

Metaproteomics reveals potential signatures of disease-specific alterations in the gut microbial proteolytic events in inflammatory bowel disease

Zhixiang Yan

Sun Yat-Sen University

Fei Xiao

Sun Yat-Sen University

Huanhuan He

Sun Yat-Sen University

Dan Li

Sun Yat-sen University

Li Cong

Sun Yat-Sen University

Feixiang He

Sun Yat-Sen University

Lu Lin

Sun Yat-Sen University

Yanyan Wu

Sun Yat-Sen University

Ru Yan

University of Macau

Xiaofeng Li

Sun Yat-Sen University

Hong Shan (✉ shanhong@mail.sysu.edu.cn)

Sun Yat-Sen University <https://orcid.org/0000-0003-1676-3320>

Research

Keywords: Metaproteomics, Gut microbial proteolysis, Inflammatory bowel disease

Posted Date: July 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38249/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background:

Proteolysis regulation allows gut microbes to respond rapidly to dynamic intestinal environments by fast degradation of misfolded proteins and activation of regulatory proteins. However, alterations of gut microbial proteolytic signatures under complex disease status such as inflammatory bowel disease (IBD, including Crohn's disease (CD) and ulcerative colitis (UC)) have not been investigated. Here, we explored the microbial proteolysis landscape using two public datasets containing 623 metaproteomes from 447 fecal and 176 mucosal luminal interface (MLI) samples from IBD patients and healthy individuals using a semi-tryptic peptide centric mining approach.

Results:

Fecal metaproteome revealed a global alteration of gut microbial proteolysis signatures in IBD, while MLI metaproteome demonstrated remarkable disease-specific and location-specific alterations at taxonomic, functional, and cleavage site motif levels. The functional alterations in IBD mainly involved microbial carbohydrate transport and metabolism, oxidative stress, cell motility, protein synthesis and maturation. Altered microbial proteolysis signatures of CD and UC mainly occurred in terminal ileum and descending colon, respectively. Microbial proteolysis patterns exhibited low correlations with β -diversity and moderate correlations with microbial protease transcriptome, respectively. Human protease inhibitors and immunoglobulins were mainly negatively associated with microbial proteolysis patterns, suggesting the inhibitory effects of these host factors on gut microbial proteolysis events.

Conclusions:

Our findings highlight the complex and diverse proteolytic alterations of gut microbiome in IBD, providing a unique layer of information beyond taxonomic and proteomic abundance.

Introduction

Gut microbiota lives in a dynamic environment, facing proteotoxic and metabolic stresses from drugs, diet, microbial competitors, and host endogenous chemical components. Regulated proteolysis allows microbes to respond to stress conditions rapidly by fast and specific proteolytic degradation of misfolded or damaged proteins, activation of regulatory proteins, and production of signals [1–3]. Several studies have investigated the regulation of certain bacterial proteolytic events in single bacterial species under simple environment stressors [3–5]. However, no study has been performed to explore alterations of gut microbial proteolytic signatures under complex disease status.

Inflammatory bowel disease (IBD), mainly consisted of Crohn's disease (CD) and ulcerative colitis (UC), is a chronic inflammatory disease influenced by genetic and environmental factors. Reports have confirmed that IBD is associated with a gut microbial dysbiosis. Metagenomics and 16S rRNA gene sequencing represented the vast majority in gut microbiome researches in IBD [6–8]. However, metatranscriptomics and/or metaproteomics approaches are needed to pinpoint functional and metabolic activities by direct measuring RNAs and proteins, respectively [9–11]. Furthermore, there are important additional regulations at protein level such as controlled proteolysis that are not captured in RNA measurements but could be monitored using metaproteomics.

In a routine (meta)proteomics data analysis [12–15], to reliably quantify proteins, it is necessary to select the most representative peptides and usually semi-tryptic peptides are not considered, because the expended search space will increase database search time and detection of semi-tryptic peptides was less consistent than that of fully tryptic peptides. In general, semi-tryptic peptides are mainly derived from endogenous proteolysis while the impacts of other factors such as in-source fragmentation and sample degradation are negligible [16, 17]. With these concepts in mind, we performed semi-tryptic peptide centric metaproteomics analysis to investigate the proteolytic events of gut microbiome. Utilizing two large-scale metaproteome datasets [7, 15], we demonstrated disease-specific and region-specific alterations of microbial proteolytic signatures in IBD.

