lysis buffer (5% sodium deoxycholate (SDC), 8 M urea, 100 mM triethylammonium bicarbonate buffer pH 8.5 (TEAB)) in MagNA Lyser Green Bead tubes (Roche Diagnostics AG, Rotkreuz, Switzerland). The biopsies were subsequently homogenized with a MagNA Lyser Instrument (Roche Diagnostics AG, Rotkreuz, Switzerland) for 35 seconds at 6500 rpm. The homogenates were stored at -70 °C until further processing.

3.4   Protein content measurement

A bicinchoninic acid (BCA) assay (Pierce™ BCA Protein Assay Kit, Thermo Fisher Scientific) was performed according to the manufacturer's protocol to determine the protein content in the samples, with standards and samples containing 5 mM TEAB.

3.5   Reduction and alkylation

Aliquots corresponding to 60 µg protein were transferred to Protein LoBind tubes (Eppendorf AG, Hamburg, Germany).

1,4-dithiothreitol (DTT) was added to a concentration of 5 mM, followed by incubation at 54 °C for 30 min to reduce cysteins. The cysteins were alkylated by adding iodoacetamide (IAA) to a concentration of 15 mM and incubating for 30 min at room temperature in the dark. Excess IAA was removed by adding DTT to a concentration of 5 mM.

3.6   Predigestion with Lys-C

0.6 µg endoproteinase Lys-C (enzyme-to-protein ratio 1:100, w/w, Wako Chemicals GmbH, Neuss, Germany) were added. Calcium chloride solution, water, and 1 M TEAB were added to a concentration of 1 mM calcium chloride, 6 M urea, and 100 mM TEAB. The predigestion was performed for 8 hours at 37°C under gentle agitation.

3.7   Digestion with trypsin

3 µg trypsin (enzyme-to-protein ratio 1:20, w/w, Sequencing Grade Modified Trypsin, Promega Corporation, Madison, USA) were added. The solution was then diluted with water, calcium chloride solution, and 1M TEAB to a concentration of 1 mM calcium chloride, 1 M urea, and 100 mM TEAB. The tryptic digestion was performed for 16 hours at 37°C under gentle agitation.

3.8 Isobaric labeling with TMT

The standard was produced by mixing equal amounts of peptide from each sample, based on BCA results. This mixture was then labeled with tandem mass tag (TMT) 126 according to the manufacturer's protocol (Thermo Fisher Scientific, Rockford, USA).

25 μg peptides of each sample were labeled with the other TMTsixplex isotopes according to the manufacturer's protocol. 25 μg peptides were labeled with 0.4 mg TMT reagent, when the starting protein concentration was ≥ 2200 μg/ml in the respective sample. Samples with lower starting concentrations were labeled with more TMT reagent to ensure complete labeling (see supplementary information).

16.7 μg labeled peptides from the standard and from 5 differently labeled samples were mixed for each fractionation.

3.9 High pH reversed-phase fractionation

50% formic acid (FA) solution was added to the mixtures to a final concentration of 2.5% FA. The pH was measured to ensure that it is ≤ 2. The precipitate was removed by centrifugation at 16,000 g for 15 min and carefully transferring the supernatants to new tubes. The acetonitrile in the supernatants was removed in a vacuum concentrator.

10% trifluoroacetic acid (TFA) solution was added to a concentration of 0.1% TFA. The samples were fractionated according to the manufacturer's protocol with Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher, Rockford, USA). The samples were loaded stepwise on the fractionation columns by centrifugation for 2 minutes after each loading of 300 μl, if necessary. The solvents were evaporated from the resulting fractions in a vacuum concentrator. The dried samples were then redissolved in 10 μl of 0.1% TFA for LC-MS/MS analyses.

The peptide concentration in these solutions was determined by measuring the absorption at 205 nm on a Nanodrop 2000 (Thermo Fisher Scientific, Bremen, Germany), based on an extinction coefficient of 31 mg/ml.(17)

3.10 Nano-LC-MS/MS

The peptides were separated by an EASY-nLC 1000 that was coupled to a Q Exactive mass spectrometer for analysis (both Thermo Fisher Scientific, Bremen, Germany).

2.0 μg peptides per sample were injected. The peptides were concentrated on a reversed-phase trap column. They were then separated on a reversed-phase main column with a gradient of the solvents 0.1% FA in water (solvent A) and 0.1% FA in acetonitrile (solvent B). The proportion of solvent B was

increasing from 0% to 5% over 19 min, to 30% at 180 min, and to 100% at 200 min. The LC-MS/MS conditions are described in more detail in the supporting information.

3.11 Processing of mass spectrometry data

The mass spectrometry data was analyzed with MaxQuant.(18) Subsequent analyzes of the resulting quantitative protein data were performed with Perseus(19) and ClueGO(20), an app for Cytoscape.(20, 21) See supporting information for details on the data analyzes.

3.12 Filtering and intensity normalization

A log2-transformation was performed in Perseus on the protein intensities. Proteins with the labels "potential contaminant", "reverse" and "only identified by site" were removed. The intensities of the standard were subtracted from the samples of the same experiment. Z-score normalization (with matrix access: columns) was performed on the resulting values.

3.13 Statistical analyses

Shortly, the samples were grouped according to their clinical state (Remission from UC, active UC, healthy controls) and analysis of variance (ANOVA) testing (S0 = 2, FDR = 0.01) was performed on the normalized intensities. A post hoc test (false discovery rate (FDR) = 0.01) was performed on the proteins with ANOVA significance. The post hoc test returns the significant pairs. Based on the combinations of significant pairs, abundance patterns were assigned to the proteins as categorical values. These categorical values were added to the list of all proteins with >70% valid values. The gene ontology (GO) annotations GOBP (biological process), GOCC (cellular component), GOMF (molecular function),(22) GSEA,(23) InterPro,(24) KEGG,(25) Pfam,(26) PRINTS,(27) PROSITE,(28) Reactome,(29) SMART,(30) and UniProt keywords(31) were added to this matrix. A Fisher exact test was performed to determine enriched functions among proteins with the different abundance patterns. GOBP annotations(22) were used for the ClueGO analyses. See supporting information for more details.

4 Ethical considerations

All enrolled subjects were informed and signed a written consent form. The study and the storage of biological material were approved by the Regional Committee of Medical Ethics of North Norway and the Norwegian Social Science Data Services under the number REK NORD 2012/1349.

5 Results and discussion

5.1 Exploratory analysis

The colon biopsy proteome of patients in remission differed from both active UC and healthy controls (figure 1A and B). Scatter plots of averaged profiles and the corresponding Pearson correlations showed a comparably low correlation between the remission groups. This correlation was only slightly higher than those between remission samples and healthy controls. The correlation of remission samples with UC samples was lower (figure 1A). Hierarchical clustering separated active UC samples entirely from healthy and remission samples. We observed limited separation with some overlap between remission and healthy control samples (figure 1B). These findings indicate that the remission proteome is more similar to that of healthy controls than to the active UC proteome.
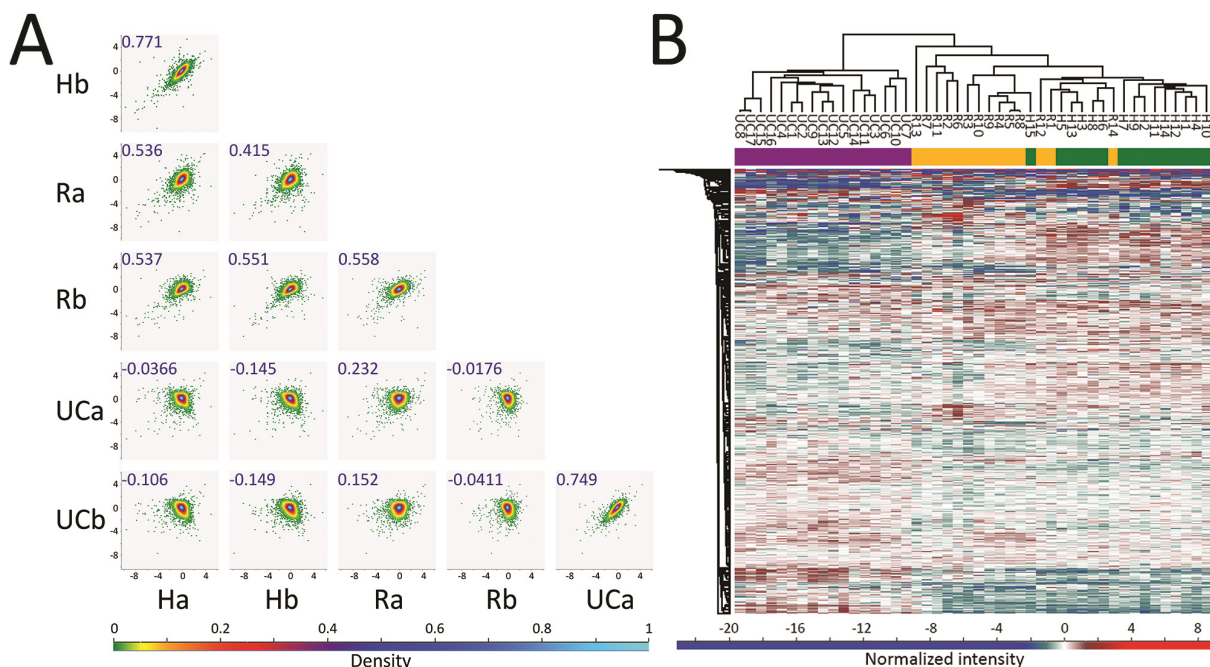

Figure 1: Exploratory analysis of the remission proteome. (A) Multi-scatter plot of averaged normalized protein intensities. The averages are the mean of half of the samples from each state (Ha: H1, H3, H4, H10, H12, H13, H14, H15; Hb: H2, H5, H6, H7, H8, H9, H11; Ra: R3, R4, R5, R8, R9, R10, R11; Rb: R1, R2, R6, R7, R12, R13, R14; UCa: UC1, UC2, UC4, UC5, UC6, UC9, UC10, UC11; UCb: UC3, UC7, UC8, UC12, UC13, UC14, UC15, UC16, UC17). Other random groupings give similar results. Included are all 6775 proteins that are quantified in more than 70% of the samples. Dark blue numbers indicate the Pearson correlation of the respective pairs. Healthy and UC samples show high in-group correlations, the correlations between Healthy and UC are very low. The remission samples show a comparably low in-group correlation, which is only slightly higher than the correlation between remission and healthy samples. The correlation between healthy and remission is higher than between UC and remission, which indicates that the remission state is more similar to the healthy state than the UC state. (B) Hierarchical clustering of protein intensities and samples. Included are the 5675 proteins that are quantified in all samples. The heatmap of normalized intensities visualizes the proteome changes occurring in the different conditions. UC samples (purple) separate entirely from remission (yellow) and healthy (green) samples. A trend for separate clustering of remission and healthy samples shows, but an overlap between the conditions occurs. This supports the finding

further that remission samples are more similar to healthy controls than to UC samples. H, Healthy controls; R, Remission; UC, Ulcerative colitis.

