

Metaproteomics of fecal samples of Crohn's disease and Ulcerative Colitis

T. Schlegel^a, K. Schallert^a, R. Vilchez-Vargas^b, D. Benndorf^{a, c}, S. Püttker^a, S. Sydor^b, C. Schulz^{b, d}, L. Bechmann^b, A. Canbay^b, B. Heidrich^e, U. Reichl^{a, c}, A. Link^b, R. Heyer^{a, *}

^a Bioprocess Engineering, Otto von Guericke University, Universitätsplatz 2, Magdeburg 39106, Germany

^b Department of Gastroenterology, Hepatology and Infectious Diseases, Otto von Guericke University, Magdeburg Leipzig Str. 44, Magdeburg 39120, Germany

^c Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, Magdeburg 39106, Germany

^d Department of Medicine II, University Hospital, LMU Munich, Marchioninstr. 15, Munich 81377, Germany

^e Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Carl-Neuberg Str. 1, Hannover 30625, Germany

ABSTRACT

Crohn's Disease (CD) and Ulcerative Colitis (UC) are chronic inflammatory bowel diseases (IBD) of the gastrointestinal tract. This study used non-invasive LC-MS/MS to find disease specific microbial and human proteins which might be used later for an easier diagnosis.

Therefore, 17 healthy controls, 11 CD patients and 14 UC patients but also 13 Irritable Bowel Disease (IBS) patients, 8 Colon Adenoma (CA) patients, and 8 Gastric Carcinoma (GCA) patients were investigated. The proteins were extracted from the fecal samples with liquid phenol in a ball mill. Subsequently, the proteins were digested tryptically to peptides and analyzed by an Orbitrap LC-MS/MS. For protein identification and interpretation of taxonomic and functional results, the MetaProteomeAnalyzer software was used.

Cluster analysis and non-parametric test (analysis of similarities) separated healthy controls from patients with CD and UC as well as from patients with GCA. Among others, CD and UC correlated with an increase of neutrophil extracellular traps and immune globulins G (IgG). In addition, a decrease of human IgA and the transcriptional regulatory protein RprY from *Bacillus fragilis* was found for CD and UC. A specific marker in feces for CD was an increased amount of the human enzyme sucrose-isomaltase.

Significance: Crohn's Disease and Ulcerative Colitis are chronic inflammatory diseases of the gastrointestinal tract, whose diagnosis required comprehensive medical examinations including colonoscopy. The impact of the microbial communities in the gut on the pathogenesis of these diseases is poorly understood. Therefore, this study investigated the impact of gut microbiome on these diseases by a metaproteome approach, revealing several disease specific marker proteins. Overall, this indicated that fecal metaproteomics has the potential to be useful as non-invasive tool for a better and easier diagnosis of both diseases.

1. Introduction

Intestinal microbiota participates in food degradation and synthesize several vitamins. Furthermore, the microbiome in acts with the host, directly influencing its immune system [1]. The gastrointestinal

tract (GIT) is colonized by about 10^{13} – 10^{14} microorganisms [2,3]. As an accepted concept homeostasis between microbiota and the immune system is a precondition for human health [4]. However, the GIT could be affected by several diseases such as inflammatory bowel disease (IBD) [5,6], which pathophysiology is linked with an overactivation of the immune system.

Crohn's Disease (CD) and Ulcerative Colitis (UC) as the two main types of IBD are chronic, recurrent inflammations with different clinical, morphological and histopathological pattern. Their prevalence surpass 0.3% in western countries with an increase also in industrialized countries whose societies have become more westernized [7]. CD causes transmural, granulomatous inflammations, which can occur discontinuously in the entire GIT. UC causes superficial inflammation, bleeding and mucosa atrophy in the distal rectum and the colon [8]. The recurrent inflammation in both diseases counts as a risk factor for premalignant transformation. Patients with IBD are under increased risk to develop colon adenoma (CA) and colorectal carcinomas [9] diminishing during that last decades in western countries. In contrast to CD and UC, irritable bowel syndrome (IBS) is a chronic functional gastrointestinal disorder dominated by abdominal discomfort and altered bowel habit which is not linked with structural or biochemical changes [10].

Unfortunately, IBD comprise different diseases with multifactorial pathogeneses, which are currently poorly understood. Genetic alterations represent a risk factor for IBD. Additionally dysfunction in the intestinal inflammatory cascade [4] and environmental factors such as smoking and increased hygienic standards are mentioned to be involved in the pathogenesis, too. Furthermore, several studies [4,11–13] showed the correlation of IBD with a dysbiosis in the microbiome. Frank et al. 2007 [11] observed a reduction of short chain fatty acid producing bacteria in patients with IBD but it remains unclear whether the dysbiosis is a risk factor for the disease or a result. In contrast to IBD, for mono-infections caused by bacteria such as *Helicobacter pylori* it is known that they have the potency to trigger precancerous lesions and thus, the development of Colon Adenoma (CA) or Gastric Carcinoma (GCA) [14].

The diagnosis of IBD is based on endoscopy as well as on clinical presentation, histopathological and laboratory findings. In addition, physicians may support the diagnosis by measuring single protein markers such as calprotectin and lactoferrin as unspecific inflammatory indicators [15]. Unfortunately, these tests focus on surrogate parameters for unspecific inflammation and do not allow to discriminate between bacterial inflammation and IBD.

Within the last years the development of high throughput methods to identify genes (metagenomics/metatranscriptomics) [16] or proteins (metaproteomics) [17–19] enabled the examination of the taxonomic and functional composition of the microbial communities in the human gut. Thereby, metagenomics/metatranscriptomics focus on the taxonomic and functional inventory of the microbial communities in the human. In contrast metaproteomics studies the expression of the proteins from the microbial communities and from the host. Identification of host proteins provides additional knowledge about the patients' health status, e.g. by monitoring proteins associated to the immune system or secreted enzymes from liver and pancreas, or apoptotic cells from the surface of the digestion system.

Up to now, several alterations within the taxonomic and functional composition of the microbiome of patients with IBD were observed based on metagenome and metaproteome studies [4,11–13,18,20,21]. For instance, Casen et al. 2015 [22] correlated CD with a shift from predominantly beneficial bacteria to potentially pathogenic bacteria. However, our understanding of the pathogenesis of IBD is still limited due to the complexity of the diseases and of the microbiome. Most of the studies investigating the microbiome of patients with IBD focused only on one specific disease. This study investigated 71 fecal metaproteomes from healthy individuals and patients with CD, UC, IBS, CA and GCA. The aim of this experimental design was to identify universal marker proteins for the diseases in samples from a representative clinical background and to proof whether non-invasive metaproteome analysis may distinguish between the different diseases.

2. Methods

Ethics statement

The ethics committees at the Otto-von-Guericke University, Magdeburg, Germany and the Hannover Medical School approved this study by an amendment to previous studies (Number 42/08, Number 47/15 and 2087-2013). The study was performed in accordance with the Declaration of Helsinki. All patients received comprehensive information about the studies and gave their written consents.

2.1. Subject population and samples

The patient cohort comprised 37 females and 34 males between 20 and 89 years of age with a median age of 54 years (Supplementary Table 1). The body mass index (BMI) ranged from 17.7–48.9 kg m⁻² with the median at 25.8 kg m⁻². Sixteen participants had an allergy and seven were smokers. A comprehensive study of confounding factors could be found in Supplementary Fig. 2.

The samples were taken at the university hospital Magdeburg or Hannover and stored immediately at –20 °C. Of the total of 71 samples, 17 were people without any gastrointestinal symptoms (named in the following healthy individuals), 11 were patients with CD, 6 were patients with UC in remission stage (UCr), 8 were patients with UC in active stage (UCa), 13 were patients with IBS, 8 were patients with CA and 8 were patients with GCA. The diseases were diagnosed by a clinic physician in accordance with general guidelines. Patients with UC were classified as UCa and UCr based on the Mayo-Score as previously described with remission defined by DAI ≤5 [23]. The concentration of fecal calprotectin was measured by enzyme-linked immune sorbent assay (ELISA) (BÜHLMANN fCAL™, Bühlmann Laboratories AG, Schönenbuch, Switzerland/ PhilCal Calprotectin ELISA MRP8/14 S100A8/9, Immundiagnostik, Bensheim, Germany) as indicator for the level of gut inflammation. A level higher than 200 µg g⁻¹_{feces} was taken as indication threshold for an active inflammation of the large bowel.

2.2. Metaproteomics workflow

All used chemicals were at least of analysis grade purity and the solvents used for LC-MS/MS were of MS grade purity.