Methods

Datasets

We analyzed two published datasets of healthy and IBD gut metaproteomes. Dataset 1 (PXD008675) was comprised of 447 fecal metaproteomes from 89 subjects aged 6-58 with a median of 22.8 years, including 24 non-IBD controls, 39 individuals with CD, and 26 with UC [7]. Of these, 272 and 184 samples have matched metagenomes and metatranscriptomes, respectively. Dataset 2 (PXD007819) came from 176 mucosal luminal interface (MLI) aspirate metaproteomes collected from the ascending colon (AsC), descending colon (DeC), or terminal ileum (TI) from 71 pediatric patients (<18 years old) including 25 CD, 22 UC, and 24 non-IBD controls [15]. Fecal samples were directly analyzed while the MLI aspirate samples were subjected to differential centrifugation to enrich microbial cells. MLI Aspirate samples were first centrifuged at 700 g, 4 °C for 5 min and the supernatant was further centrifuged at 14,000 g, 4 °C for 20 min with the pellet fraction harvested for metaproteomics analysis. Both datasets were generated by higher energy collisional dissociation and Orbitrap detection (HCD-OT) with metadata indicating age, gender, and disease status for each sample (Additional file 1: Tables S1 and S2).

Sequence database

A comprehensive human gut microbial protein database was generated by combining the following parts: (1) the integrated gene catalog (IGC) of human gut microbiome based on 1,267 gut metagenomes from 1,070 individuals (760 European, 368 Chinese, and 139 American samples) [18]; (2) the sequence data of 215 bacteria isolates cultured from healthy adult human faeces [19]; (3) the Culturable Genome

Reference (CGR) of 1,520 nonredundant, high-quality draft genomes generated from >6,000 bacteria cultivated from fecal samples of healthy humans [20]; (4) all Archaea, Bacteria, and Fungi sequences in UniProtKB (Release 2017_06) and NCBI RefSeq (Release 90). The microbial sequence database was appended by a UniProt human reference proteome (downloaded on 2017_06), a food database of dietary organisms *Triticum aestivum* (wheat), *Oryza sativa subsp. japonica* (rice), *Glycine max* (soybean), *Zea mays* (Maize), *Arachis hypogaea* (Peanut), *Solanum tuberosum* (Potato), *Solanum lycopersicum* (Tomato), *Sus scrofa* (pig), *Bos taurus* (Bovine), *Gallus gallus* (chicken), *Ovis aries* (sheep), *Salmo salar* and *Oncorhynchus mykiss* (fish), *Artemia sp.*, and *Litopenaeus vannamei* (shrimp), and a common contaminants database (<http://maxquant.org/contaminants.zip>). Proteins were dereplicated at 100% amino acid identity using USEARCH v11.0.667 (-fastx_uniques) [23], resulting in a total number of 130,975,891 non-redundant sequences.

Database searching

To handle the increased search space and time in metaproteomic semi-tryptic peptide identification, search was first performed using PEAKS DB (version X) [24] on a local 156-core cluster configured with Intel(R) Xeon(R) CPU @ 3.00GHz and 1.5TB 2666MHz RAM. The software first performed *de novo* sequencing followed by database search using the following parameters. Mass tolerance was set to 10 ppm for the precursor ion and 0.02 Da for the fragment ion. Carbamidomethylation of Cys was set as a fixed modification. The maximum number of variable posttranslational modifications per peptide was three, including acetylation of protein N-terminus, oxidation of Met, deamidation of Asn and Gln as well as Pyro-glu from Gln. For database search, enzyme was trypsin, digest mode was semispecific, and max missed cleavages were three. The two-step strategy [14] was employed to increase the sensitivity of metaproteomics database searching. Proteins identified by at least one peptides (1% false discovery rate (FDR) using the decoy fusion approach) in the first step search were reserved for the second round multiengine database search using PEAKS DB, MaxQuant (version 1.6.2) [25], and pFind (version 3.1.5) [26].