5.2 Functional analysis

The multiple-sample test (ANOVA, S0 = 2, permutation-based FDR = 0.01) returned a significance finding for 678 proteins. 631 of these proteins were differently abundant in remission compared to UC, healthy controls, or both (post-hoc test, FDR = 0.01).

39 (6%) proteins were higher abundant in remission compared to both active UC and healthy controls. We found no proteins to be lower abundant in remission than in both UC and healthy controls. The remaining 592 (94%) proteins were similarly abundant as in either UC or healthy controls, or the abundances in remission were between those observed in UC and healthy controls. Hence, the largest part of the remission proteome could be viewed as a combination of features from UC and healthy controls.

359 (61%) of these 592 proteins showed similar abundances in remission as in healthy controls. 82 proteins (14%) had a similar abundance in remission as in active UC. 151 (26%) of the protein abundances in remission lay in between healthy control and active UC abundances. The described groups could further be divided into proteins with higher and lower, respectively, abundances in healthy controls compared to active UC. In total, the proteome of remission patients seemed to resemble more the proteome of healthy controls than patients with an active UC. We present the relative protein abundances of the respective groups in figure 2.

Table 2: Selection of enriched functional terms in groups of proteins that are differently abundant in remission than in UC and/or healthy controls, including enrichment characteristics. Terms are listed in the order as they are discussed in the article. (H, healthy control; R, remission from UC; UC, active UC; "<", significantly lower abundant in group to the left; ">", significantly higher abundant in group to the left; "≈", not significantly different abundant between groups to both sides).

Abundance pattern	Number of proteins with pattern	Enriched functions	Number of proteins with pattern and function	Enrichment factor	p-value of enrichment
H < R > UC (group 1)	39	Hormone	3	57.9	1.44E-05
		Hormone metabolic process	5	14.5	2.00E-05
		Vitamin transport	3	57.9	1.44E-05
		High-density lipoprotein particle receptor binding	3	173.7	1.76E-07
		Regulation of very-low-density lipoprotein particle remodeling	4	173.7	9.38E-10
		Serpin	5	45.7	5.25E-08
		Muscle protein	6	41.7	3.93E-09
H ≈ R > UC (group 2)	140	Fibrillar collagen	5	34.6	7.08E-08
		Glycoside catabolic process	3	29	8.29E-05
	219	Neutrophil mediated cytotoxicity	3	23.2	0.000129

H ≈ R < UC (group 3)		NADPH oxidase complex	5	30.9	3.38E-08
		Superoxide anion generation	5	22.1	6.65E-07
		B cell receptor signaling pathway	8	11.8	1.42E-07
		Activated T cell proliferation	3	23.2	0.000129
		Natural killer cell mediated cytotoxicity	12	6.6	1.37E-07
		Collagen degradation	4	17.7	3.38E-05
		Cellular response to type I interferon	7	6.6	6.26E-05
		Protein processing in endoplasmic reticulum	23	5.8	3.56E-12
		Endoplasmic reticulum unfolded protein response	17	9.2	1.02E-12
		Signal peptidase complex	3	30.9	3.33E-05
H > R ≈ UC (group 4)	50	Metallothionein	3	135.5	3.78E-07
		Prostaglandin metabolic process	3	27.1	0.000158
H < R ≈ UC (group 5)	32	Immunoglobulin domain	14	19.4	1.60E-15
H > R > UC (group 6)	135	Fatty acid beta-oxidation	10	16.189	2.12E-10
		Butanoate metabolism	7	18.5	4.28E-08
		Ketone body biosynthetic process	4	50.2	1.51E-07
		Reactome mitochondrial fatty acid beta oxidation	6	25.1	4.61E-08
		Valine, leucine and isoleucine degradation	10	12.5	3.42E-09
		Tryptophan metabolism	6	11.2	1.11E-05
		Citrate cycle (TCA cycle)	7	11	2.23E-06
		Electron transport chain (ETC)	11	5.5	3.88E-06
		Fatty acid binding	7	11.3	1.77E-06
		PPAR signaling pathway	7	9.2	7.45E-06
		Nuclear receptors in lipid metabolism and toxicity (Biocarta_Nuclearrs_Pathway)	3	50.2	7.74E-06
		Xenobiotic metabolic process	10	6.7	1.73E-06
		Glucuronate metabolic process	4	22.3	1.72E-05
		Bile acid and bile salt transport	3	25.1	0.000146
H < R < UC (group 7)	16	Immunoglobulin domain	7	19.4	2.46E-08

Enrichment analysis of these groups returned, among others, the functions presented in table 2. We discuss these functions in the following sections.

This article does not cover all results from the enrichment analysis. Furthermore, single proteins that are not covered by the enrichment analyses can be of importance. The article format allows the discussion of only a small selection of single proteins. Hence we encourage the reader to consult the supplementary data content 1 which provides the complete lists of proteins and enriched functions with fold changes and p-values, also for the non-significantly changed proteins.
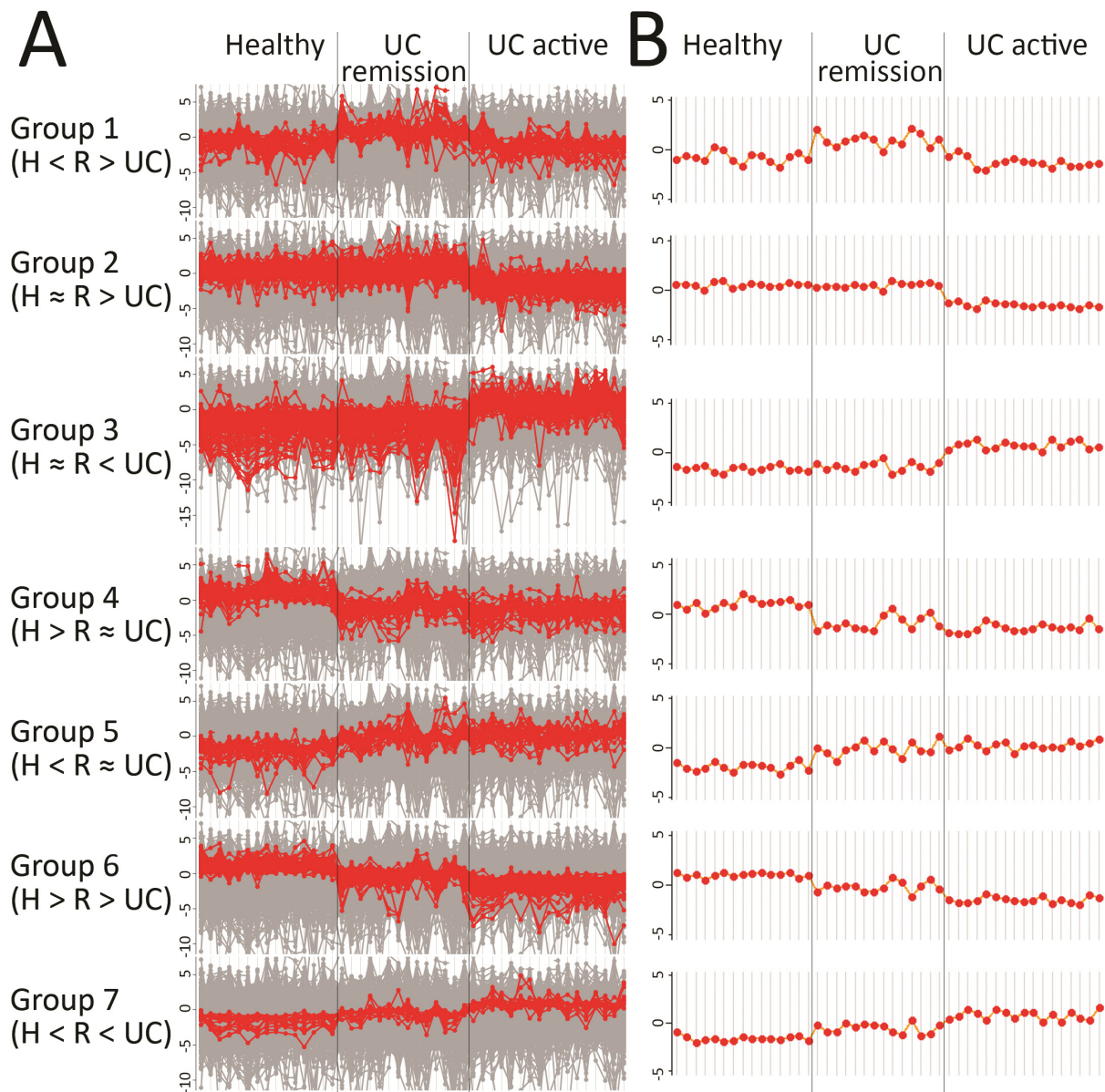
Figure 2: Normalized protein intensities of the protein groups that are differently abundant in remission compared to active UC and/or healthy controls. (A) Profile plots of individual protein intensities. Red = proteins of the respective groups, Grey = other proteins. (B) Profile plots of the averaged protein intensities from the respective groups. ("<", significantly lower abundant in group to the left; ">", significantly higher abundant in group to the left; "≈", not significantly different abundant between groups to both sides)

5.2.1 Proteins with higher abundances in remission than in active UC and healthy controls (H<R>UC, group 1)

The 39 proteins in group 1 showed comparable levels in active UC and healthy controls, but were higher abundant in remission tissue (Figure 2). This is remarkable because most proteins showed similar abundances in remission as in patients with active UC and healthy controls, respectively, or

their abundances lay in between those. Abundance changes in this group were possibly effects of the administered drugs, either as side effects or as mechanisms of action.