Protein extraction was carried out as described by Heyer et al. (2013) [24] and the LC-MS/MS measurement as described by Heyer et al. (2016) [25]. In brief, proteins from approx. 200 mg fecal samples were extracted by combined cell lysis and phenol extraction in a ball mill (FastPrep-96, MP Biomedicals, Eschwege, Germany). Extracted proteins were subsequently dissolved in 100 µL to 790 µL buffer (7 M urea, 2 M thiourea and 0.01 g mL⁻¹ 1, 4-dithiothreitol). Protein amounts were quantified by amido black assay [26]. 30 µg of each protein extract were loaded on a SDS-PAGE and were entered into the first 5 mm of the separation gel [27]. The complete SDS-PAGE fraction was cut out of the gel, digested tryptically into peptides [28] and dried using a vacuum centrifuge (Digital Series SpeedVac SPD121P, Thermo Scientific, Waltham, USA). For the LC-MS/MS measurement, the samples were resolved in 12 µL of solvent A (98% LC-MS Water, 2% ACN, 0.05% TFA). After centrifugation (30 min, 13,000 xg, 4 °C), 10 µL of this peptide solution were transferred into a HPLC vial. The LC-MS/MS analysis was carried out using an UltiMate 3000 RSLCnano splitless liquid chromatography system, coupled online to an Orbitrap Elite™ Hybrid Ion Trap-Orbitrap MS (both from Thermo Fisher Scientific, Bremen, Germany). Initially, for concentrating and desalting 5 µL injected peptides were loaded isocratically on a trap column (Dionex Acclaim, nano trap column, 100 µm i.d. x 2 cm, PepMap100 C18, 5 µm, 100 Å, nanoViper) with a flow rate of 7 µL/min using 4% solvent A (solvent A: 98% LC-MS Water, 2% ACN, 0.05% TFA, solvent B: 80% ACN, 20% LC-

MS water, 0.05% TFA). Next, the chromatographic separation was operated on a Dionex Acclaim PepMap C18 RSLC nano reversed phase column (2 µm particle size, 100 Å pore size, 75 µm inner diameter and 250 mm length) at 40°C column temperature. The used flow rate of 300 nL/min was applied by a binary A/B-solvent gradient (nano solvent A: 98% LC-MS Water, 2% acetonitrile, 0.1% formic acid; nano solvent B: 80% acetonitrile, 10% LC-MS Water, 10% trifluoroethanol, 0.1% formic acid) beginning with 4% of solvent B for 4 min followed with a linear increase to 55% of solvent B within 120 min. Afterwards, the column was washed with 90% B for 5 min and re-adjusted with 4% of solvent B for 25 min. A data-dependent MS/MS method was chosen for the MS acquisition using the positive ion mode while precursor ions were acquired in the orbital trap of the hybrid MS at a resolution of 30,000 and an m/z range of 350–2000. Subsequently, the fragment ion scan proceeded in the linear ion trap of the hybrid MS with a mass range and a scan rate with default parameter settings for the top 20 most intense precursors selected for collision-induced dissociation.

2.3. Data handling

For protein identification the acquired LC-MS/MS spectra were submitted to a comprehensive bioinformatics workflow. The raw mass spectral data (*.raw-files) were converted to *.mgf-files using the Proteome Discoverer Software (Thermo Fisher Scientific, Bremen, Germany, version 1.4.1.14). For protein database search the mgf-files were uploaded into the MetaProteomeAnalyzer [29] (MPA) (version 2.12, www.mpa.ovgu.de) and searched with X!Tandem [30] and Omssa [31] against a protein database containing UniProtKB/SwissProt (version: 23.10.2014) and several metagenomes from other microbiome studies [32]. Additionally, protein database searches with Mascot (Matrix Science, London, England, version 2.5.1) were performed using the ProteinScape software (Bruker Daltonics, Bremen, Germany, version 3.1.3461) and the results were imported as *.dat files into the MPA.

The following parameters for the protein database search were applied: trypsin as used enzyme, one missed cleavage, monoisotopic mass, carbamidomethylation (cysteine) as fixed modification, oxidation (methionine) as variable modification, ± 10 ppm as peptide tolerance and ± 0.5 Da MS/MS fragment tolerance, ^{13}C , +2/+3 charged peptide ions and a false discovery rate of 1%.

Identified unknown protein sequences from the metagenomes were annotated by Basic Local Alignment Search Tool (BLAST) search (NCBI-Blast-version 2.2.31) [33] against UniProtKB/SwissProt requiring a maximum e-value of 10^{-4} . All BLAST hits with best e-value were considered for further processing. Afterwards, all identified proteins were linked with their associated UniProt entries [34] covering the biological processes (UniProtKB keywords), the NCBI taxonomies [35], the UniProt Reference Clusters (UniRef) [36], the KEGG orthologies [37] and the enzyme commission numbers [38]. In order to decrease the number of redundant, homologous protein identifications, these proteins were grouped to metaproteins according to the UniRef50 clusters.

Metaprotein profiles were exported as comma separated value files and cluster analysis was performed using Matlab (The MathWorks GmbH, Ismaningen, Germany, version 8.3.0.532 (R2014a)). Krona plots [39] were created to visualize taxonomic results. Finally, all result files were submitted for long term storage to PRIDE [40] with the accession number PXD010371.

2.4. Statistical analysis

Analysis of similarities (ANOSIM) and hierarchical clustering (Bray Curtis distance, paired group linkage) were performed with Past3 (Version: 3.17) using the spectral counts of all human and microbial metaproteins. For the one-way ANOSIM p -values smaller than 0.01

were considered statistically significant and R-values were defined as follows: 0.75–1 highly different; 0.50–0.75 different; 0.25–0.49 different with some overlap; 0–0.24 highly similar [41].

Significant differences (p -value < 0.01) in the spectral abundance of taxonomies and metaproteins between the healthy individuals and the five diseases were identified by two-tailed Mann-Whitney U Tests using Microsoft Excel® (Version 2016). Only, the metaproteins and microbial phyla representing $> 0.01\%$ of the identified spectra were considered for statistical analysis.

3. Results

All together this study investigated the microbial and human metaproteins in 71 fecal samples from patients with CD, UCa, UC, IBS, GCA and CA and healthy individuals as controls. Therefore, we searched for disease specific metaprotein patterns as well as certain taxonomic phyla and metaproteins which could be used as diagnostic markers for a specific disease.

3.1. Protein extraction and identification

From 200 mg of fecal samples approx. 0.12 mg to 4.06 mg protein amount were extracted. Protein database searches from several human fecal samples identified in average 11,966 spectra. All identified spectra were assigned to 2969 metaproteins and 23 microbial phyla (Supplementary Table 2). On average, and based on the spectral abundance, 74.7% of the identified metaproteins were assigned to the host and 25.3% to the microbial communities. Among the assignable microbial communities and summed over samples (Supplementary Fig. 1) the four phyla Proteobacteria, Firmicutes, Bacteroidetes and Actinobacteria constituted the biggest groups as observed also in other studies [42]. For the entire taxonomy profile of all samples please refer to Supplementary Table 2.

3.2. Identification of disease specific metaprotein patterns

To identify disease specific metaprotein patterns a hierarchical clustering (Fig. 1A) using Past3 and the Bray-Curtis similarity index were performed.

Consistency of the workflow was confirmed after finding that the highest similarities were between the technical replicates (P6; P17; P18). Furthermore, four main clusters were detected with $< 30\%$ similarity within the metaprotein profiles. The cluster 1 comprised only patients with GCA (4 out of 4, 100% of patients). Cluster 2 was associated with patients with different diseases (17 out of 18, 94% of patients) including IBD, GCA, CA and in particular IBS, whereas cluster 3 mainly comprised healthy individuals and patients with CA (21 out of 25, 84% of patients). Cluster 4 was linked with CD and UC (19 out of 24, 79% of patients). The protein profiles showed a separation between samples of IBD and GCA versus samples from IBS, CA and healthy individuals. However, no specific clustering was observed for UCa and UC. Furthermore, metaprotein profiles from IBS samples and CA appeared to be more similar to the healthy individuals than to IBD.

After assigning the calprotectin levels measured by ELISA (D) to the different diseases increased amounts for CD, UC and GCA were observed in comparison to healthy controls as well as IBS and CA patients. Similar results for most samples were also observed by metaproteome analysis focusing on the calprotectin subunits S100A8 (UniRef50_P05109) and protein S100A9 (UniRef50_P06702) (Supplementary Note 1). However, for certain samples such as P03 and P52 metaproteome analyses revealed lower amounts of calprotectin, matching better to the specific clinical picture.

To apply metaproteomics for the diagnosis of individual diseases a clear differentiation between the respective metaprotein profiles is re-