MaxQuant (version 1.6.2.10) was performed using the Andromeda search engine [27]. Mass tolerance was set to 20 ppm for the first and 4.5 ppm for the main search. Enzyme was trypsin, digest mode was semispecific, max missed cleavages were two. Carbamidomethylation of Cys was set as a fixed modification. The maximum number of variable posttranslational modifications per peptide was five, including acetylation of protein N-terminus, oxidation of Met, deamidation of Asn and Gln as well as Pyro-glu from Gln. Peptide-to-spectrum matches, peptide and site FDR were set to 0.01 based on the target-decoy strategy. Second peptides option was enabled to search for co-fragmented peptides in the MS/MS spectra. The "match between runs" option was enabled (without matching unidentified features) using a match time window of 0.7 min with an alignment window of 20 min. For protein quantification, a separate database searching was performed with digest mode set to specific. Protein and peptide quantification was performed using the label-free quantification (LFQ) algorithm with a minimum ratio count of 1, and minimum and average number of neighbors of 3 and 6, respectively. Reverse decoy and common contaminants matches were removed from the identification list. Peptides with local FDR

(posterior error probability (PEP)) < 0.05 were kept for further analysis. Database search using pFind was performed using parameters similar to those in PEAKS and MaxQuant with a 1% global FDR based on the target-decoy strategy. Only peptides identified by all three searching engines were kept for further analysis.

Semi-tryptic peptide mining

Peptides that do not have R or K (excluding protein N-terminal peptides, ~0.6% in our cases) in the amino acid before identified sequences were selected as semitryptic N-term peptides. Semitryptic C-term peptides were selected if the last amino acid of identified sequences lacks an R or K (excluding C terminus of the protein, ~2.2% in our cases). The in-source fragments were distinguished from proteolytic-derived semi-tryptic peptides based on elution time as previously reported [16]. The majority of in-source fragments showed different retention time as compared to their theoretical retention time (predicted using SSRCalc [28]) and gave the same retention time of their fully tryptic parental peptides. In addition, in-source fragments exhibit lower charge states than their corresponding parental peptides because the charge of the parental peptide is divided between fragments. Microbial semi-tryptic peptides were distinguished from human and food peptides based on the corresponding accession numbers from fasta sequence entries.

Taxonomy and functional analysis of peptides

Analysis was performed with UniPept (version 4.3.5) [29, 30] using UniProt 2020.01 based on the lowest common ancestor (LCA) algorithm with the following parameters: Equate I and L, Filter duplicate peptides, Advanced missing cleavage handling. The taxonomy information was visualized using a sunburst view provided by UniPept. Functional peptide annotations were performed using Gene Ontology (GO) terms and Enzyme Commission (EC) numbers.

Combining semi-tryptic and fully tryptic peptide data to quantify the degree of proteolysis

We determined alterations in the degree of proteolysis based on the normalized relative abundance of semi-tryptic peptides (NRASP) by normalizing semi-tryptic peptide based relative abundance to fully tryptic peptide based relative abundance. The logic is similar to that employed in other post-translational modification (PTM) studies, where PTM occupancy determination is achieved by measuring the abundances of both PTM and non-PTM peptides.

Nomenclature and sequence motif of cleavage sites

According to the nomenclature of Schechter and Berger [31], amino acids around the cleavage sites were numbered as P6-P5-P4-P3-P2-P1 ↓ P1'-P2'-P3'-P4'-P5'-P6', with the downward-pointing arrow indicating the cleaved peptide bonds between the P1 and P1' sites. We retrieved amino acids in P6-P6' of each *semi-tryptic peptide* from MaxQuant outputs (N-term cleavage window and C-term cleavage window of the peptides.txt file) using Microsoft Excel formulas and functions. The relative frequency of each amino acid at P6-P6' were calculated to determine the inter-group differences. To visualize conserved and

frequently occurring amino acids at positions flanking the cleavage site, sequence motif logos were generated using Weblogo [32]. Each logo consists of stacks of symbols, one stack for each position in the sequence. The height of symbols within the stack indicates the relative frequency of each amino acid at that position.

Statistical analysis

Multivariate analyses of the amino acid frequencies around the cleavage sites was performed using principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) with missing value imputed by Bayesian PCA (BPCA) [33]. Dunn-Bonferroni post-hoc procedure following Kruskal-Wallis test with a threshold of adjusted P -value <0.05 was employed to detect significantly different variables (present in at least 75% of samples) among groups using R (version 3.5.3) and RStudio (version 1.1.383). Beta-diversity of multi-omics data were determined using principal coordinate analysis (PCoA) using the Bray-Curtis distance [34]. The correlation between microbial semi-tryptic peptide intensity and multi-omics data was evaluated by Spearman rank correlation of their top three PCos (PCo1-PCo3).