The hormone precursor angiotensinogen and the hormones glucagon and insulin-like peptide 5 (INSL5) were higher abundant in remission than in both active UC and healthy controls. Angiotensinogen (SERPINA8) is the precursor of angiotensin I, which is further converted to angiotensin II. Hence, angiotensinogen is part of the renin-angiotensin-aldosterone system (RAAS). This system regulates the blood pressure and the reabsorption of water and electrolytes. Glucagon regulates glucose and lipid homeostasis, energy expenditure, and thermogenesis, and decreases the sense of hunger and the food intake.(32) Insulin-like peptide 5 regulates appetite and glucose metabolism.(33) The hormone transport proteins corticosteroid-binding globulin (SERPINA6) and transthyretin (TTR) were in group 1 as well. Corticosteroid-binding globulin is the main transport protein of glucocorticoids and progesterone.(34) Transthyretin transports the thyroid hormone thyroxine.

Also vitamin transport functions were enriched in group 1. This affected the protein Afamin (AFM) which transports vitamin E (35) as well as retinol-binding protein 4 (RBP4) and transthyretin (TTR) which transport vitamin A (retinol). Transthyretin is, beside its function in thyroxine transport (see above), the carrier protein of RBP4. RBP4 has a low molecular weight and the binding of transthyretin to RBP4 prevents an extensive loss of RBP4 through glomerular filtration.(36, 37)

Serpins were enriched in group 1. Serpins are a protein superfamily with a highly conserved core structure, but overall low sequence homology between family members. Most, but not all, family members are serine protease inhibitors, hence the term serpin was coined as an acronym.(38) Serpins with different functions showed an increased abundance in remission. Several of these differently abundant serpins (alpha-1-antitrypsin, alpha-1 antichymotrypsin, kallistatin) inactivate proteins of the immune system. The other serpins with increased abundance in remission were the hormone precursor angiotensinogen (SERPINA8, see above), and corticosteroid-binding globulin (SERPINA6, see above).

Alpha-1-antitrypsin (SERPINA1) inhibits neutrophil elastase. Neutrophil elastase is a protease which, when unopposed, can degrade tissue.(39) Alpha-1-antichymotrypsin (SERPINA3, ACT) inhibits cathepsin G. Cathepsin G is, like neutrophil elastase, a protease which is secreted by neutrophils and can degrade pathogens but also host tissue.(40, 41) The abundances of neutrophil elastase and cathepsin G were, like the abundances of most other inflammatory proteins, not differently abundant in remission compared to healthy controls, but higher abundant in active UC (group 3). Kallistatin (SERPINA4) inhibits kallikrein. Kallikrein is part of the plasma kallikrein-kinin-system. It

cleaves kininogens to release kinins such as bradykinin.(42, 43) Our data shows that kallikrein abundances were significantly higher abundant in healthy controls and significantly lower abundant in active UC than in remission (group 6). The increased levels of serpins directed against inflammatory and tissue-degrading enzymes were possibly a mechanism of action of the administered drugs.

We furthermore observed enrichment of muscle proteins (actin, myosin, tropomyosin, caldesmon) and of proteins with functions in lipoproteins in group 1. The enrichment of lipoprotein functions was mostly due to apolipoprotein classes A (A-I, A-II, A-IV) and C (C-I, C-II, C-III). The single apolipoproteins of these classes are associated with chylomicrons, high-density lipoproteins (HDL), and/or very-low-density lipoproteins (VLDL), but not with low-density lipoproteins (LDL).(44)

5.2.2 Proteins with different abundances in remission compared to active UC and similar abundances in remission as in healthy controls

In this section we discuss proteins with abundances that were in remission significantly different from active UC, but not from healthy controls (Groups 2 and 3, Figure 2). This indicates that under remission the abundances of the respective proteins were fully restored to levels as in healthy controls. Hence it is less likely that the proteins in this group, and presumably their functions, contribute to the symptoms and the risk of relapses observed in remission tissue.

The 140 proteins in group 2 were not significantly different abundant in remission compared to healthy tissue, but lower abundant in active UC (Healthy ≈ Remission > UC). This indicates that the functions of these proteins were restored in patients in remission as opposed to active UC.

Fibrillar collagens were enriched in group 2. They provide three-dimensional frameworks in the tissue.(45) Metalloproteinases, which were higher abundant in active UC than in both remission and healthy controls, can degrade these proteins, see group 3 (section 5.2.2.2).

Glycoside catabolic processes were enriched in group 2 as well. This was due to the sugar hydrolases alpha-L-fucosidase 1 and 2 (FUCA1, FUCA2), and alpha-N-acetylgalactosaminidase (NAGA). Alpha-L-fucosidases 1 and 2 remove fucose residues from glycoproteins and glycolipids, while alpha-N-acetylgalactosaminidase removes terminal α-linked *N*-acetylgalactosamine residues.(46) Mutations of FUCA1 cause fucosidosis(47), a mutation of the NAGA gene causes Schindler disease.(48) Moreover, sialate O-acetylesterase (SIAE) was in group 2. Sialate O-acetylesterase cleaves O-acetyl groups in 9-position from sialic acid residues. Sialic acid residues occur in glycosilated proteins. The removal of these 9-O-acetyl groups from CD22 is a prerequisite for its inhibitory functions on the B cell receptor activation. A lack of sialate O-acetylesterase in mice causes enhanced B cell receptor

signaling, defects in B cell differentiation, expression of autoantibodies against chromatin, and the deposition of immune complexes in the glomeruli.(49).

We summarize as group 3 the 219 proteins that showed similar abundances in healthy controls and patients in remission, but higher abundances in active UC (Healthy ≈ Remission < UC). This indicates that the respective functions returned in remission from high levels in active UC to levels as observed in healthy controls.

Naturally, proteins of the immune system were overrepresented in group 3. These proteins were associated with neutrophils (e.g. NADPH oxidase complex), B cells, T cells, and natural killer cells. The immune system proteins in group 3 comprised proteins which degrade collagen (e.g. metalloproteinases). The lower abundance of these proteins in remission and healthy tissue could explain the finding that fibrillar collagen was higher abundant in both remission and healthy controls than in active UC (see group 2). Moreover, cellular responses to type I interferons in remission was at a similar level as in healthy controls. Hence, our findings indicate that immune system processes – with exception of immunoglobulins (see group 5) – were contained in remission.

Proteins related to protein processing in the endoplasmic reticulum are higher abundant in active UC than in healthy controls.(ref paper 2) These proteins were enriched in group 3, which shows that their abundances in remission decreased to the levels found in healthy controls. This included the associated unfolded protein response (UPR) and the signal peptidase complex (SPC).

Beside these enriched functions, the inducible nitric oxide synthase (NOS2), which produces nitric oxide (NO) and shows increased abundances in UC(50), returned in remission to abundances comparable to those in healthy controls.

5.2.3 Proteins with similar abundances in remission as in active UC and differently abundant in healthy controls

This section discusses proteins with abundances that were in remission significantly different from healthy controls, but not from active UC, i.e. abundances of proteins in this group remained in remission at levels as observed in active UC (Groups 4 and 5, Figure 2).

These proteins are of particular interest, because their persisting abundance changes could cause the symptoms observed under remission, e.g. IBS-like symptoms(3), and may be involved in the occurrence of relapses. They may further reflect which underlying changes cause the transition from healthy tissue to inflamed tissue in UC in the first place.

Group 4 summarizes the 50 proteins that were similarly abundant in remission as in active UC samples, and higher abundant in healthy controls (Healthy > Remission ≈ UC).