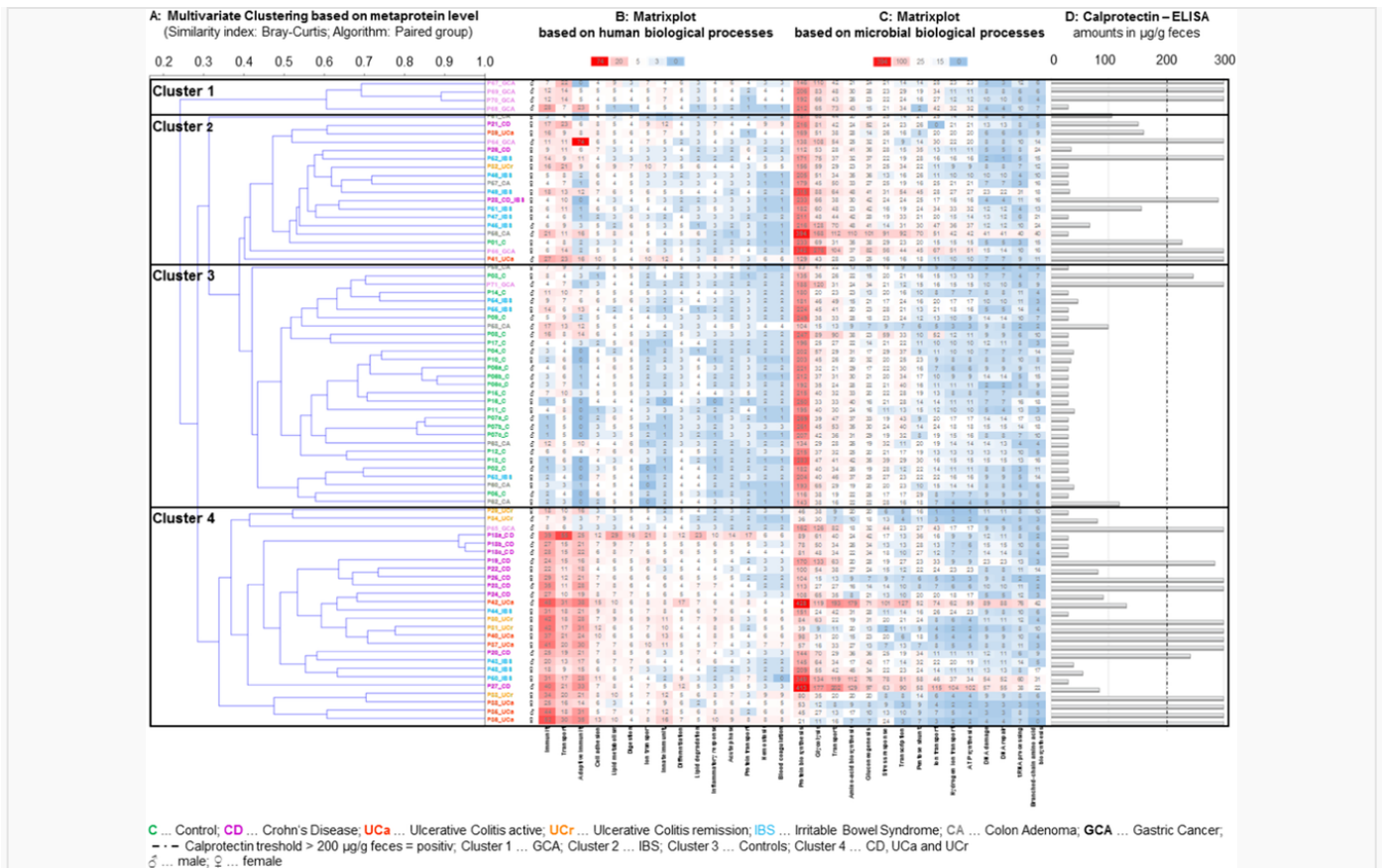


Fig. 1. Identification of disease specific metaprotein patterns by hierarchical clustering. This figure shows the cluster analysis of all samples (A) by hierarchical clustering using Past 3. For the clustering all human and microbial metaproteins were considered (Supplementary Table 2). Names of samples are shown in the middle of the figure. Applied colors are explained in the legend. After visual assignment, four main clusters were proposed: namely GCA (Cluster 1), mainly IBS and other diseases (Cluster 2), mainly healthy individuals (Cluster 3) as well as mainly Crohn's Disease (CD) and Ulcerative Colitis (UC) (Cluster 4). Subsequently, the metaproteins were grouped according to their human (B) and microbial (C) biological processes (UniProtKB keyword) and for both groups the 15 most abundant processes were shown. Red indicates a high, white an average and blue a low spectral abundance. In D the calprotectin amounts, quantified by ELISA are shown. Amounts of $>200 \mu\text{g/g}$ feces were considered as active inflammation and highlighted by a blue threshold line. CD: Crohn's Disease (CD); UC: Ulcerative Colitis; IBS: Irritable Bowel Syndrome (IBS); CA: Colon Adenoma; GCA: Gastric Carcinoma. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Identification of disease specific metaprotein patterns by hierarchical clustering. This figure shows the cluster analysis of all samples (A) by hierarchical clustering using Past 3. For the clustering all human and microbial metaproteins were considered (Supplementary Table 2). Names of samples are shown in the middle of the figure. Applied colors are explained in the legend. After visual assignment, four main clusters were proposed: namely GCA (Cluster 1), mainly IBS and other diseases (Cluster 2), mainly healthy individuals (Cluster 3) as well as mainly Crohn's Disease (CD) and Ulcerative Colitis (UC) (Cluster 4). Subsequently, the metaproteins were grouped according to their human (B) and microbial (C) biological processes (UniProtKB keyword) and for both groups the 15 most abundant processes were shown. Red indicates a high, white an average and blue a low spectral abundance. In D the calprotectin amounts, quantified by ELISA are shown. Amounts of $>200 \mu\text{g/g}$ feces were considered as active inflammation and highlighted by a blue threshold line. CD: Crohn's Disease (CD); UC: Ulcerative Colitis; IBS: Irritable Bowel Syndrome (IBS); CA: Colon Adenoma; GCA: Gastric Carcinoma. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

quired. Deeper statistical analysis was performed by ANOSIM (Fig. 2). As indicated by the low p -values and high R -values samples of the diseases differed significantly from the healthy individuals as already shown by the cluster analysis. The highest differences to the healthy individuals were observed for UCr and UCa patients ($p = 0.0002$, $R = 0.9211$), followed by patients with CD ($p = 0.0001$, $R = 0.7761$), GCA ($p = 0.0001$, $R = 0.7716$), IBS ($p = 0.0001$, $R = 0.4722$), and finally by patients with CA ($p = 0.0005$, $R = 0.4679$).

Furthermore, samples from IBS patients were statistically different from that samples of patients with UCa ($p = 0.0001$, $R = 0.6159$) as well as UCr ($p = 0.0005$, $R = 0.5125$) and partially from patients with CD ($p = 0.0017$, $R = 0.3282$). On contrary, no statistically differences were found between differentiation between CD and UCr ($p = 0.3662$, $R = 0.0223$), CD and UCa ($p = 0.1217$, $R = 0.0967$) and between UCr and UCa ($p = 0.5165$, $R = -0.0184$).

3.3. Identification of disease specific human and microbial functional patterns

Metaproteins were separated into human and microbial metaproteins and then grouped according to their biological processes (UniProtKB keyword) (Supplementary Table 2 "Human biological processes", "Microbial biological processes"). The 15 most abundant microbial and human biological processes were visualized separately as a matrix plot (Fig. 1B and C). Clustering of the biological processes did not result in a better separation of the diseases (data not shown). Nevertheless, the matrix plot enabled a first functional overview about the functional differences, which were linked with the diseases.

Identified human metaproteins were mainly assigned to the immune system and to transport processes. Samples of IBD patients (UC + CD) contained larger amounts of human metaproteins than samples from all other diseases or healthy individuals. In particular, metaproteins for immunity, adaptive immunity, innate immunity and

p-values, uncorrected significance							
	Control	CD	UCr	UCa	IBS	CA	GCA
Control		0.0001	0.0002	0.0001	0.0001	0.0005	0.0001
CD	0.0001		0.3662	0.1217	0.0017	0.0005	0.0008
UCr	0.0002	0.3662		0.5165	0.0005	0.0011	0.0133
UCa	0.0001	0.1217	0.5165		0.0001	0.0003	0.0005
IBS	0.0001	0.0017	0.0005	0.0001		0.0065	0.0001
CA	0.0005	0.0005	0.0011	0.0003	0.0065		0.0102
GCA	0.0001	0.0008	0.0133	0.0005	0.0001	0.0001	
R-values							
	Control	CD	UCr	UCa	IBS	CA	GCA
Control		0.7761	0.9211	0.942	0.4722	0.4679	0.7716
CD	0.7761		0.0223	0.0967	0.3282	0.484	0.4439
UCr	0.9211	0.0223		-0.0184	0.5125	0.5533	0.3866
UCa	0.942	0.0967	-0.0184		0.6159	0.6669	0.5469
IBS	0.4722	0.3282	0.5125	0.6159		0.2828	0.5499
CA	0.4679	0.484	0.5533	0.6669	0.2828		0.3583
GCA	0.7716	0.4439	0.3866	0.5469	0.5499	0.3583	

Fig. 2. ANOSIM of different diseases. P- and R-values for the one-way ANOSIM based on the metaprotein profiles of the healthy individuals as well as the different diseases (CD: Crohn's Disease; UCr: Ulcerative Colitis in remission stage; UCa: Ulcerative Colitis in active stage; IBS: Irritable Bowel Syndrome; CA: Colon Adenoma; GCA: Gastric Carcinoma). P-values < 0.01 were considered statistical significant and are highlighted with a grey background. R-values are coded as follows: > 0.75; 0.50–0.75; 0.25–0.49; < 0–0.24.

acute phase were more abundant in UC and CD patients. In contrast to samples from CD patients, samples from UC patients contained more metaproteins for hemostasis and blood coagulation. The main difference between samples from the healthy individuals and patients with IBS, CA and GCA was a slight increase of the immunity and transport processes.

Identified microbial metaproteins from all samples were mainly assigned to metabolic pathways such as protein biosynthesis, glycolysis and transport as also found in similar studies [42,43]. In samples from UC patients a decreased amount of metaproteins for ion or hydrogen ion transport as well as ATP synthesis was found.

3.4. Identification of disease specific biomarkers

In order to identify disease specific metaproteins as well as microbial phyla Mann-Whitney *U* tests were carried out for all diseases based on spectral abundances. Over all 5 phyla, 45 microbial metaproteins and 58 human metaproteins showed significant alterations in their abundance in the disease groups in comparison to the healthy individuals. In the following the most important alteration are present. For more details please refer to supplementary Table 3A3A, B, C.