Results

Identification of semi-tryptic peptides in metaproteomes

We analyzed two large-scale published datasets including 447 fecal (PXD008675, [7]) and 176 MLI metaproteomes (PXD007819, [15]). Both datasets were generated using high-resolution MS/MS, which allowed searching a large sequence space with a low FDR for semi-tryptic peptide identification. To increase microbial taxonomic coverage, we assembled sequences from a variety of culture-dependent sources such as UniProtKB, NCBI, CGR [20] and culture-independent sources such as IGC [18]. A total number of 12,828,005, 3,133,023, 2,948,562, and 2,757,990 MS/MS spectra were searched for the feces, ascending colon (AsC), descending colon (DeC), and terminal ileum (TI) MLI metaproteomes respectively, from which 3,804,903 (29.66%), 2,035,847 (64.98%), 1,917,761 (65.04%), and 1,808,732 (65.58%) peptide-spectrum matches (PSMs) were identified using PEAKS (1% FDR) in the first step large database search. To increase the sensitivity and confidence of large sequence space based metaproteomic analysis, we employed the two-step database searching strategy [14] and multiple search engines. Proteins identified with at least one peptide in the first step were reserved for the second round search using MaxQuant, PEAKS, and pFind. Only peptides identified by all three software were kept for further analysis, resulting in 125,494, 103,170, 106,243, and 92,784 peptides identified in the fecal, AsC, DeC, and TI metaproteomes respectively (Additional file 1: Tables S3-S18), among which 108,784 (86.68%), 76,325 (73.97%), 77,341 (72.79 %), and 65,002 (70.06%) peptides were assigned as microbial unique peptides (not shared by human or food sequences). Using UniPept, a total of 85,126 (78.25%), 67,288 (88.16%), 70,492 (91.14%), and 57,955 (89.16%) microbial peptides could get taxonomic and/or functional annotations in the fecal, AsC, DeC, and TI metaproteomes respectively. Despite of the comprehensiveness of IGC, which were frequently employed in previous metaproteomics studies, 11,540 (10.61%), 9,025 (11.82%), 9,129

(11.80%), and 7308 (11.24% TI) microbial peptides were only captured by UniProt/NCBI/CGR in the fecal, AsC, DeC, and TI metaproteomes respectively. This was probably because the UniProt/NCBI/CGR database are largely based on the translation of a completely sequenced single microorganism genome, the depth and assembly quality of which are significantly increased compared with that of gut microbial metagenomes. Among all identified microbial peptides, 28,525 (26.22%), 10,650 (13.95%), 9,357 (12.10%), and 10,614 (16.33%) peptides were semi-tryptic in the fecal, AsC, DeC, and TI metaproteomes respectively (Additional file 1: Tables S3 and S5-S7).

We identified 7,969 (6.35%), 14,869 (19.48%), 15,128 (14.24%), and 15,360 (16.55%) human specific peptides in the fecal, AsC, DeC, and TI metaproteomes respectively, among which, 5,104 (64.05%), 5,724 (38.50%), 5,254 (34.73%), and 6,825 (44.43%) peptides were semi-tryptic (Additional file 1: Tables S4 and S8-S10). Gene ontology (GO) analysis reveals that 84.13%, 79.97%, 81.74%, and 80.18 % of human semi-tryptic peptides were derived from potential extracellular proteins while only 1.16%, 0.80%, 0.73%, and 0.76% of microbial semi-tryptic peptides were assigned to potential extracellular proteins in the fecal, AsC, DeC, and TI metaproteomes respectively. The higher percentage of extracellular proteins, which are more susceptible to the gut luminal and mucosal proteases, contributed to the higher proportion of semi-tryptic peptides for human proteins.

Because many food resource such as pig, bovine, and other mammals share a large number of sequences with humans, the number of food unique semi-tryptic peptides were negligible, generally below 50 per sample (thus was not considered for further analysis), after excluding peptides shared by human. In addition, food semi-tryptic peptides represent complex proteolytic events across the entire digestive system because food proteins could be extensively proteolyzed by gastric pepsin, pancreatic proteases, and small intestinal exopeptidases before they reaches the large intestine (colon), where most gut microorganisms live.