All identified metallothioneins (Metallothionein-1F (MT1F), Metallothionein-1H (MT1H), Metallothionein-2 (MT2A)) were in group 4. This shows that the abundance decreases of metallothioneins, which were observed in active UC compared to healthy controls (ref paper 2), persisted in non-inflamed mucosa in remission. Metallothioneins are cystein-rich, heavy metal-binding proteins. They have a role in the detoxification of heavy metals and the homeostasis of essential metal ions such as zinc (figure 3A). They furthermore function as antioxidants against reactive oxygen and nitrogen species.(51)
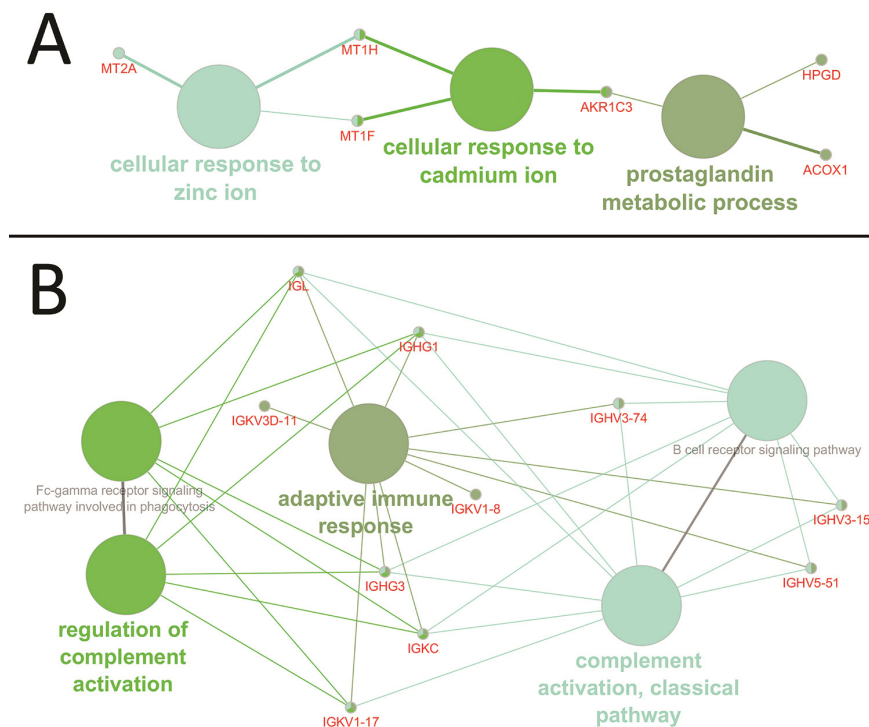

Figure 3: Functional networks of proteins with similar abundances in remission as in active UC. (A) Functional network of metallothioneins (Metallothionein-1F (MT1F), Metallothionein-1H (MT1H), Metallothionein-2 (MT2A)) and prostaglandin metabolism proteins in the protein group H > R ≈ UC. (B) Functional network of immunoglobulins in the protein group H < R ≈ UC. Generated with Cytoscape/ClueGO. Standard settings except for activated GO fusion and custom reference set with all proteins identified in >70% of samples. Annotations: GOBP, updated 04.09.2018.

Prostaglandin metabolism was enriched in group 4. This affected aldo-keto reductase family 1 member C3 (AKR1C3), 15-hydroxyprostaglandin dehydrogenase (HPGD), and peroxisomal acyl-coenzyme A oxidase 1 (ACOX1) (figure 3A). AKR1C3 acts, beside its function in testosterone synthesis, as a prostaglandin F synthase.(52) HPGD and ACOX1 are involved in the prostaglandin catabolism. HPGD inactivates prostaglandins by oxidizing their 15-hydroxy group to a 15-keto group.(53) ACOX1 oxidizes CoA esters of prostaglandins, e.g. prostaglandin E2, in the first step of the chain-shortening β-oxidation. Chain-shortened prostaglandins are excreted in urine.(54, 55)

Group 5 contains the 32 proteins that showed similar abundances in remission as in active UC and lower abundances in healthy controls than in remission (Healthy < Remission ≈ UC). Most of the significantly different immunoglobulins are in group 5 (table 2, figure 3B). However, group 7 (Healthy < Remission < UC) contains further immunoglobulins. The immunoglobulins thereby constitute the only major immune system related group which did not reach abundance levels comparable to healthy controls in remission, with many of them even remaining at the same level as in UC patients.

5.2.4 Proteins with remission abundances between those in active UC and healthy controls

Remission abundances of the proteins discussed in this section lay between those in active UC and healthy controls and were significantly different from both (Groups 6 and 7, Figure 2). This indicates that tissue in remission recovered the respective properties of healthy tissue only partially; hence these properties may have a role in the remission pathophysiology and in relapses.

The abundances of the 135 proteins in group 6 were in remission between the decreased levels in UC and those in healthy controls ((Healthy > Remission > UC).

Metabolic functions were enriched in group 6 which indicates that these functions recover only partially under remission. The abundance decreases affected the metabolism of fatty acids (such as the short-chain fatty acid butyrate), ketone bodies and amino acids as well as the tricarboxcylic acid (TCA) cycle and the electron transport chain (ETC).

Butyrate is the main energy source for colonocytes. Colonocytes metabolize butyrate partially to ketone bodies. A dysfunctional butyrate utilization has previously been identified as a possible factor in UC.[56]

Beside proteins involved in fatty acid metabolism, fatty acid binding proteins (e.g. FABPs) were enriched in group 6. This may explain that peroxisome proliferator-activated receptor (PPAR)-inducible proteins were enriched in this group as well, because FABPs facilitate the transport of fatty acids and thereby contribute to the PPAR signaling.[57, 58]

The partial recovery of functions observed in group 6 also affected the xenobiotic metabolism, such as glucuronidation (e.g. glucuronosyltransferases), sulfation (sulfotransferases), and monoamine oxidation (MAOA).

Moreover, the proteins involved in steroid hormone biosynthesis (e.g. hydroxysteroid dehydrogenases) were enriched in group 6. A role of endogen steroids was suggested earlier as a potential factor in UC, because their metabolism was found to be decreased in UC.(ref paper 2) Our data shows that this decrease persisted to a smaller degree in remission.

Decreases of bile acid transport proteins in active UC were repeatedly shown.(59) (+ref paper 2) The enrichment of these proteins in group 6 show that their abundances in remission lay between those observed in active UC samples and healthy controls.

In group 7 we summarize the 16 proteins with remission abundances between the increased levels of UC and those of healthy controls (Healthy < Remission < UC). Immunoglobulins were enriched in this group, even though group 5 (Healthy < Remission ≈ UC) contains more immunoglobulins.

5.2.5 Proteins differently abundant in healthy controls compared to active UC, but not significantly different abundant compared to remission (group 8)

Group 8 comprises the 33 proteins that are not differently abundant in remission compared to active UC or healthy tissue, but significantly different between active UC and healthy controls. The majority of the protein abundances in group 8 lies in remission tissue between UC and healthy controls, in which they resemble the groups Healthy < Remission < UC and Healthy > Remission > UC, respectively. Accordingly, immunoglobulins are enriched among the proteins in this group which are higher abundant in UC compared to healthy controls. Monocarboxylic acid binding proteins that are related to fatty acid binding are enriched among the proteins with lower abundance in UC compared to healthy tissue, as they are in the group healthy > remission > UC.

5.2.6 Further protein groups and functions of interest

5.2.6.1 S100 proteins

S100 proteins are a family of 24 proteins with EF-hand calcium-binding properties.(60) Decreased abundances of S100 proteins have earlier been found in ulcerative colitis for protein S100-A14 (S100A14) and protein S100-A16 (S100A16). Protein S100-A12 (S100A12), protein S100-A8 (S100A8), and protein S100-A9 (S100A9) are inflammatory proteins and higher abundant in UC than in healthy controls.(5) S100A8 and S100A9 together form the calprotectin complex. Our data shows that the abundances of all these S100 proteins were restored in remission to levels similar to those in healthy controls (groups 2 and 3). Protein S100-P (S100P) was the only of the differently abundant S100 proteins, which under remission retained similar abundances as in active UC. S100P showed similarly high abundances as in active UC, instead of returning to the lower abundances observed in healthy controls. S100P is involved in multiple pathways, e.g. via Receptor for Activated Glycation End Products (RAGE). Most publications on S100P discuss its role of promoting tumor growth, drug resistance and metastasis in various forms of cancer.(60, 61)

5.2.6.2 Carbonic anhydrases

Carbonic anhydrases (CAs) catalyze at physiological pH the hydration of carbon dioxide to hydrogencarbonate and protons. The reaction contributes to the pH homeostasis and the absorption

of salt and water in the intestine.(62) CAs have been found to be enriched among proteins with higher abundance in healthy controls compared to active UC tissue.(ref paper 2) The abundance changes in remission however were less uniform. CA1, CA2, CA4, and CA12 were significantly different abundant in remission tissue compared to healthy controls, active UC, or both. CA1 was similar abundant in remission and healthy controls, but significantly higher abundant in remission compared to active UC. CA4 on the other hand was similar abundant in remission and active UC, but significantly lower abundant compared to healthy controls. The CA2 and CA12 abundances in remission lay between those observed in healthy controls and the decreased abundances in active UC tissue, and they were significantly different from both.

5.3   Foreseeable applications

The major contribution of the present study for future application is the identification of protein functions that remain different in remission tissue compared to healthy controls. This has implications for the prioritization in the development of potential treatments. The functional changes that remain in remission similar to the state in active UC are of particular interest for treatment development, because this persistence indicates that these changes are not a result of the acute inflammation. Instead they may be underlying causes, which makes these functions potential treatment targets for keeping patients in remission. Metallothioneins and prostaglandin metabolism proteins showed no changes in remission compared to active UC. The effects of the abundance decrease of metallothioneins may be ameliorated by chelation of toxic heavy metals, the supplementation of essential metal ions, and the reduction of oxidative stress. Metabolic studies to determine the actual prostaglandin abundance changes could propel the identification of treatment targets related to the changes in the prostaglandin metabolism. Functions that in remission have abundances between those in healthy controls and active UC, e.g. proteins of the nutrient and xenobiotic metabolism, are potentially interesting targets for retaining the remission state as well. The functions that reach abundances as in healthy tissue however seem to be sufficiently addressed with current treatments.

The present study also has implications for future diagnostic developments. As the functions that are similar in active UC and remission are independent of the acute inflammation, it can be hypothesized that these changes may as well occur before the inflammation and could hence serve as early biomarkers for developing UC. However, testing of this hypothesis demands further studies of patients without symptoms and developing UC. Such samples are hard to obtain.