3.4.1. Microbial taxonomies and metaproteins

CD patients showed a significant decrease of the class Acidobacteria (Supplementary Table3A, Fig. 3). Patients with UC displayed reduced abundances of the phyla Bacteroidetes, Spirochaetes, and Tenericutes. IBS was also linked with a decrease of Tenericutes whereas GCA patient possessed less microorganisms belonging to the phyla Actinobacteria.

Nearly all microbial metaproteins showing a significant alteration in their abundance were associated with the microbial metabolism or were housekeeping proteins (Supplementary Table 3B).

As already observed by ANOSIM (Fig. 2) patients with CD, UCa and UCr showed quite similar metaprotein expression profiles. Overall, samples of these patients contained lower amounts of microbial metaproteins such as elongation factor G (Bacteria, UniRef50_P08068, CD/C: $p = 0.001$, UCr/C: $p = 0.001$, UCa/C: $p = 0.001$), 30S ribosomal protein S8 (Bacteria, UniRef50_P14826, CD/C: $p = 0.003$, UCr/C: $p = 0.005$, UCa/C: $p = 0.001$), and DNA-directed RNA polymerase (Bacteroidetes, UniRef50_Q8A469, CD/C: $p = 0.001$, UCr/C: $p = 0.007$, UCa/C: $p = 0.001$). Of note is the transcriptional regulatory metaprotein RprY from *Bacteroides fragilis* (UniRef50_Q9AE24, CD/C: $p = 0.001$, UCr/C: $p = 0.003$, UCa/C: $p = 0.002$), which was less abundant for CD and UC patients (Fig. 4). Only significantly decreased

in CD patients were the proteins fructose-biphosphate aldolase (UniRef50_P14540, $p = 0.003$) from the order Saccharomycetales, two elongation factors (Bacteria, UniRef50_Q9P9Q9, $p = .005$; Bacteroidales, UniRef50_Q8A0Z3, $p = 0.009$) and an uncharacterized ABC transporter ATP-binding protein (Bacteria, UniRef50_P10907, $p = 0.005$). Specific for UCa were lower amounts of xylose isomerase (Bacteroides, UniRef50_Q2SW40, $p = 0.006$).

In contrast to UC and CD samples, samples from IBS and CA patients showed only minor alteration in comparison to the healthy individuals. For example in CA patients less phosphoenol pyruvate carboxykinase (Bacteria, UniRef50_O09460, $p = 0.000$) and 30S ribosomal protein S2 (Thermoanaerobacter sp. X514, UniRef50_Q8RA21, $p = 0.010$) were observed in comparison to the healthy individuals. At least specific for IBS samples was an increased amount of pyruvate-flavodoxin oxidoreductase (Enterobacteriaceae, UniRef50_P03833, $p = 0.001$).

Among others for GCA, increased amounts of alkyl hydroperoxide reductase (Proteobacteria, UniRef50_P80239, $p = 0.005$) (Fig. 4), malate dehydrogenase (Gammaproteobacteria, UniRef50_P44427, $p = 0.003$), L-asparaginase (Proteobacteria, UniRef50_P00805, $p = 0.006$) (Fig. 4) as well as outer membrane proteins (Proteobacteria, UniRef50_Q8ZG77, $p = 0.005$) were observed.

3.4.2. Human metaproteins

Most significant alteration in human metaproteins between patient cohorts compared to the healthy individuals could be linked to immunoglobulins (Igs), antimicrobial proteins, cell integrity proteins and proteins involved in inflammatory processes as well as proteins involved in digestion (Supplementary Table3C).

In comparison to healthy individuals samples from patients with CD and UC (IBD) showed both the upregulation of some metaproteins belonging to the main components of neutrophil extracellular traps (NETs) The detected NET metaproteins were neutrophil elastase (UniRef50_P08246, CD/C: $p \leq 0.001$, UCr/C: $p = 0.003$, UCa/C: $p = 0.001$) (Fig. 4), azurocidin (UniRef50_P20160, CD/C: $p = 0.001$, UCr/C: $p = 0.002$, UCa/C: $p = 0.001$), cathepsin G (UniRef50_P08311, CD/C: $p = 0.009$, UCa/C: $p = 0.001$), myeloblastin (UniRef50_P24158, CD/C: $p = 0.003$, UCa/C: $p = .001$), myeloperoxidase (UniRef50_P05164, CD/C: $p = 0.001$, UCr/C: $p = 0.004$, UCa/C: $p = 0.001$), neutrophil gelatinase-associated lipocalin (UniRef50_P80188, CD/C: $p = 0.001$, UCr/C: $p = 0.001$, UCa/C: $p = 0.001$), protein-arginine deiminase type-2 (UniRef50_Q9Y2J8, UCa/C: $p = 0.002$) as well as metaproteins S100-A8/A9 (UniRef50_P05109, CD/C: $p = 0.001$, UCr/C: $p = 0.004$, UCa/C: $p = 0.001$); UniRef50_P06702, CD/C: $p = 0.002$,

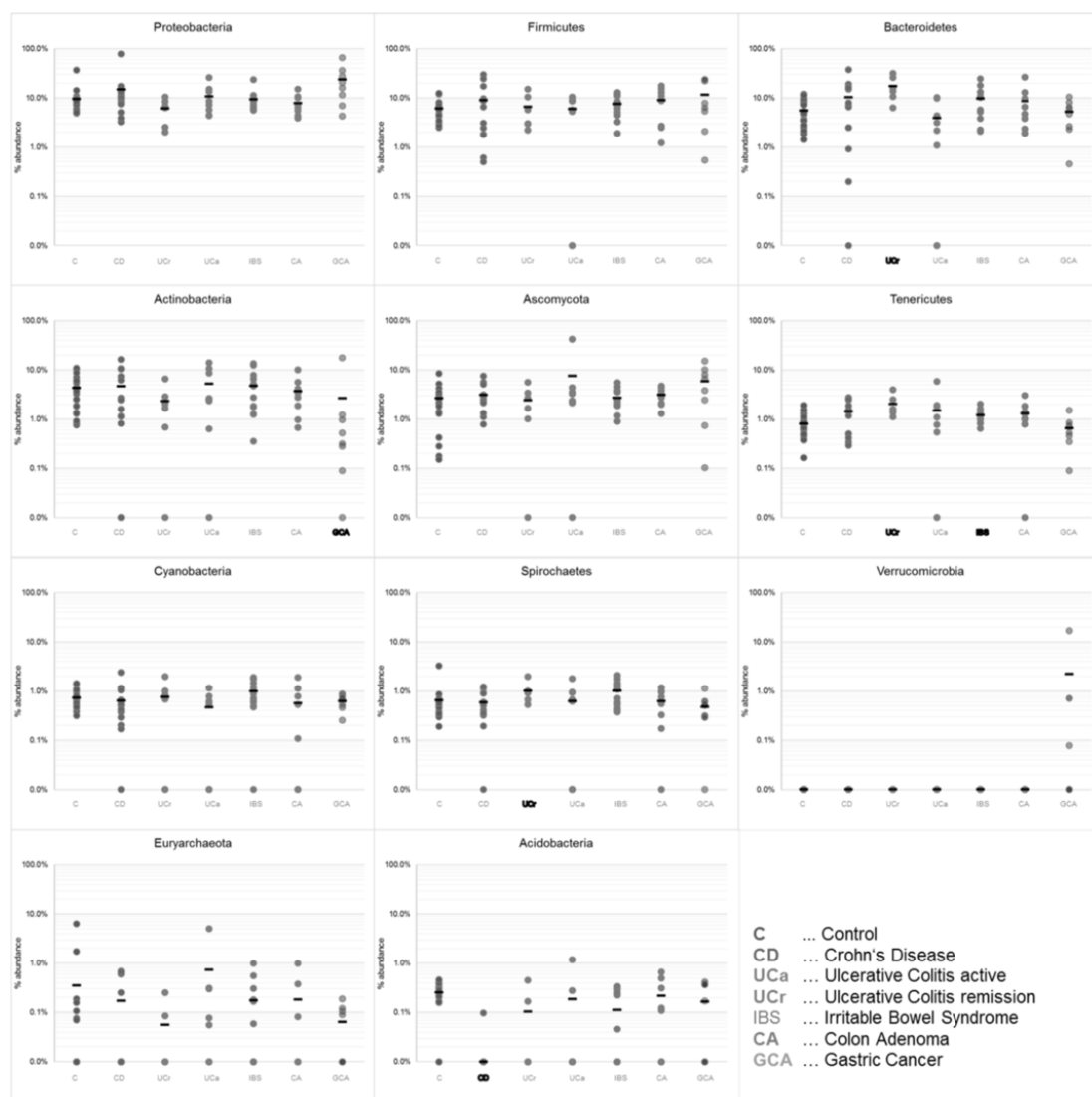


Fig. 3. Main microbial phyla. Relative abundance (log10-transformed) of the dominant phyla in fecal samples from healthy individuals and patients (CD: Crohn's Disease; UCr: Ulcerative Colitis in remission stage; UCa: Ulcerative colitis in active stage; IBS: Irritable Bowel Syndrome; CA: Colon Adenoma; GCA: Gastric Carcinoma). The means are illustrated as black bars.