Fecal metaproteomics reveals altered microbial proteolysis at different levels

We comprehensively investigated gut microbial proteolysis in terms of the taxonomic and functional distributions as well as the cleavage motif (Fig. 1 and 2). We first calculated the normalized relative abundance of semi-tryptic peptides (NRASP) to determine the relative degree of proteolysis of different microbial and functional groups. Fig. 1 illustrated the NRASP of 20 major taxonomic sub-groups, 35 major biological processes, and 32 enzyme sub-classes (Additional file 1: Tables S19-S21) identified in at least 75% of the 447 fecal metaproteomes from CD (n=204), UC (n=123), and control (n=120) groups.

The median values of NRASP of two dominant phyla *Firmicutes* and *Bacteroidetes*, two dominant classes *Bacteroidia* and *Clostridia*, two major orders *Bacteroidales* and *Clostridiales*, family *Bacteroidaceae*, as well as genus *Bacteroides* were around 1 with a very low individual variation, suggesting the relative abundance of the corresponding semi-tryptic peptides were comparable to that of fully tryptic peptides (Fig. 1a). However, the median of NRASP increased to approximately 1.25 for families *Lachnospiraceae* and *Ruminococcaceae*, and 1.5 for genera *Roseburia* and *Prevotella* as well as two abundant species *Faecalibacterium prausnitzii* and *Prevotella copri*, respectively, but reduced to

approximately 0.5 for phylum *Actinobacteria* and order *Bifidobacteriales* (including its single family member, *Bifidobacteriaceae*). Four out of 20 major taxonomic sub-groups exhibited significant inter-group difference (Dunn-Bonferroni post hoc, $P < 0.05$) in their NRASP (Additional file 1: Tables S19): family *Ruminococcaceae* and species *Prevotella copri* in CD have increased NRASP compared with the control group; genus *Faecalibacterium* and its sole known species *Faecalibacterium prausnitzii* in UC have reduced NRASP compared with the CD group (Fig. 2a).

The median of NRASP of most biological processes also fluctuated around 1 (Fig. 1b). However, these values increased to 1.75-2 for isoleucine biosynthetic process, valine biosynthetic processes, bacterial-type flagellum-dependent cell motility, protein transport, carboxylic acid metabolic process, fucose metabolic process, and glucose metabolic process, and further increased to 2.5 for fatty acid metabolic process and L-threonine catabolic process to glycine, but reduced to approximately 0.75 for polysaccharide catabolic process, carbohydrate transport and transmembrane transport, and further reduced to 0.3 for metabolic process, respectively. Six out of 35 major biological processes exhibited significant inter-group difference (Dunn-Bonferroni post hoc, $P < 0.05$) in their NRASP (Additional file 1: Tables S20): bacterial-type flagellum-dependent cell motility, polysaccharide catabolic process, anaerobic cobalamin biosynthetic process, and fructose 1,6-bisphosphate metabolic process in CD have increased NRASP compared with the control group, and translation and glycolytic process in UC have reduced NRASP compared with the control and UC group, respectively (Fig. 2b).

At enzyme level, 3-hydroxybutyryl-CoA dehydrogenase which is involved in butyrate metabolism showed the highest NRASP (median value > 3), followed by 3-hydroxyacyl-CoA dehydrogenase involved in fatty acid beta-oxidation, glycine C-acetyltransferase involved in L-threonine degradation, phosphoenolpyruvate carboxykinase (ATP) involved in gluconeogenesis, ketol-acid reductoisomerase (NADP(+)) involved in the biosynthesis of branched-chain amino acids (BCAA), and superoxide dismutase involved in tolerance to oxidant stress (median NRASP of 2-3, Fig 1c). Four out of 32 major enzyme sub-classes exhibited inter-group difference (Dunn-Bonferroni post hoc, $P < 0.05$) in their NRASP (Additional file 1: Table S21): superoxide dismutase in UC, protein-synthesizing GTPase in CD, and short-chain acyl-CoA dehydrogenase in both CD and UC have increased NRASP compared with the control group (Fig 2c).