The insights into the UC remission pathophysiology allow for hypothesis generation regarding the role of the respective functions in the pathophysiology and the disease etiology. Hence, the present study has potential to shape the direction of future research on disease mechanisms in UC.

6   Conclusions

Our study presents the first comprehensive functional analysis of the remission proteome of UC and is on the to-date deepest proteome data set. The results indicate that the remission state is in its majority a combination of features from healthy controls and active UC, and that even though the clinical scores of the remission patients resemble the healthy state, the proteome has some important differences. In total 61% of the differently abundant proteins in UC patients compared to controls, returned in remission to levels as observed in healthy controls, e.g. fibrillary collagen and immune system proteins. However, there are also some proteins that do not return to the same levels as controls, e.g. metallothionins, proteins involved in prostaglandin synthesis and immunoglobulins, which in patients in deep remission all had abundancies closer to the patients with active disease than the controls.

These functions that did not recover or that recovered only partly are of particular interest. These may contribute to the observed symptoms in remission and the risk of relapses. These characteristics may cause the transition from uninflamed mucosa to inflammation and are of interest for our understanding of the disease pathophysiology and potentially the etiology of UC.

7   Acknowledgements

8   Conflict of interest

The authors declare no conflict of interest.

9 References

1.	Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology. 2012;142:46-54.e42; quiz e30

2.	Bryant RV, Winer S, Spl T, et al. Systematic review: Histological remission in inflammatory bowel disease. Is 'complete' remission the new treatment paradigm? An IOIBD initiative. Journal of Crohn's and Colitis. 2014;8:1582-1597

3.	Fukuba N, Ishihara S, Tada Y, et al. Prevalence of irritable bowel syndrome-like symptoms in ulcerative colitis patients with clinical and endoscopic evidence of remission: prospective multicenter study. Scandinavian Journal of Gastroenterology. 2014;49:674-680

4.	Bennike TB, Carlsen TG, Ellingsen T, et al. Neutrophil Extracellular Traps in Ulcerative Colitis: A Proteome Analysis of Intestinal Biopsies. Inflamm Bowel Dis. 2015;21:2052-2067

5.	Schniers A, Anderssen E, Fenton Christopher G, et al. The Proteome of Ulcerative Colitis in Colon Biopsies from Adults - Optimized Sample Preparation and Comparison with Healthy Controls. PROTEOMICS - Clinical Applications. 2017;11:1700053

6.	Starr AE, Deeke SA, Ning Z, et al. Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate Crohn's disease from UC. Gut. 2016

7.	Poulsen NA, Andersen V, Møller JC, et al. Comparative analysis of inflamed and non-inflamed colon biopsies reveals strong proteomic inflammation profile in patients with ulcerative colitis. BMC Gastroenterology. 2012;12:76

8.	Deeke SA, Starr AE, Ning Z, et al. Mucosal–luminal interface proteomics reveals biomarkers of pediatric inflammatory bowel disease-associated colitis. The American Journal of Gastroenterology. 2018;113:713-724

9.	Bennike T, Birkelund S, Stensballe A, et al. Biomarkers in inflammatory bowel diseases: Current status and proteomics identification strategies. World Journal of Gastroenterology : WJG. 2014;20:3231-3244

10.	Andersson E, Bergemalm D, Kruse R, et al. Subphenotypes of inflammatory bowel disease are characterized by specific serum protein profiles. PloS one. 2017;12:e0186142

11.	Planell N, Lozano JJ, Mora-Buch R, et al. Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. 2013;62:967-976

12.	Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Letters. 2009;583:3966-3973

13.	Stange EF, Travis SP, Vermeire S, et al. European evidence-based Consensus on the diagnosis and management of ulcerative colitis: Definitions and diagnosis. J Crohns Colitis. 2008;2:1-23

14.	Sutherland LR, Martin F, Greer S, et al. 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. Gastroenterology. 1987;92:1894-1898

15.	Olsen T, Goll R, Cui G, et al. Tissue levels of tumor necrosis factor-alpha correlates with grade of inflammation in untreated ulcerative colitis. Scand J Gastroenterol. 2007;42:1312-1320

16.	Geboes K, Riddell R, Öst A, et al. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut. 2000;47

17.	Scopes RK. Measurement of protein by spectrophotometry at 205 nm. Analytical Biochemistry. 1974;59:277-282

18.	Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech. 2008;26:1367-1372

19.	Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Meth. 2016;13:731-740

20.   Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics (Oxford, England). 2009;25:1091-1093

21.   Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13:2498-2504

22.   Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25:25-29

23.   Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102:15545

24.   Finn RD, Attwood TK, Babbitt PC, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic acids research. 2017;45:D190-D199

25.   Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28:27-30

26.   Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Research. 2014;42:D222-D230

27.   Attwood TK, Beck ME, Bleasby AJ, et al. PRINTS--a database of protein motif fingerprints. Nucleic acids research. 1994;22:3590-3596

28.   Sigrist CJA, de Castro E, Cerutti L, et al. New and continuing developments at PROSITE. Nucleic Acids Research. 2013;41:D344-D347

29.   Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. Nucleic acids research. 2018;46:D649-D655

30.   Schultz J, Milpetz F, Bork P, et al. SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America. 1998;95:5857-5864

31.   The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Research. 2017;45:D158-D169

32.   Habegger KM, Heppner KM, Geary N, et al. The metabolic actions of glucagon revisited. Nature Reviews Endocrinology. 2010;6:689

33.   Patil NA, Rosengren KJ, Separovic F, et al. Relaxin family peptides: structure–activity relationship studies. British Journal of Pharmacology. 2017;174:950-961

34.   Hammond GL. Plasma steroid-binding proteins: primary gatekeepers of steroid hormone action. The Journal of Endocrinology. 2016;230:R13-R25

35.   Jerkovic L, Voegele AF, Chwatal S, et al. Afamin Is a Novel Human Vitamin E-Binding Glycoprotein Characterization and In Vitro Expression. Journal of Proteome Research. 2005;4:889-899

36.   Naylor HM, Newcomer ME. The Structure of Human Retinol-Binding Protein (RBP) with Its Carrier Protein Transthyretin Reveals an Interaction with the Carboxy Terminus of RBP. Biochemistry. 1999;38:2647-2653

37.   Hamilton JA, Benson MDJC, CMLS MLS. Transthyretin: a review from a structural perspective. 2001;58:1491-1521

38.   Huntington JA. Serpin structure, function and dysfunction. Journal of thrombosis and haemostasis : JTH. 2011;9 Suppl 1:26-34

39.   Borel F, Sun H, Zieger M, et al. Editing out five Serpina1 paralogs to create a mouse model of genetic emphysema. 2018

40.   Baker C, Belbin O, Kalsheker N, et al. SERPINA3 (aka alpha-1-antichymotrypsin). Frontiers in bioscience : a journal and virtual library. 2007;12:2821-2835

41.   Pham CTN. Neutrophil serine proteases: specific regulators of inflammation. Nature Reviews Immunology. 2006;6:541

42.   Chai KX, Chen LM, Chao J, et al. Kallistatin: a novel human serine proteinase inhibitor. Molecular cloning, tissue distribution, and expression in Escherichia coli. 1993;268:24498-24505

43.   Colman RW, Sartor RB, Adam AA, et al. The plasma kallikrein-kinin system in sepsis, inflammatory arthritis, and enterocolitis. Clinical reviews in allergy & immunology. 1998;16:365-384

44.   Mahley RW, Innerarity TL. Lipoprotein receptors and cholesterol homeostasis. Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes. 1983;737:197-222

45.   Bella J, Hulmes DJS. Fibrillar Collagens. In: Parry DAD, Squire JM, eds. Fibrous Proteins: Structures and Mechanisms. Cham: Springer International Publishing; 2017:457-490

46.   Zhu A, Wang ZK, Beavis R. Structural Studies of α-N-Acetylgalactosaminidase: Effect of Glycosylation on the Level of Expression, Secretion Efficiency, and Enzyme Activity. Archives of Biochemistry and Biophysics. 1998;352:1-8

47.   Willems PJ, Seo H-C, Coucke P, et al. Spectrum of mutations in fucosidosis. European Journal Of Human Genetics. 1999;7:60

48.   Wang AM, Schindler D, Desnick R. Schindler disease: the molecular lesion in the alpha-N-acetylgalactosaminidase gene that causes an infantile neuroaxonal dystrophy. The Journal of Clinical Investigation. 1990;86:1752-1756

49.   Cariappa A, Takematsu H, Liu H, et al. B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. 2009;206:125-138

50.   Leonard N, Bishop AE, Polak JM, et al. Expression of nitric oxide synthase in inflammatory bowel disease is not affected by corticosteroid treatment. Journal of Clinical Pathology. 1998;51:750-753

51.   Ruttkay-Nedecky B, Nejdl L, Gumulec J, et al. The Role of Metallothionein in Oxidative Stress. 2013;14:6044

52.   Penning TM, Byrns MC. Steroid Hormone Transforming Aldo-Keto Reductases and Cancer. 2009;1155:33-42

53.   Cho H, Huang L, Hamza A, et al. Role of glutamine 148 of human 15-hydroxyprostaglandin dehydrogenase in catalytic oxidation of prostaglandin E2. Bioorganic & Medicinal Chemistry. 2006;14:6486-6491

54.   Van Veldhoven PP, Vanhove G, Assselberghs S, et al. Substrate specificities of rat liver peroxisomal acyl-CoA oxidases: palmitoyl-CoA oxidase (inducible acyl-CoA oxidase), pristanoyl-CoA oxidase (non-inducible acyl-CoA oxidase), and trihydroxycoprostanoyl-CoA oxidase. 1992;267:20065-20074