UCr/C: $p = 0.007$, UCa/C: $p = 0.001$) (Fig. 4). Furthermore, in the samples of CD and UC patients an increased amount of IgG (UniRef50_P01859) and a decreased amount of IgA (UniRef50_P01876, CD/C: $p = 0.001$, UCr/C: $p = 0.001$, UCa vs. C: $p = 0.001$) (Fig. 4) was observed in comparison to the healthy controls.

Interestingly, significantly more abundance of titin (UniRef50_Q8WZ42, $p = 0.002$), intestinal sucrase-isomaltase (UniRef50_P14410, $p = 0.002$) (Fig. 4), and cadherin-17 (UniRef50_Q9R100, $p = 0.009$) could be detected exclusively in CD patients whereas more abundance of annexin A3 (UniRef50_P12429, $p = 0.001$), complement C4-A (UniRef50_P0C0L4, $p = 0.002$) (Fig. 4), resistin (UniRef50_Q9HD89, $p = 0.001$), and plasma protease C1 inhibitor (UniRef50_P05155, $p = 0.001$) as well as less abundance of chymotrypsin C (UniRef50_Q99895, $p = 0.008$) were only detected in UC patients.

Comparison between UCa and UCr showed similar regulated metaproteins, but less strong in UCr.

An increase of IgG and a decrease of IgA were observed in IBS patients in comparison to the healthy individuals, similar to CD and UC patients, but to a lesser degree. However, larger amounts of pancreatic secretory granule membrane major glycoprotein GP2 (UniRef50_P55259, $p = 0.001$), and enteropeptidase (UniRef50_P98073, $p = 0.001$) were observed exclusively in IBS samples.

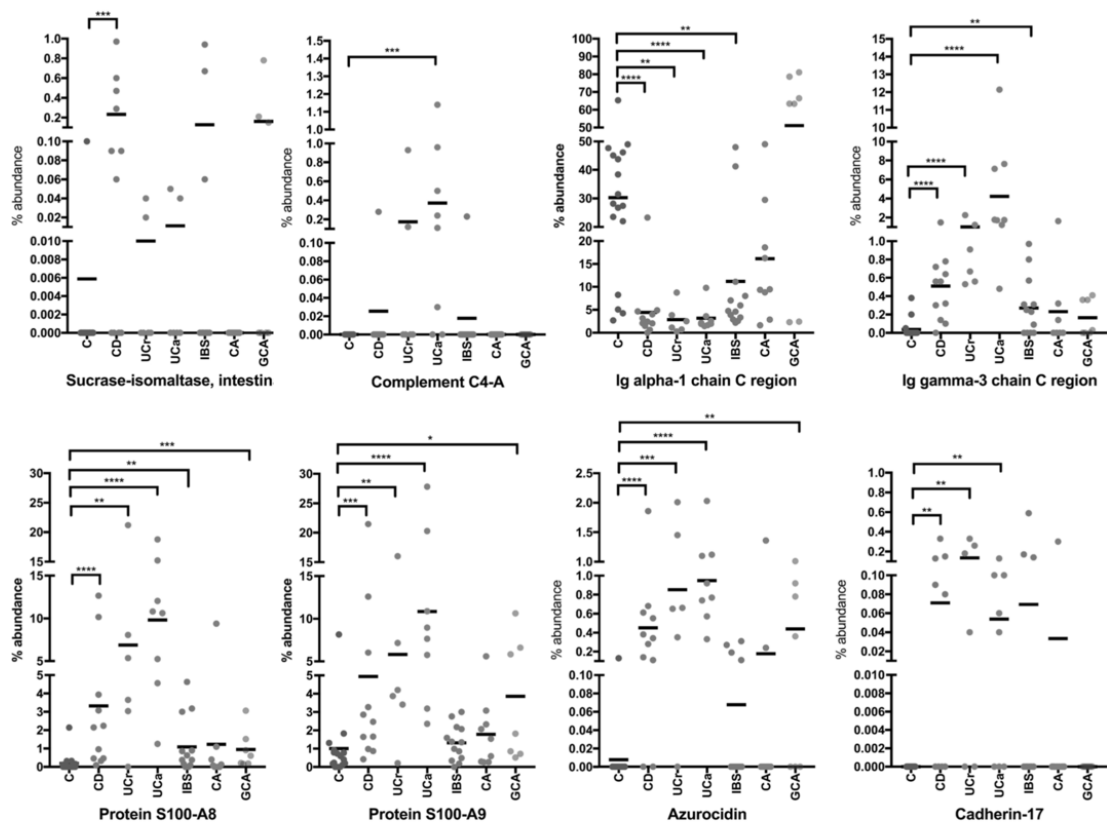
Additionally, an increase in trypsin-2 (UniRef50_P07478, $p = 0.001$) was only observed in samples of patients with GCA.

4. Discussion

The main goal of this study was to proof the concept whether metaproteomics may distinguish between patients with different diseases and healthy individuals to support the diagnosis of GIT diseases. Therefore, the microbial and human metaproteins in fecal samples of patients with CD, UC, IBS, CA, GCA and a control group of healthy individuals were examined. In total 2969 metaproteins were identified, revealing interdependencies between diseases and metaprotein profiles. Additionally, identified metaproteins helped to reconstruct the taxonomic composition of the microbiome and its metabolic reactions.

Highly similar technical replicates confirm the reproducibility of the metaproteome workflow (Fig. 1). Human metaproteins were found highly abundant in all measurements (average $\sim 75\%$). Proteomics only identifies the proteins with the highest abundance. Microbial proteins are spread among multiple species and are consequently less abundant

A: Potential human marker metaproteins



B: Potential bacterial marker metaproteins

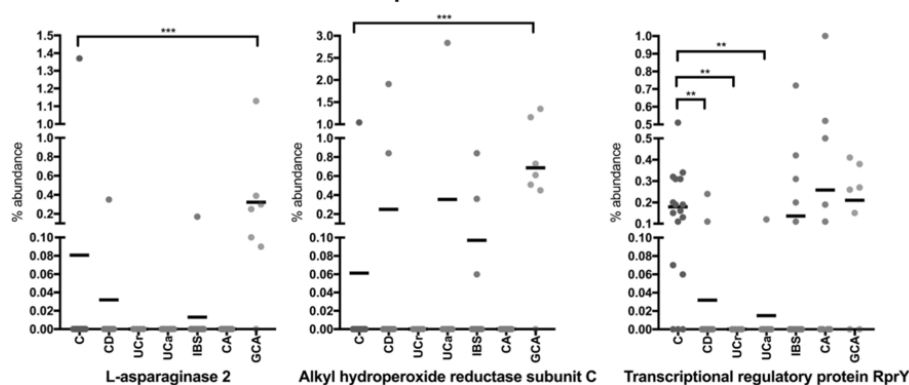


Fig. 4. Panel of potential marker metaproteins. Relative abundance (log10-transformed) of the marker metaproteins in fecal samples from healthy individuals and patients (CD: Crohn's Disease; UC: Ulcerative Colitis in remission stage; UC*: Ulcerative colitis in active stage; IBS: Irritable Bowel Syndrome; CA: Colon Adenoma; GCA: Gastric Carcinoma). The means are illustrated as black bars. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

than the few human proteins found in human feces. To increase the amount of identified microbial metaproteins, several solutions exist including sample fractionation [27,44], more sensitive mass spectrometry equipment [45] or the depletion of highly abundant human proteins [46].

Diagnosis of diseases of the GIT is difficult and requires experienced physicians and in most cases invasive diagnostic tools such as endoscopy. Previous non-invasive methods such as the detection of calprotectin in feces by ELISA gave some hints about mucosal inflammation of the large bowel but could be improved since the sensitivity for diagnosis of IBD was only 89% and the specific only 79% [47] (Supplementary Note 1, Fig. 1D). In contrast to the single-protein-analysis of calprotectin by ELISA, metaproteome analysis revealed abundances of

many metaproteins providing comprehensive information about the patient and the microbiome. This will become especially useful as soon as the impact of the different microorganism on the IBD pathogenesis is better understood.

Hierarchical cluster analysis was used to separate patient cohorts based on their metaprotein profiles. Overall, four main clusters were observed in the clustergram, which were linked with GCA (cluster 1), different diseases of the GIT (cluster 2), healthy individuals and IBS (cluster 3) as well as UC and CD (cluster 4). Apparently, the investigated diseases had a bigger impact on the fecal metaproteome in the clustergram (Fig. 1) than inter and intra patient variations such as sample time points or the gender. However, not all five analyzed diseases could be clearly separated and some patient's metaproteome profiles

fell into the wrong cluster. Still, our results may already provide valuable information for the diagnosis of the diseases. Furthermore, one-way ANOSIM confirmed the separation of the diseased groups from the healthy individuals (Fig. 2). However, the results should be verified by a further study on an independent cohort since the protein profiles of the different samples possessed high intra and interindividual variability as also observed in similar studies [42,43,48,49]. At least the correlation of the human enzyme calprotectin detected by the ELISA with the MS results (Supplementary Note 1) proved the validity of the study.

Based on the clustering of the metaprotein profile (proteotyping [25]), physicians may distinguish whether a new sample belongs to healthy individuals or to people with only a functional disorder (IBS) or to patients that potentially have a severe disease such as GCA, UC or CD.