We further investigated the microbial proteolytic cleavage site motif by calculating the amino acid frequencies at P6 - P6' position (Fig. 2d). Generally, alanine and valine were the most abundant amino acids in P1 position in different samples, while amino acids in other positions exhibited higher individual variation. The fecal metaproteome revealed a global alteration of microbial proteolytic motif in IBD (Fig. 2e and Additional file 1: Table S22). In the unsupervised hierarchical clustering, CD and UC clustered together (separated from control), sharing many alterations such as increased frequencies of leucine (P5 and P5') and tryptophan (P3 and P4') as well as decreased frequencies of alanine (P1'), methionine (P1' and P2'), aspartic acid (P3), and serine (P4). Cleavage motif of CD and UC also showed their unique alteration patterns in many cases. For instance, increased frequencies of leucine (P3, P2', and P4') and phenylalanine (P6, P3', and P5') were only observed in CD, while increased frequencies of histidine (P1

and P5'), serine (P1'), and arginine (P4) as well as decreased frequencies of valine (P2) and cysteine (P3' and P6') were only observed in UC. However, the large inter-individual variation did not allow for clear separation between groups in the principal component analysis (PCA, Fig. 2f) or partial least squares discriminant analysis (PLS-DA, Fig. 2g) of cleavage motif.

We also performed principal coordinates analysis (PCoA) of semi-tryptic peptide LFQ abundance using Bray-Curtis distance or Jaccard-based dissimilarity but did not result in clear group separation either (Additional file 1: Fig. S1). Although we used the “match between runs” of option MaxQaunt to increase transferred identification between separate LC-MS runs, the percentage of semi-tryptic peptides identified in more than 75% of samples were less than 0.05%.

Altered cleavage motif of human proteins were also observed (Additional file 1: Fig. S2 and Table S23). Similar to microbial cleavage motif, CD and UC clustered together and separated from control for human protein cleavage motif. Microbial proteins and human proteins can exhibit similar or reversed alteration trends in certain motif positions around the cleavage site in IBD (Additional file 1: Fig. S3). For instance, frequencies of glutamic acid, histidine, and three aromatic amino acids (phenylalanine, tryptophan and tyrosine) were increased in P5' in IBD for both microbial and human proteins.

MLI metaproteomics reveals location-specific alterations of microbial proteolysis

The fecal microbiome serve as a proxy for the gut luminal microbiota but is not fully representative of the mucosa-associated microbiota at the site of disease. Important complementary knowledge was acquired by systematically characterizing the MLI metaproteomes from different sites, which revealed remarkable location-specific alterations of microbial proteolysis signatures. A total of 57, 58, and 51 biological processes were identified in at least 75% of samples in AsC, DeC, and TI MLI metaproteome, respectively, among which 7 (12.28%), 10 (17.86%), and 10 (19.61%) biological processes exhibited significant inter-group difference in their NRASP (Additional file 1: Tables S24-S26). In the AsC metaproteomes, NRASP of regulation of translation, carbohydrate transport, DNA repair, protein secretion, generation of precursor metabolites and energy, and carboxylic acid metabolic process significantly increased in CD and/or UC (Fig. 3a). In the DeC metaproteomes, most alterations occurred in UC, including increased NRASP of ribosome biogenesis, terpenoid biosynthetic process, 'de novo' UMP biosynthetic process, cell division, and translational termination, as well as reduced NRASP of gluconeogenesis (Fig. 3b). In contrast, in the TI metaproteomes, most alterations occurred in CD, including increased carbohydrate transport, L-fucose catabolic process, translational termination, and ATP synthesis coupled proton transport as well as reduced NRASP of transcription, translation, protein folding, and carbohydrate metabolic process (Fig. 3c).

The microbial cleavage motif also revealed remarkable location-specific alterations in MLI metaproteomes (Fig. 4a-c). Similar to NRASP, microbial cleavage motif differed more in DeC and TI than AsC, where 41, 57 and 32 amino acid frequencies at a specific site exhibited significant inter-group differences, respectively ($P < 0.05$, Kruskal-Wallis and Dunn-Bonferroni test, Additional file 1: Tables S27-S29). In the unsupervised hierarchical clustering, CD and UC clustered together, separated from control in

the ascending colon (Fig. 4a); whereas UC and control clustered together, and separated from CD in the terminal ileum (Fig. 4c). The supervised partial least squares discriminant analysis (PLS-DA) also revealed that, on PC1 axis, UC partially separated from CD and control in the descending colon and CD clearly separated from the other two groups in the terminal ileum (Fig. 4d-f). Altered cleavage motif of human proteins were also observed in the MLI metaproteomes (Additional file 1: Fig. S4 and Tables S30-32). Similar to cleavage motif of microbial proteins, cleavage motif of human proteins in CD separated from the other two groups in the terminal ileum. Microbial proteins and human proteins can exhibit similar or reversed alteration trends in certain motif positions in IBD (Additional file 1: Fig. S5). For instance, increased frequencies of valine and leucine (P1) and reduced frequencies of two aromatic amino acids phenylalanine and tyrosine (P1) in IBD were observed in both microbial and human motif in MLI metaproteomes from different locations.