55.   Diczfalusy U, Alexson SE. Identification of metabolites from peroxisomal beta-oxidation of prostaglandins. 1990;31:307-314

56.   Roediger WEW. The colonic epithelium in Ulcerative-Colitis - An Energy-Deficiency Disease. Lancet. 1980;2

57.   Wolfrum C, Borrmann CM, Börchers T, et al. Fatty acids and hypolipidemic drugs regulate peroxisome proliferator-activated receptors α- and γ-mediated gene expression via liver fatty acid binding protein: A signaling path to the nucleus. 2001;98:2323-2328

58.   Tan N-S, Shaw NS, Vinckenbosch N, et al. Selective Cooperation between Fatty Acid Binding Proteins and Peroxisome Proliferator-Activated Receptors in Regulating Transcription. Molecular and Cellular Biology. 2002;22:5114-5127

59.   Jahnel J, Fickert P, Hauer AC, et al. Inflammatory bowel disease alters intestinal bile acid transporter expression. Drug metabolism and disposition: the biological fate of chemicals. 2014;42:1423-1431

60.   Donato R, Cannon BR, Sorci G, et al. Functions of S100 Proteins. Current molecular medicine. 2013;13:24-57

61.   Prica F, Radon T, Cheng Y, et al. The life and works of S100P - from conception to cancer. American Journal of Cancer Research. 2016;6:562-576

62.   Kivelä AJ, Kivelä J, Saarnio J, et al. Carbonic anhydrases in normal gastrointestinal tract and gastrointestinal tumours. World Journal of Gastroenterology : WJG. 2005;11:155-163

Paper IV

Prediction of the 1-year outcome for Ulcerative Colitis from the proteomic profile of treatment naïve patients

Armin Schniers[1], Rasmus Goll[2,3], Sveinung Wergeland Sørbye[4], Hans Stenlund[5], Jon Florholmen[2,3], Terkel Hansen[1]*

[1]Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[2]Department of Medical Gastroenterology, University Hospital of North Norway, Tromsø, Norway

[3]Gastroenterology and Nutrition Research Group, Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[4]Department of Clinical Pathology, University Hospital of North Norway, Tromsø, Norway

[5]Swedish Metabolomics Centre, Umeå, Sweden.

Abstract

Clinicians have several options to treat ulcerative colitis, dependent on the severity of the disease. So far, the most used approach is a step-up approach based on the clinical findings and the patient reported problems. However, there is some evidence that treatment with TNF-alpha inhibitors at an early stage improves the outcome for patients with a severe course of the disease. However, predicting the outcome for the patients has so far been difficult. We present a statistical model that includes 50 proteins for predicting the 1-year outcome from proteomic data. The data was obtained from 19 previously untreated ulcerative colitis patients. The statistical model was built through selecting the most important proteins from a PLS model, and further validated through building a PCA model from these proteins. The proteins used in the statistical model show biological importance with regard to inflammation, which strengthens the validity of the statistical model.

Introduction

UC has a complex and not fully understood pathogenesis that involves genetic factors, environmental influences and epithelial barrier defects resulting in a dysregulated immune response.[1,2] The treatment aim in UC is to achieve remission. Remission can be defined in several different ways, but most common is the resolution of clinical symptoms and endoscopic mucosa healing. However, these two do not necessarily correlate well.[2–5] The resolution of clinical symptoms is important for the patients' quality of life while mucosal healing is associated with improved long-term clinical remission.[6] Therefor endoscopic mucosa healing must be assessed independently of the symptoms.

The treatment approach has recently been under debate, and several clinicians now report that patients can benefit from an early aggressive treatment with biological TNF-alpha inhibitors instead of the standardized step-up treatment.[7] The step-up treatment, which starts with mild treatment like 5-ASA before stronger immunomodulators or biological TNF-alpha inhibitors are used, seems to work best for patients with a mild to moderate course of the disease. However, for the patients that have a more severe course of the disease the data released to-date seems to indicate that a top-down approach, i.e. an aggressive biological anti TNF-alpha treatment, seems to be the best choice.[8] It is currently difficult to identify which patients will have a mild and which will have a severe course at an early stage of the disease.[9]

Biomarkers are measurable bio-indicators of specific biological states and can be used clinically in diagnosis of diseases, monitoring disease activity, guiding the therapy, to assess the therapeutic response, or to predict the disease progression.[10,11] Efforts of biomarker discovery in UC with diagnostic purpose have so far been mainly aimed to differentiate UC from CD or healthy controls.[11] This is extremely important, both clinically and in order to understand the pathology of the disease. However, clinically more important would be a biomarker that could predict the further course of the disease or the response to treatment.[11] This biomarker could be a DNA mutation, an mRNA or protein, a small molecule, or a combination of any of the mentioned. Ideally, biomarkers should be easily obtainable with non-invasive methods, their measurement should be cost-effective and they should, most importantly, be both sensitive and specific.

Most potential clinical applications of proteomics are believed to require the assessment of multiple proteins. The rationale behind this is that a single protein has multiple functions in a cell and interacts with other proteins to exert these functions, and that as a result a change in the abundance of one protein may only be relevant when other protein abundances change as well.[11] A common challenge when trying to investigate prognostic disease markers is to separate the markers for the disease from the biological changes induced by treatment of the disease.

In the following publication, we present the first prognostic model for predicting the 1-year outcome for patients with ulcerative colitis based on proteomics data.

Methods

Patient data:

Table 1: Patient data. *: indicates significant difference between groups (p<0.05)

1-year outcome	Average age (Range)	TNF-α (Range)	Average Geboes index (Range)*	Sex (F/M)
Mild/Moderate	40.3 (20-62)	11,315 (4,600-31,700)	6.5 (3-13)	4/10
Severe	40.0 (18-68)	37,880 (17,200-96,000)	12.0 (8-16)	1/4

Biopsy lysis:

Colon biopsies were lyzed in 250 µl cold buffer (5% sodium deoxycholate (SDC), 8 M urea, 100 mM triethylammonium bicarbonate buffer pH 8.5 (TEAB)) in MagNA Lyser Green Bead tubes (Roche Diagnostics AG, Rotkreuz, Switzerland) with a MagNA Lyser Instrument (Roche Diagnostics AG, Rotkreuz, Switzerland) for 35 seconds at 6500 rpm. The samples were frozen and stored at -70 °C until further preparation.

Protein concentration measurement:

The protein concentration was determine with a bicinchoninic acid (BCA) assay (Pierce™ BCA Protein Assay Kit, Thermo Fisher Scientific) according to the manufacturer's protocol. The TEAB concentration for the BCA assay in standards and samples was 5 mM TEAB.

Reduction, alkylation, and quenching of alkylating agent:

1,4-dithiothreitol (DTT) was added to a concentration of 5 mM to 60 µg protein aliquots in Protein LoBind tubes (Eppendorf AG, Hamburg, Germany). Cysteines were reduced under incubation at 54 °C for 30 min. The alkylating agent iodoacetamide (IAA) was added to a concentration of 15 mM and cysteines were alkylated during incubation for 30 min at room temperature in the dark. DTT was added to a concentration of 5 mM to quench remaining IAA.

Lys-C/trypsin tandem digestion:

0.6 µg endoproteinase Lys-C (Wako Chemicals GmbH, Neuss, Germany) were added for a 1:100 enzyme-to-protein ratio (w/w). The sample compositions were adjusted to 1 mM calcium chloride, 6 M urea and 100 mM TEAB with calcium chloride solution, water, and 1 M TEAB. The Lys-C predigestion was facilitated under gentle agitation for 8 hours at 37°C.

3 µg trypsin (Sequencing Grade Modified Trypsin, Promega Corporation, Madison, USA) were added to the predigested samples for a 1:20 enzyme-to-protein ratio (w/w). The sample compositions were adjusted to 1 mM calcium chloride, 1 M urea, and 100 mM TEAB with calcium chloride solution, water, and 1 M TEAB. Tryptic digestion was facilitated for 16 hours at 37°C under gentle agitation.

TMT labelling:

Equal peptide amounts from each sample were mixed to produce a standard. This standard mixture was labeled with the tandem mass tag (TMT) reagent TMT6-126 as described in the manufacturer's protocol (Thermo Fisher Scientific, Rockford, USA).

For each sample, 25 µg peptides were labeled with the remaining TMTsixplex reagents as described in the manufacturer's protocol. 0.4 mg TMT reagent were used to label 25 µg peptides, if the starting protein concentration was ≥ 2200 µg/ml in the respective sample. Samples with lower starting concentrations were labeled with more TMT reagent to achieve complete labeling.

The labeled standard and 5 differently labeled samples were mixed in equal amounts to in total 100 µg peptides for each fractionation mixture.

Detergent removal and high pH reversed-phase fractionation:

The detergent SDC was precipitated by adding 50% formic acid (FA) solution to a final concentration of 2.5% FA. A pH ≤ 2 was confirmed with pH indicator paper. The samples were centrifuged at 16,000 g for 15 min and the supernatant was transferred to new Protein LoBind tubes to separate it from the pelleted detergent. Acetonitrile, which was added to the samples during TMT labeling, was removed in a vacuum concentrator because it would interfere with the fractionation procedure.

The samples were adjusted to a concentration of 0.1% TFA by adding 10% trifluoroacetic acid (TFA) solution. The fractionation into 8 fractions was performed with Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher, Rockford, USA) as described in the manufacturer's protocol. Samples exceeding 300 µl were loaded stepwise onto the fractionation columns by centrifugation for 2 minutes after each loading. The collected fractions were dried in a vacuum concentrator and redissolved in 10 µl 0.1% TFA for subsequent LC-MS/MS analyses.