A differentiation between CD, UC and UCa and between the healthy individuals and IBS or CA was not possible at this time. The similarity between the healthy individuals, CA and IBS samples can be explained by the small number of samples or by the fact that all three cases do not cause big functional changes of the GIT and microbiome. Furthermore, variations between IBS patients might be caused by IBS subtypes, which display a wide range of symptoms such as diarrhea and obstipation. In contrast, CD and UC cause severe inflammations which resulted even under remission in increased levels of inflammation markers. However, it was difficult to distinguish the diseases since the severity of the inflammation and the progress of the therapies varied between the patients.

In order to improve the diagnosis of the diseases specific marker metaproteins were searched by Mann-Whitney *U* test. Furthermore, the regulated metaproteins were linked with pathogenesis.

CD and UC represent main inflammatory disease of the GIT. As expected for the inflammation processes in both diseases a significant increase of IgG and NET metaproteins were observed. NET proteins are produced by neutrophil granulocytes as part of their immune response. According to He et al. (2016) [50] and Bennike et al. (2015) [51] an increased number of neutrophil granulocytes migrate in IBD patients from the blood into the GIT. There, they excrete their DNA and several proteins in order to kill and trap microorganisms. Consequently, inhibition of neutrophil granulocytes migration and NET represent a promising strategy in IBD therapy [52].

Comparison of NET metaprotein abundance based on the metaproteome analysis against the calprotectin abundances measured by ELISA showed several significant differences. Since all other NET proteins exclusively identified by metaproteome analysis show the same trend, the results of the calprotectin ELISA might be biased by unspecific binding of the ELISA antibodies.

A main consequence of CD and UC is that the cells in the affected tissue partially lose their function due to inflammation processes. The mucosal membranes of the GIT produced less IgA [52,53] as observed by our metaproteome analysis [53,54]. Usually, IgA protects the GIT from microorganisms and can inhibit the immune response [55]. However, as previously mentioned an increase in IgG was observed. IgG activates the innate immune response and triggers an even stronger inflammation. Therefore, the dysregulation of IgA and IgG constitutes a feedback loop and might explain the chronic nature of UC and CD. Further evidence for this hypothesis is a reduced antibody dysbiosis in patients with UC under remission.

However, it remained unclear whether the exaggerated immune reaction is a dysfunction of the immune system or targets specific microorganisms which cause the diseases. Still, some differences in taxonomic composition were observed. Metaprotein taxonomies revealed e.g. a decrease of the phyla Spirochaetes, in UC patients and a decrease for the phyla Acidobacteria in CD patients. A good example for the correlation of UC and CD with the microbiome is the decrease of transcriptional regulatory protein RprY from *Bacteroides fragilis* (UniRe-

f50_Q9AE24) in UC and CD patients. It is known that the proteins RprX and RprY of *Bacillus fragilis* inhibit the expression of a proteinase, which is toxic for humans. Due to the decrease of RprY, more toxins can be expressed. The toxin may cleave E-cadherin molecules in the gut, which induces subsequently the release of cytokines in intestinal epithelial cells [56,57]. Finally, the cytokines may promote inflammatory processes.

The regulated metaproteins discussed so far, correlated with both CD and UC equally. A specific marker for CD is sucrose-isomaltase (UniRef50_P14410), which was found in increased amounts in CD patients. Usually this enzyme is attached to the brush border of the small intestine and is responsible for the degradation of carbohydrates [58]. During inflammation processes, this sucrose-isomaltase is likely detached from the cells. The reduced sugar degradation activity can also explain the weight loss in CD patients. Sucrose-isomaltase is not increased in UC patients, since the small intestine is not affected in UC. Therefore, this marker metaprotein allows the clear differentiation between both diseases. An alternative explanation for the increase of sucrose-isomaltase is the increase of tumor necrosis factor (TNF)-alpha in CD, which was correlated to increase the expression of sucrose isomaltase [59].

Additionally, several other regulated metaproteins were observed in CD and UC patients, which were already described as potential marker proteins, but could not be linked to coherent explanation for the diseases. This included the metaproteins gamma-glutamyl hydrolase (UniRef50_Q92820), resistin 50 (UniRef50_Q9HD89) olfactomedin-4 [60] (UniRef50_Q6UX06) and trefoil factor 252 (UniRef50_Q03404) as well as the metaproteins of the complement system C4-A (UniRef50_P0C0L4), and plasma protease C1 inhibitor (UniRef50_P05155) [61]. At least for the activation of the complement system it might be speculated that a dysfunction of it may be correlated with the diseases.

A comparison of our results with two recent IBD studies [42,43] showed several accordances such as the increased amounts of calprotectin, but also some discrepancies. For example Mills et al. (2019) [43] observed increased amounts of s-IgA in contrast to decreased amounts in our study. Contrary, Mills et al. (2019) observed also a strong dynamics of s-IgA. Unfortunately, a proper comparison of the studies is hindered by the different sampling time points, different laboratory workflows, different databases for protein identifications and different protein grouping strategies.

An increase of IgG and a decrease of IgA were observed for IBS patients, a similar but less intense case to what was found for CD and UC patients. This also indicates an activation of the immune system, and might correlate with a dysbiosis of the GIT or the microbiome. Furthermore, the proteins pancreatic secretory granule membrane major glycoprotein GP2 (UniRef50_P55259) and enteropeptidase (UniRef50_P98073) were upregulated. Latter enzyme is produced by the intestinal mucosa [62]. The increase of intestinal mucosa metaproteins can be explained by overworking resp. detachment of the intestinal mucosa due to diarrhea or constipation, which are the typical symptoms of IBS.

This study found no meaningful alterations of the metaprotein profile in the fecal samples of CA patients in comparison to the healthy individuals. In CA patients the intestinal epithelial is locally restricted proliferated, but the GIT and microbiome are still intact and working properly. Contrary to other publications, which found a relation between CA and the microbiome associated to the intestinal epithelial, this study could not reproduce these results [63]. A more in-depth investigation, including more samples and using a higher resolution is necessary to resolve the disagreement between these studies. Furthermore, this study should also include fecal samples from patients with colon carcinomas for which several potential marker proteins were already identified by other groups [48,49].

In samples of GCA patients an increased amount of enzymes from granulocytes was observed, which might be linked to tumor specific background inflammation. Furthermore, a massive decrease of the pancreas enzymes trypsin-2 (UniRef50_P07478) and kallikrein-1 (UniRef50_P07288) was observed. The decreased amount of trypsin might be an artefact of the tryptic digest during sample preparation, but this is unlikely since kallikrein-1 was similarly affected. The decrease of these enzymes indicates reduced pancreas functionality. The cause for this remains speculative, and maybe associated to the proliferation of the cancer [64], the cancer treatment, the medical history e.g. alcohol consume or other reasons.

Within the microbial community of GCA patients an upregulation of the enzymes alkyl hydroperoxide reductase (Proteobacteria, UniRef50_P80239) and L-asparaginase 2 (Proteobacteria, UniRef50_P00805) was observed. Unfortunately these metaproteins could be only assigned to the phyla Proteobacter and not to a more specific taxonomy level. Among others *Helicobacter pylori* is one important member of the phyla Proteobacter. Furthermore, *Helicobacter pylori* infections are a major cause of GCA, which may promote the tumor development through DNA-damaging reactive oxygen species [65]. Therefore, the presence of alkyl hydroperoxide reductase, an enzyme which protects the microorganisms against radical oxygen species, may be a potential marker metaprotein. L-asparaginase converts asparagine and glutamine to aspartate and glutamate, respectively. Previous studies [66,67] showed that this enzyme enables *Helicobacter pylori* to utilize extracellular asparagine and may promote the colonization of the gastric niche by the inhibition of normal lymphocyte function.

5. Conclusion

The study presented in this paper benefitted from the metaproteomics approach, by enabling the combined identification of multiple human and microbial metaproteins from a single fecal sample. Cluster analysis and specific marker metaproteins, i.e. human sucrose-isomaltase (UniRef50_P14410) and microbial RprY (UniRef50_Q9AE24), could be shown to differentiate between healthy individuals and patients of several GIT diseases.

Consequently, non-invasive metaproteome analysis of fecal samples may support the diagnosis of CD, UC, IBS, CA and GCA. Altered metaprotein expression provided new insights into the pathogenesis of GIT diseases, thereby, supporting the identification of drug targets. Before the presented method, the metaproteome analysis of fecal samples, can be applied in diagnostics, validation according to medical guidelines is required.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2019.04.009>.

Authors contribution

C.S., B.H.; Sample Collection & Clinical data recording: C.S., A.L., T.S.; Disease Diagnosis: C.S., A.L., L.B., B.H.; Medical support: A.L., L.B., S.S., A.C.; Conceptual design and project managing, T.S., R.H., R. V.-V.; Design & Performance of experiments, T.S., R.H.; MS-Analytics: S.P., R.H.; Computing Support, K.S.; Data Analysis, T.S., R.H., R. V.-V., D.B.; Writing of manuscript, T.S., R.H., R. V.-V., D.B., K.S.; Funding Acquisition, R.H.; Supervision, R.H.

Conflict of interest

All authors declare no conflicts of interest.

Patient consent

Obtained.

Ethical approval

The samples for this study were taken at the Otto-von-Guericke University Magdeburg and the Hannover Medical School while different projects (Magdeburg: Number 42/08, Number 47/15 and Hannover 2087-2013). Amendments to re-use the samples for this study were given by the local Ethics Committees.

Acknowledgements and funding

We thank the physicians for the diagnoses of the disease and for the medical support of this work as well as the patients for their permission to participate in this study. We acknowledge Corina Siewert for the laboratory support. This project was financed by the German Research Society (DFG) under number HE 8077/1-1.