Association with microbial composition and proteases

To explore the potential relevance of gut microbial community structure (Additional file 1: Table S33) to proteolysis pattern, we associated the top three principal coordinates (PCo1-PCo3) of microbial semi-tryptic peptide LFQ intensity and the top three microbiome β -diversity principal coordinates (PCo1-PCo3, Fig. 5a) computed using Bray-Curtis distance. We found low correlations between semi-tryptic peptide LFQ intensity (PCo1) and β -diversity (PCo2 and PCo3) at different taxonomic levels ($-0.40 < \text{Spearman's rank correlation coefficient } (R) < 0.42$, $P < 5.6e-9$ for all pairwise associations, $n = 272$, Fig. 5c).

To investigate the association with microbial proteases, we resorted to the transcriptional abundance of microbial protease/peptidase (Fig. 5b, Additional file 1: Table S34) because protein-level abundance was inaccessible due to the limited sensitivity of current metaproteomics methodology. The microbial protease/peptidase transcriptome at feature (PCo3), species (PCo1), genus (PCo2), and family (PCo2) levels showed moderate correlations with semi-tryptic peptide LFQ intensity (PCo1), stronger than those of β -diversity (Fig. 5d, $-0.55 < R < 0.54$, $P < 2.6e-11$, $n = 184$). However, there was only low correlations at higher taxonomic levels (order, class, and phylum).

Association with host protease inhibitors and immunoglobulins

In addition to microbial variables, we also investigated the involvement of host factors. Human endogenous protease inhibitors are particularly present in the intestinal tract. Four human protease inhibitors (serpin A1, A3, B1, and B6) were identified in fecal metaproteomes. Serpin A1, A3, and B6 exhibited negative correlations ($-0.41 < R < -0.25$, $P < 7.9e-8$, $n = 447$) with the semi-tryptic peptide LFQ intensity (PCo1 and PCo2). To further investigate the effect of host factors on gut microbial proteolysis, we also analyzed the correlations between gut microbial proteolysis pattern and host immunoglobulins. IgG1, IgG4, and IgM were negatively correlated with PCo1 ($-0.44 < R < -0.25$, $P < 8.4e-8$, $n = 447$), and IgA was positively correlated with PCo2 ($R = 0.33$, $P < 8.4e-13$, $n = 447$) of semi tryptic peptide LFQ intensity, respectively.

Discussion

This study represents the first effort to discover IBD-associated alterations in gut microbial proteolytic signatures using a semi-tryptic peptide centric metaproteome mining approach, providing a different layer of information beyond taxonomic and protein abundances. The analysis was based on the hypothesis that similar degree of proteolysis should lead to similar abundance of semi-tryptic peptides. Using NRASP as an indicator, we observed that microbial semi-tryptic peptides in 447 fecal metaproteomes were enriched in several biological processes including fatty acid, carboxylic acid, glucose, and fucose metabolic processes, BCAA biosynthesis process, protein transport, and bacterial-type flagellum-dependent cell motility, suggesting they underwent more extensive proteolytic regulation. BCAAs (isoleucine, leucine, and valine) are important nutrients in bacterial physiology and BCAA biosynthesis pathway is essential for optimal growth of many bacteria [35-37]. In this study, we found that NRASP of BCAA biosynthesis process (1.75-2) is higher than that of non-BCAA (~1) in fecal metaproteomes. The higher NRASP may offer bacteria more adaptive flexibility in BCAA biosynthesis by proteolysis regulation. The higher NRASP of protein transport than carbohydrate transport and transmembrane transport could be rationalized by the fact that transport of proteins involved more proteolytic events.