The absorption of these solutions at 205 nm was measured with a Nanodrop 2000 (Thermo Fisher Scientific, Bremen, Germany). The corresponding peptide concentration was calculated based on an extinction coefficient of 31 mg/ml.[12]

LC-MS/MS:

The fractions were analyzed with an EASY-nLC 1000 coupled to a Q Exactive mass spectrometer (both Thermo Fisher Scientific, Bremen, Germany).

2.0 µg peptides of each fraction were injected and concentrated on a reversed-phase trap column. Separation was performed with a gradient of 0.1% FA in water and 0.1% FA in acetonitrile on a reversed-phase main column. The proportion of the acetonitrile solvent was increased over 19 min from 0% to 5%, to 30% at 180 min, and up to 100% at 200 min. More details on the LC-MS/MS settings are provided in the supporting information.

Processing of mass spectrometry data:

Proteins were identified and quantified from the mass spectrometry data with MaxQuant.[13]

Protein intensities were log2-transformed in Perseus.[14] The identifications with labels "potential contaminant", "reverse" or "only identified by site" were removed. The standard intensities were subtracted from the sample intensities of the same experiment and the resulting intensities were Z-score normalized (matrix access: columns).

Multivariate analysis of data:

All statistical multivariate handling of data was conducted in SIMCA P+ 15.0 (Umetrics, Sweden). Normalized protein levels were imported and all proteins with missing data were removed from the dataset before further data handling.

Ethical considerations:

Informed and written consent forms were signed by all enrolled subjects. The Regional Committee of Medical Ethics of North Norway and the Norwegian Social Science Data Services approved the study and the storage of biological material under the number REK NORD 2012/1349.

Results and discussion

Pinpointing the important features in a prospective omics analysis is challenging and requires strict evaluation for the clinical usefulness of the biomarkers. The process from the first promising data to the ready-to-use is long with many obstacles. However, we present in this article the first promising prospective results from a proteomics study from colon biopsies of IBD patients.
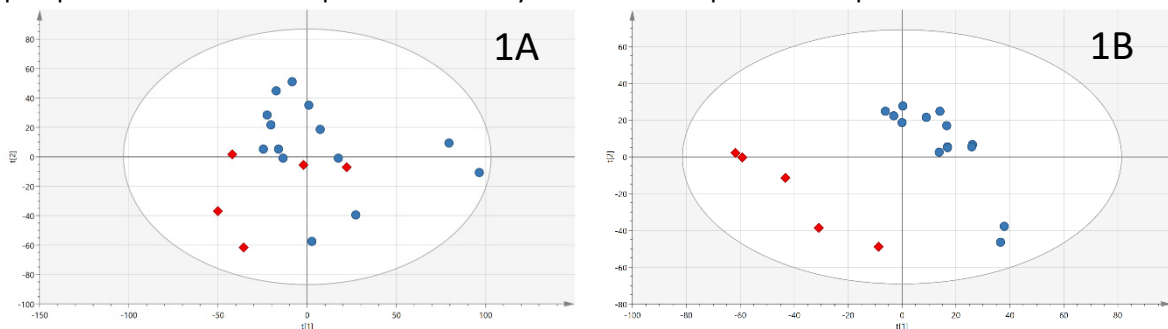

Figure 1: Principal component analysis (PCA) (1A) and Partial least squares regression analysis (PLS) (1B) of samples with all identified proteins taken into account, blue circles indicate mild or moderate outcome after one year, red diamonds indicate severe outcome after one year.

In an unsupervised PCA model (figure 1A) the data is quite spread, however there is a trend that the patients with severe 1-year outcome seems to cluster slightly more towards the lower left side than the rest. This indicates that there are generally differences expressed in patients that will have a more severe course of the disease already at time of diagnosis. In a supervised PLS model, where we want to separate the group with the severe 1-year outcome from those with mild to moderate outcome, the separation between the groups becomes very clear (figure 1B). The model also shows excellent separation of the patient's outcome severity (Y-variable) with a R2Y of 0.995, and an acceptable predictive Q2 value of 0.673. However, as the model is based on a total of almost 8.500 proteins, there is a necessity to reduce the number of proteins before the data can be utilized in a clinical environment. In addition, most of the identified proteins are probably not important for the statistical model; hence, these proteins contribute only noise, but no additional predictive strength, to the model.

To circumvent the problems with noise and to reduce the number of proteins, several PLS models were built with different numbers of the most important proteins, choosing the most important proteins in the PLS model in figure 1B. Reducing the number of variables in a predictive model is a trade-off between reducing noise and keeping important information. In this dataset, it seems that the 50 most important proteins are essential for building the best predictive model. With the resulting PLS model (Figure 2A), we see an excellent separation of the 1-year outcome status of the patients. As seen in the graphical representation of the model, the patients are clearly separated by the t1 compartment. Upon reduction of the number of proteins from approximately 8.500 to 50, the predictive Q2 value increased from 0.673 to 0.869. The relatively high Q2 value testifies that the model has good predictive power for the 1-year outcome.

In order to investigate the model further, we also created unsupervised PCA plots of the included patients, with only the 50 most important proteins taken into account (figure 2B). The resulting unsupervised PCA shows that the groups open up a bit more, especially the group with mild to moderate outcome, however, the separation is still excellent which strengthens the postulation that this model can predict the 1-year outcome of IBD patients.
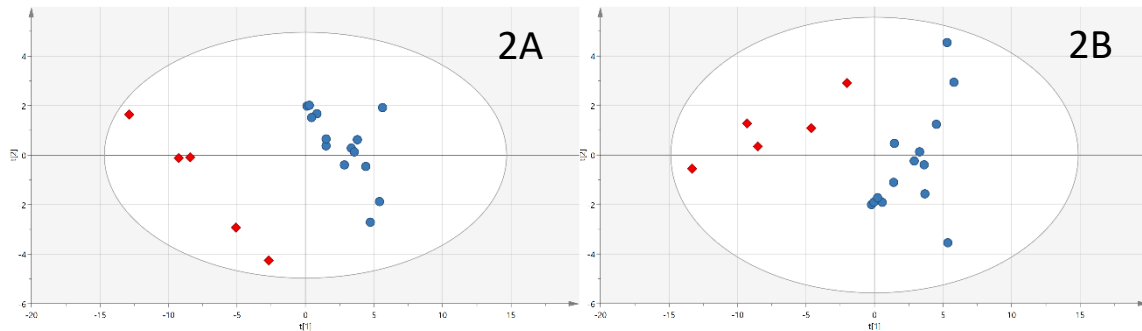
Figure 2: Partial least squares regression analysis (PLS) (2A) and Principal component analysis (PCA) (2B) of samples with the 50 most important proteins taken into account, blue circles indicate mild or moderate outcome after one year, red diamonds indicate severe outcome after one year.

Biological relevance of proteins:

The 50 most important proteins from the model shown in figure 1B, and further used in the models presented in figure 2, show distinct networks when processed through functional processing association network searches. In general there are two separate networks (figure 3), where the major network consists of 3 interlinked networks (Figure 3, A, B and C).
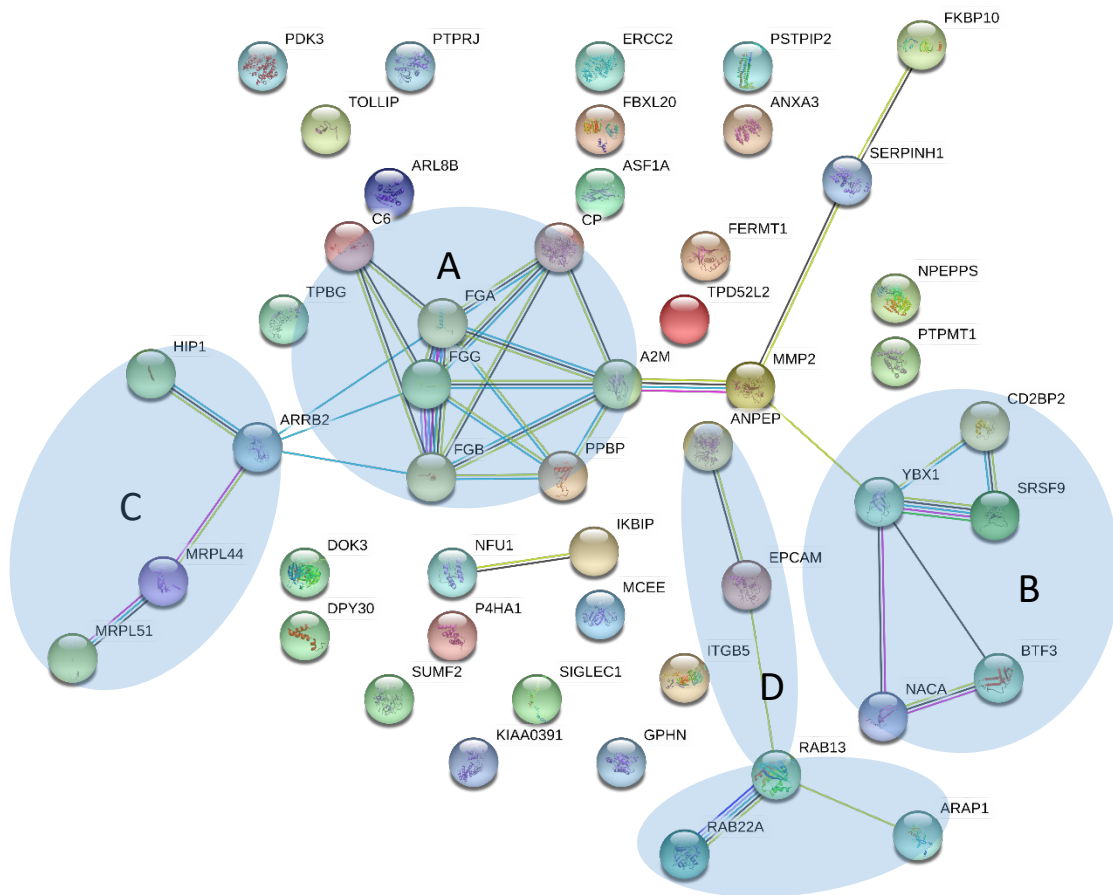

Figure 3: Network analysis of the 50 most important proteins in the PLS model. Interactive link: https://version-11-0.string-db.org/cgi/network.pl?networkId=XFQRhO2seBbQ

The core of the network, labelled A in figure 3, is dominated by the fibrinogen proteins Fibrinogen Alpha, Beta and Gamma chain (FGA, FGB and FGG). Fibrinogen is most commonly recognized for the involvement in blood clotting and wound healing[15] as an acute phase protein.