References

- [1] F. Guarner, J.R. Malagelada, Gut flora in health and disease, *Lancet* 361 (9356) (2003) 512–519.
- [2] P.D. Cani, Human gut microbiome: hopes, threats and promises, *Gut* 67 (9) (2018) 1716–1725.
- [3] R. Sender, S. Fuchs, R. Milo, Revised estimates for the number of human and Bacteria cells in the body, *PLoS Biol.* 14 (8) (2016), e1002533.
- [4] N. Kaur, C.C. Chen, J. Luther, J.Y. Kao, Intestinal dysbiosis in inflammatory bowel disease, *Gut Microbes* 2 (4) (2011) 211–216.
- [5] C. Maaser, A. Sturm, S.R. Vavricka, T. Kucharzik, G. Fiorino, V. Annese, E. Calabrese, D.C. Baumgart, D. Bettenworth, P. Borralho Nunes, J. Burisch, F. Castiglione, R. Eliakim, P. Ellul, Y. González-Lama, H. Gordon, S. Halligan, K. Katsanos, U. Kopylov, P.G. Kotze, E. Krustinš, A. Laghi, J.K. Limdi, F. Rieder, J. Rimola, S.A. Taylor, D. Tolan, P. van Rheenen, B. Verstockt, J. Stoker, C.S. European, O. Colitis, the European Society of, G.; Abdominal, R, ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications, *J. Crohn's Colitis* (2018) jyy113.
- [6] A. Sturm, C. Maaser, E. Calabrese, V. Annese, G. Fiorino, T. Kucharzik, S.R. Vavricka, B. Verstockt, P. van Rheenen, D. Tolan, S.A. Taylor, J. Rimola, F. Rieder, J.K. Limdi, A. Laghi, E. Krustinš, P.G. Kotze, U. Kopylov, K. Katsanos, S. Halligan, H. Gordon, Y. González Lama, P. Ellul, R. Eliakim, F. Castiglione, J. Burisch, P. Borralho Nunes, D. Bettenworth, D.C. Baumgart, J. Stoker, C.S. European, O. Colitis, the European Society of, G.; Abdominal, R, ECCO-ESGAR guideline for diagnostic assessment in IBD part 2: IBD scores and general principles and technical aspects, *J. Crohn's Colitis* (2018) jyy114.
- [7] S.C. Ng, H.Y. Shi, N. Hamidi, F.E. Underwood, W. Tang, E.I. Benchimol, R. Panaccione, S. Ghosh, J.C.Y. Wu, F.K.L. Chan, J.J.Y. Sung, G.G. Kaplan, Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies, *Lancet* 390 (10114) (2018) 2769–2778.
- [8] C. Mowat, A. Cole, A. Windsor, T. Ahmad, I. Arnott, R. Driscoll, S. Mitton, T. Orchard, M. Rutter, L. Younge, C. Lees, G.T. Ho, J. Satsangi, S. Bloom, Guidelines for the management of inflammatory bowel disease in adults, *Gut* 60 (5) (2011) 571–607.
- [9] R. Francescone, V. Hou, S.I. Grivennikov, Microbiome, inflammation, and cancer, *Cancer J.* 20 (3) (2014) 181–189.
- [10] F. Mearin, B.E. Lacy, L. Chang, W.D. Chey, A.J. Lembo, M. Simren, R. Spiller, Bowel disorders, *Gastroenterology* (2016).
- [11] D.N. Frank, A.L. St Amand, R.A. Feldman, E.C. Boedeker, N. Harpaz, N.R. Pace, Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases, *Proc. Natl. Acad. Sci. U. S. A.* 104 (34) (2007) 13780–13785.
- [12] H. Sokol, P. Seksik, L. Rigottier-Gois, C. Lay, P. Lepage, I. Podglajen, P. Marteau, J. Dore, Specificities of the fecal microbiota in inflammatory bowel disease, *Inflamm. Bowel Dis.* 12 (2) (2006) 106–111.
- [13] D.A. Peterson, D.N. Frank, N.R. Pace, J.I. Gordon, Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases, *Cell Host Microbe* 3 (6) (2008) 417–427.
- [14] P. Correa, M.B. Piazuelo, *Helicobacter pylori* infection and gastric adenocarcinoma, *US Gastroenterol. Hepatol. Rev.* 7 (1) (2011) 59–64.
- [15] R. Caccaro, R. D'Inca, S. Pathak, G.C. Sturmiolo, Clinical utility of calprotectin and lactoferrin in patients with inflammatory bowel disease: is there something new from the literature?, *Expert. Rev. Clin. Immunol.* 8 (6) (2012) 579–585.
- [16] Human Microbiome Project, C, Structure, function and diversity of the healthy human microbiome, *Nature* 486 (7402) (2012) 207–214.
- [17] C.A. Kolmeder, W.M. de Vos, Metaproteomics of our microbiome - developing insight in function and activity in man and model systems, *J. Proteome* 97 (2014) 3–16.
- [18] A.R. Erickson, B.L. Cantarel, R. Lamendella, Y. Darzi, E.F. Mongodin, C. Pan, M. Shah, J. Halfvarson, C. Tysk, B. Henrissat, J. Raes, N.C. Verberkmoes, C.M. Fraser, R.L. Hettich, J.K. Jansson, Integrated metagenomics/metaproteomics reveals hu-

- man host-microbiota signatures of Crohn's disease, *PLoS One* 7 (11) (2012), e49138.
- [19] A. Heintz-Buschart, P. May, C.C. Laczny, L.A. Lebrun, C. Bellora, A. Krishna, L. Wampach, J.G. Schneider, A. Hogan, C. de Beaufort, P. Wilmes, Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes, *Nat. Microbiol.* 2 (2016) 16180.
- [20] C. Manichanh, L. Rigottier-Gois, E. Bonnaud, K. Gloux, E. Pelletier, L. Frangeul, R. Nalin, C. Jarrin, P. Chardon, P. Marteau, J. Roca, J. Dore, Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach, *Gut* 55 (2) (2006) 205–211.
- [21] A.D. Kostic, R.J. Xavier, D. Gevers, The microbiome in inflammatory bowel disease: current status and the future ahead, *Gastroenterology* 146 (6) (2014) 1489–1499.
- [22] C. Casen, H.C. Vebo, M. Sekelja, F.T. Hegge, M.K. Karlsson, E. Ciemniejewska, S. Dzankovic, C. Froyland, R. Nestestog, L. Engstrand, P. Munkholm, O.H. Nielsen, G. Rogler, M. Simren, L. Ohman, M.H. Vatn, K. Rudi, Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD, *Aliment. Pharmacol. Ther.* 42 (1) (2015) 71–83.
- [23] C. Schulz, T. Wex, U.V. Arnim, P. Malfertheiner, Validation of two Calprotectin rapid tests in daily routine, *Clin. Lab.* 62 (7) (2016) 1249–1254.
- [24] R. Heyer, F. Kohrs, D. Benndorf, E. Rapp, R. Kausmann, M. Heiermann, M. Klocke, U. Reichl, Metaproteome analysis of the microbial communities in agricultural biogas plants, *New Biotechnol.* 30 (6) (2013) 614–622.
- [25] R. Heyer, D. Benndorf, F. Kohrs, J. De Vrieze, N. Boon, M. Hoffmann, E. Rapp, A. Schluter, A. Sczyrba, U. Reichl, Proteotyping of biogas plant microbiomes separates biogas plants according to process temperature and reactor type, *Biotechnol. Biofuels* 9 (2016) 155.
- [26] N. Popov, M. Schmitt, S. Schulzeck, H. Matthies, Eine störungsfreie Mikromethode zur Bestimmung des Proteingehaltes in Gewebehomogenaten, *Acta biologica et medica Germanica* 34 (9) (1975) 1441–1446.
- [27] F. Kohrs, R. Heyer, A. Magnussen, D. Benndorf, T. Muth, A. Behne, E. Rapp, R. Kausmann, M. Heiermann, M. Klocke, U. Reichl, Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and thermophilic biogas plants, *Anaerobe* 29 (2014) 59–67.
- [28] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal. Chem.* 68 (5) (1996) 850–858.
- [29] T. Muth, A. Behne, R. Heyer, F. Kohrs, D. Benndorf, M. Hoffmann, M. Lehteva, U. Reichl, L. Martens, E. Rapp, The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation, *J. Proteome Res.* 14 (3) (2015) 1557–1565.
- [30] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, *Bioinformatics* 20 (9) (2004) 1466–1467.
- [31] L.Y. Geer, S.P. Markey, J.A. Kowalak, L. Wagner, M. Xu, D.M. Maynard, X. Yang, W. Shi, S.H. Bryant, Open mass spectrometry search algorithm, *J. Proteome Res.* 3 (5) (2004) 958–964.
- [32] J. Qin, R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D.R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H.B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, H.I.T.C. Meta, P. Bork, S.D. Ehrlich, J. Wang, A human gut microbial gene catalogue established by metagenomic sequencing, *Nature* 464 (7285) (2010) 59–65.
- [33] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (3) (1990) 403–410.
- [34] S. Patient, D. Wieser, M. Kleen, E. Kretschmann, M. Jesus Martin, R. Apweiler, UniProt/JAP: a remote API for accessing UniProt data, *Bioinformatics* 24 (10) (2008) 1321–1322.
- [35] S. Federhen, The NCBI taxonomy database, *Nucleic Acids Res.* 40 (Database issue) (2012) D136–D143.
- [36] B.E. Suzek, H.Z. Huang, P. McGarvey, R. Mazumder, C.H. Wu, UniRef: comprehensive and non-redundant UniProt reference clusters, *Bioinformatics* 23 (10) (2007) 1282–1288.
- [37] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* 44 (D1) (2016) D457–D462.
- [38] A. Bairoch, The ENZYME database in 2000, *Nucleic Acids Res.* 28 (1) (2000) 304–305.
- [39] B.D. Ondov, N.H. Bergman, A.M. Phillippy, Interactive metagenomic visualization in a web browser, *BMC Bioinformatics* 12 (2011) 385.
- [40] J.A. Vizcaino, A. Csordas, N. Del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its related tools, *Nucleic Acids Res.* 44 (22) (2016), 11033.
- [41] K.R. Clarke, Non-parametric multivariate analyses of changes in community structure, *Aust. J. Ecol.* 18 (1) (1993) 117–143.
- [42] J.A. Blakeley-Ruiz, A.R. Erickson, B.L. Cantarel, W.L. Xiong, R. Adams, J.K. Jansson, C.M. Fraser, R.L. Hettich, Metaproteomics reveals persistent and phylum-redundant metabolic functional stability in adult human gut microbiomes of Crohn's remission patients despite temporal variations in microbial taxa, genomes, and proteomes, *Microbiome* 7 (2019).
- [43] R.H. Mills, Y. Vazquez-Baeza, Q. Zhu, L. Jiang, J. Gaffney, G. Humphrey, L. Smarr, R. Knight, D.J. Gonzalez, Evaluating metagenomic prediction of the Metaproteome in a 4.5-year study of a Patient with Crohn's disease, *mSystems* 4 (1) (2019).
- [44] B.A. Petritz, O.L. Franco, Metaproteomics as a complementary approach to gut microbiota in health and disease, *Front. Chem.* 5 (2017).
- [45] F. Meier, S. Beck, N. Grassl, M. Lubeck, M.A. Park, O. Raether, M. Mann, Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device, *J. Proteome Res.* 14 (12) (2015) 5378–5387.
- [46] E. Bellei, S. Bergamini, E. Monari, L.I. Fantoni, A. Cuoghi, T. Ozben, A. Tomasi, High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins, *Amino Acids* 40 (1) (2011) 145–156.
- [47] N.E. Walsham, R.A. Sherwood, Fecal calprotectin in inflammatory bowel disease, *Clin. Exp. Gastroenterol.* 9 (2016) 21–29.
- [48] C.S. Ang, J. Rothacker, H. Patsiouras, P. Gibbs, A.W. Burgess, E.C. Nice, Use of multiple reaction monitoring for multiplex analysis of colorectal cancer-associated proteins in human feces, *Electrophoresis* 32 (15) (2011) 1926–1938.
- [49] L.J.W. Bosch, M. de Wit, T.V. Pham, V.M.H. Coupe, A.C. Hiemstra, S.R. Piersma, G. Oudgenoeg, G.L. Scheffer, S. Mongera, J.T.S. Droste, F.A. Oort, S.T. van Turenhout, I. Ben Larbi, J. Louwagie, W. van Criekeing, R.W.M. van der Hulst, C.J.J. Mulder, B. Carvalho, R.J.A. Fijneman, C.R. Jimenez, G.A. Meijer, Novel stool-based protein biomarkers for improved colorectal Cancer screening a case-control study, *Ann. Intern. Med.* 167 (12) (2017) 855.
- [50] Z. He, Y. Si, T. Jiang, R. Ma, Y. Zhang, M. Cao, T. Li, Z. Yao, L. Zhao, S. Fang, B. Yu, Z. Dong, H.S. Thatte, Y. Bi, J. Kou, S. Yang, D. Piao, L. Hao, J. Zhou, J. Shi, Phosphotyrosine exposure and neutrophil extracellular traps enhance procoagulant activity in patients with inflammatory bowel disease, *Thromb. Haemost.* 115 (4) (2016) 738–751.
- [51] T.B. Bennike, T.G. Carlsen, T. Ellingsen, O.K. Bonderup, H. Glerup, M. Bogsted, G. Christiansen, S. Birkelund, A. Stensballe, V. Andersen, Neutrophil extracellular traps in ulcerative Colitis: a proteome analysis of intestinal biopsies, *Inflamm. Bowel Dis.* 21 (9) (2015) 2052–2067.
- [52] A. Mantovani, M.A. Cassatella, C. Costantini, S. Jaillon, Neutrophils in the activation and regulation of innate and adaptive immunity, *Nat. Rev. Immunol.* 11 (8) (2011) 519–531.
- [53] A. Macpherson, U.Y. Khoo, I. Forgacs, J. Philpott-Howard, I. Bjarnason, Mucosal antibodies in inflammatory bowel disease are directed against intestinal bacteria, *Gut* 38 (3) (1996) 365–375.
- [54] L. Kazemly-Shirazi, C.H. Gasche, S. Natter, A. Gangl, J. Smolen, S. Spitzauer, P. Valent, D. Kraft, R. Valenta, IgA autoreactivity: a feature common to inflammatory bowel and connective tissue diseases, *Clin. Exp. Immunol.* 128 (1) (2002) 102–109.
- [55] N.J. Mantis, N. Rol, B. Corthesy, Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut, *Mucosal Immunol.* 4 (6) (2011) 603–611.
- [56] S. Wu, J. Shin, G. Zhang, M. Cohen, A. Franco, C.L. Sears, The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell receptor, *Infect. Immun.* 74 (9) (2006) 5382–5390.
- [57] S. Wu, J. Powell, N. Mathioudakis, S. Kane, E. Fernandez, C.L. Sears, Bacteroides fragilis enterotoxin induces intestinal epithelial cell secretion of interleukin-8 through mitogen-activated protein kinases and a tyrosine kinase-regulated nuclear factor-kappaB pathway, *Infect. Immun.* 72 (10) (2004) 5832–5839.
- [58] H.P. Hauri, A. Quaroni, K.J. Isselbacher, Biogenesis of intestinal plasma membrane: posttranslational route and cleavage of sucrase-isomaltase, *Proc. Natl. Acad. Sci. U. S. A.* 76 (10) (1979) 5183–5186.
- [59] J.M. Reimund, C. Wittersheim, S. Dumont, C.D. Muller, J.S. Kenney, R. Baumann, P. Poindron, B. Duclos, Increased production of tumour necrosis factor-alpha interleukin-1 beta, and interleukin-6 by morphologically normal intestinal biopsies from patients with Crohn's disease, *Gut* 39 (5) (1996) 684–689.
- [60] X.Y. Wang, S.H. Chen, Y.N. Zhang, C.F. Xu, Olfactomedin-4 in digestive diseases: a mini-review, *World J. Gastroenterol.* 24 (17) (2018) 1881–1887.
- [61] J. Laufer, R. Oren, I. Goldberg, A. Horwitz, J. Kopolovic, J. Chowers, J.H. Passwell, Cellular localization of complement C3 and C4 transcripts in intestinal specimens from patients with Crohn's disease, *Clin. Exp. Immunol.* 120 (1) (2000) 30–37.
- [62] J. Hermon-Taylor, J. Perrin, D.A. Grant, A. Appleyard, M. Babel, A.I. Magee, Immunofluorescent localisation of enterokinase in human small intestine, *Gut* 18 (4) (1977) 259–265.
- [63] Q. Feng, S. Liang, H. Jia, A. Stadlmayr, L. Tang, Z. Lan, D. Zhang, H. Xia, X. Xu, Z. Jie, L. Su, X. Li, X. Li, J. Li, L. Xiao, U. Huber-Schonauer, D. Niederseer, X. Xu, J.Y. Al-Aama, H. Yang, J. Wang, K. Kristiansen, M. Arumugam, H. Tilg, C. Datz, J. Wang, Gut microbiome development along the colorectal adenoma-carcinoma sequence, *Nat. Commun.* 6 (2015) 6528.
- [64] K. Yamashita, K. Mimori, H. Inoue, M. Mori, D. Sidransky, A tumor-suppressive role for trypsin in human cancer progression, *Cancer Res.* 63 (2003) 6575–6578.
- [65] K.M. Brawner, C.D. Morrow, P.D. Smith, Gastric microbiome and gastric cancer, *Cancer J.* 20 (3) (2014) 211–216.
- [66] D. Covini, S. Tardito, O. Bussolati, L.R. Chiarelli, M.V. Pasquetto, R. Digilio, G. Valentini, C. Scotti, Expanding targets for a metabolic therapy of cancer: L-asparaginase, *Recent Pat. Anticancer Drug Discov.* 7 (1) (2012) 4–13.
- [67] C. Scotti, P. Sommi, M.V. Pasquetto, D. Cappelletti, S. Stivala, P. Mignosi, M. Savio, L.R. Chiarelli, G. Valentini, V.M. Bolanos-Garcia, D.S. Merrell, S. Franchini, M.L. Verona, C. Bolis, E. Solcia, R. Manca, D. Franciotta, A. Casasco, P. Filipazzi, E.

Zardini, V. Vannini, Cell-cycle inhibition by helicobacter pylori L-asparaginase, PLoS One 5 (11) (2010), e13892.