However, fibrinogen has been linked to inflammation as well and has been reported to have pro-inflammatory properties in several diseases, e.g. multiple sclerosis, Alzheimer's disease, rheumatoid arthritis, kidney fibrosis and several types of cancer.[16] The fibrins are also linked to other well-known parts of the immune system, like complement component C6 (C6), a part of the adaptive immune response, which forms pores in the cell membrane of target cells. In addition, the fibrinogens are interlinked with the glycoprotein Ceruplasmin (CP), known for oxidizing cupper to Cu^{2+}, as well as platelet basic protein (PPBP), which among many other functions is known to stimulate the formation and secretion of plasminogen activator.

Network A is linked through a metalloproteinase (MMP2) to the network labelled B, which consists of proteins involved in splicing (YBX1, CD2BP2 and SRSF9) and other proteins involved in translation of mRNA to proteins (NACA and BTF3). The cold-shock protein YBX1 has been shown to be a key protein in inflammation through regulation of interleukin release, and has been linked to resistance against Janus kinase (JAK) inhibitors.

Network C consists of 4 proteins, all located in the mitochondria of the cells. In network C there are two upregulated proteins in patients with severe compared to mild or moderate outcome, β-Arrestin2 (ARRB2) which is a G-protein coupled receptor signalling for various cytoplasmic signalling within inflammation[17] and Huntington interacting protein (HIP1) which has been shown to regulate arthritis severity.[18] In addition, two 39S ribosomal proteins that interact with β-Arrestin2 are downregulated, MRPL44 and MRPL51. MRPL44 has previously been shown to be involved in tumour development.[19] MRPL51 on the other hand has previously been reported in a dextran sodium sulphate (DSS)-induced colitis model, where MRPL51 was shown to be regulated at transcript level but without being able to explain the role of the protein in IBD.[20]

The independent network labelled D has two Rab proteins at its core, Rab13 and Rab22A, which have their main importance in membrane trafficking. Both Rab13 and Rab22A have been reported to be involved in inflammation as well as inflammation-related diseases like cancer.[21,22] Mistargeting of Rab13 to basolateral sites has been observed in patients with Crohn´s Disease.[23] Network D seems to influence permeability of the intestine, which is known to be dysregulated in patients with IBD. These data indicate that intestinal permeability is also of importance for predicting the outcome of the patients.

In addition to the proteins described above as members of the networks, there are other proteins that are of importance for the clustering of patients. The majority of these proteins seems to be involved in immune system or energy metabolism pathways. In general, the biological functions of the proteins seem to fit to the severity of IBD, which is strengthening the background of the statistical model for predicting the outcome of IBD patients.

Conclusion

In conclusion, we created a promising statistical model for the prediction of outcome for newly diagnosed UC patients. The predictive statistics seem to be convincing and the biological functions of the most important proteins can be explained by the known pathology of the disease, which further strengthens the statistical model. However, there is a need to validate the model further by new samples and an orthogonal quantification method.

References
(1)   Danese, S.; Sans, M.; Fiocchi, C. Inflammatory Bowel Disease: The Role of Environmental Factors. *Autoimmunity Reviews*. 2004.
(2)   Ungaro, R.; Mehandru, S.; Allen, P. B.; Peyrin-Biroulet, L.; Colombel, J.-F. Ulcerative Colitis. *Lancet* **2017**, *389* (10080), 1756–1770.
(3)   Jharap, B.; Sandborn, W. J.; Reinisch, W.; D'Haens, G.; Robinson, A. M.; Wang, W.; Huang, B.; Lazar, A.; Thakkar, R. B.; Colombel, J. F. Randomised Clinical Study: Discrepancies between Patient-Reported Outcomes and Endoscopic Appearance in Moderate to Severe Ulcerative Colitis. *Aliment. Pharmacol. Ther.* **2015**.
(4)   Falvey, J. D.; Hoskin, T.; Meijer, B.; Ashcroft, A.; Walmsley, R.; Day, A. S.; Gearry, R. B. Disease

Activity Assessment in IBD: Clinical Indices and Biomarkers Fail to Predict Endoscopic Remission. *Inflamm. Bowel Dis.* **2015**.

(5) Regueiro, M.; Rodemann, J.; Kip, K. E.; Saul, M.; Swoger, J.; Baidoo, L.; Schwartz, M.; Barrie, A.; Binion, D. Physician Assessment of Ulcerative Colitis Activity Correlates Poorly with Endoscopic Disease Activity. *Inflamm. Bowel Dis.* **2011**.

(6) Shah, S. C.; Colombel, J. F.; Sands, B. E.; Narula, N. Mucosal Healing Is Associated With Improved Long-Term Outcomes of Patients With Ulcerative Colitis: A Systematic Review and Meta-Analysis. *Clinical Gastroenterology and Hepatology*. 2016.

(7) Masond, M. S.; Abd El-Hamid, O. H. Structural Chemistry of Amino Acid Complexes. *Transit. Met. Chem.* **1989**, *14* (3), 233–234.

(8) Sandborn, W. J. Step-up versus Top-down Therapy in the Treatment of Ulcerative Colitis. *Gastroenterol. Hepatol. (N. Y).* **2007**, *3* (1), 16–17.

(9) Devlin, S. M.; Panaccione, R. Evolving Inflammatory Bowel Disease Treatment Paradigms: Top-Down Versus Step-Up. *Med. Clin. NA 94*, 1–18.

(10) Rifai, N.; Gillette, M. A.; Carr, S. A. Protein Biomarker Discovery and Validation: The Long and Uncertain Path to Clinical Utility. *Nature Biotechnology*. 2006.

(11) Gisbert, J. P.; Chaparro, M. Clinical Usefulness of Proteomics in Inflammatory Bowel Disease: A Comprehensive Review. *J. Crohn's Colitis* **2018**.

(12) Scopes, R. K. Measurement of Protein by Spectrophotometry at 205 Nm. *Anal. Biochem.* **1974**, *59* (1), 277–282.

(13) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.

(14) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat Meth* **2016**, *13* (9), 731–740.

(15) Kattula, S.; Byrnes, J. R.; Wolberg, A. S. Fibrinogen and Fibrin in Hemostasis and Thrombosis. *Arterioscler. Thromb. Vasc. Biol.* **2017**, *37* (3), e13–e21.

(16) Davalos, D.; Akassoglou, K. Fibrinogen as a Key Regulator of Inflammation in Disease. *Semin. Immunopathol.* **2012**, *34* (1), 43–62.

(17) Gaffal, E.; Jakobs, M.; Glodde, N.; Schröder, R.; Kostenis, E.; Tüting, T. β-Arrestin 2 Inhibits Proinflammatory Chemokine Production and Attenuates Contact Allergic Inflammation in the Skin. *J. Invest. Dermatol.* **2014**, *134* (8), 2131–2137.

(18) Laragione, T.; Brenner, M.; Lahiri, A.; Gao, E.; Harris, C.; Gulko, P. S. Huntingtin-Interacting Protein 1 (HIP1) Regulates Arthritis Severity and Synovial Fibroblast Invasiveness by Altering PDGFR and Rac1 Signalling. *Ann. Rheum. Dis.* **2018**, *77* (11), 1627–1635.

(19) Lee, Y.-K.; Lim, J. J.; Jeoun, U.-W.; Min, S.; Lee, E.-B.; Kwon, S. M.; Lee, C.; Yoon, G. Lactate-Mediated Mitoribosomal Defects Impair Mitochondrial Oxidative Phosphorylation and Promote Hepatoma Cell Invasiveness Downloaded From. **2017**.

(20) Fang, K.; Bruce, M.; Pattillo, C. B.; Zhang, S.; Stone, R.; Clifford, J.; Kevil, C. G.; Kevil, C. G. Temporal Genomewide Expression Profiling of DSS Colitis Reveals Novel Inflammatory and Angiogenesis Genes Similar to Ulcerative Colitis. *Physiol. Genomics* **2011**, *43* (1), 43–56.

(21) Campos, J. H.; Soares, R. P.; Ribeiro, K.; Andrade, A. C.; Batista, W. L.; Torrecilhas, A. C. Extracellular Vesicles: Role in Inflammatory Responses and Potential Uses in Vaccination in Cancer and Infectious Diseases. *J. Immunol. Res.* **2015**, *2015*, 832057.

(22) Prashar, A.; Schnettger, L.; Bernard, E. M.; Gutierrez, M. G. Rab GTPases in Immunity and Inflammation. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 435.

(23) Ohira, M.; Oshitani, N.; Hosomi, S.; Watanabe, K.; Yamagami, H.; Tominaga, K.; Watanabe, T.; Fujiwara, Y.; Maeda, K.; Hirakawa, K.; et al. Dislocation of Rab13 and Vasodilator-Stimulated Phosphoprotein in Inactive Colon Epithelium in Patients with Crohn's Disease. *Int. J. Mol. Med.* **2009**, *24* (6), 829–835.