One feature of inflammation-associated gut microbiota is the enrichment of motile bacteria, which can facilitate microbiota encroachment into the mucosa and drive inflammatory disease [38]. We found NRASP of flagellin-based flagellar cell motility was significantly increased in CD in the fecal metaproteomes ($P < 0.001$, Fig. 2b), which may reflect enriched levels of motile bacteria that have high ability to penetrate the mucus layer. Cobalamin (vitamin B₁₂) is an essential nutrient for higher animals and its biosynthesis is restricted to certain bacteria. Cobalamin deficiency is common in patients with CD not in UC [39]. Accordingly, we found NRASP of anaerobic cobalamin biosynthetic process was significantly increased in CD but not in UC ($P < 0.01$, Fig. 2b), implying that the altered proteolysis may be involved in cobalamin deficiency in CD. Gut bacteria can degrade dietary polysaccharides (fiber) and endogenous glycans such as those from host mucins as nutrients [40]. NRASP of polysaccharide catabolic process (resulting in polysaccharide breakdown), was increased in both UC and CD ($P < 0.001$, Fig. 2b), suggesting altered glycan utilization in IBD. Superoxide dismutase is a critical enzyme involved in controlling oxidative stress in any living system and may undergo proteolysis following oxidative modification [41]. In this study, we found NRASP of superoxide dismutase significantly increased in UC fecal metaproteomes ($P < 0.001$, Fig. 2c), suggesting increased oxidative stress under UC conditions.

We used the MLI data to illustrate the location-specific alterations of microbial proteolysis. Importantly, our findings are consistent with clinical symptoms of IBD within the gastrointestinal tract, where CD mainly occurs in *terminal ileum* [42] and UC mainly localizes in descending colon and rectum [43]. Accordingly, our results revealed that of CD differed from UC and Control in *terminal ileum* and UC differed from the other two groups in descending colon in terms of microbial proteolysis motif and NRASP of biological processes. In *terminal ileum*, NRASP of transcription, translation, and protein folding were all decreased in CD while NRASP of translational termination increased. Similarly, decreased NRASP of translation and increased NRASP of translation termination were observed in UC in descending colon.

These results represented the dysregulations of microbial protein synthesis and maturation in IBD. Both CD and UC has increased NRASP of terpenoid biosynthetic process in descending colon ($P < 0.0001$, Fig. 3b), suggesting elevated level of secondary metabolite synthesis in IBD. Other functional alterations mainly involved carbohydrate transport and metabolism such as galactose metabolic process, L-fucose catabolic process, glycolysis, and gluconeogenesis (Fig. 3).

In terms of host factors, we focused on human protease inhibitors and immunoglobulins. Generally, intestinal protease activity increases in IBD and protease inhibition has been proposed as new therapeutic strategy for IBD [44-46]. In our analysis, we found human endogenous protease inhibitors such as Serpins A1 significantly increased in IBD, particularly CD, fecal and MLI samples (Additional file 1: Fig. S6a), suggesting the increased protease activity triggers the production of a higher level of protease inhibitors in order to control the destructive nature of protease. There are many host mechanisms that have evolved to regulate host-microbiota interactions, and among these, one of the most widely studied is immunoglobulin A (IgA) [47, 48]. In addition, a recent study has reported that IgG selectively identifies pathobionts in pediatric IBD [49]. Altered microbial composition and metabolic products can trigger mucosal immune responses that mediate IBD [50, 51]. In our study, we did observe different immunoglobulins significantly increased in IBD (Additional file 1: Fig. S6b). Interestingly, microbial proteolysis exhibited negative associations with IgG and IgM but positive associations with IgA. In most cases, we found human immunoglobulins and protease inhibitors were negatively associated with the microbial proteolytic signatures, suggesting the inhibitory effects of these host factors on gut microbial proteolysis events. Taken together, dysregulated intestinal proteolytic homeostasis and immune response in IBD trigger the host to express high levels of protease inhibitors and immunoglobulins, which could in turn alter the proteolytic regulation in the gut microbiome.

Conclusions

Proteolysis regulation is an important strategy for gut microbial adaptation to the fluctuating intestinal environment. Alterations of the gut microbial proteolytic signatures in inflammatory bowel disease are highly diverse and divergent, thus highlighting the need for broader investigations to elucidate their functions. Our data also supports metaproteomics as a valuable approach to investigate the deeper regulation rules of the gut microbiota and host-microbial interactions.

Abbreviations

BCAA: branched-chain amino acids; CD: Crohn's disease; CGR: Culturable Genome Reference; IBD: inflammatory bowel disease; IGC: integrated gene catalog; LFQ: label-free quantification; MLI: mucosal luminal interface; NRASP: normalized relative abundance of semi-tryptic peptides; PCA: principle component analysis; PCoA: principal coordinate analysis; PEP: posterior error probability; UC: ulcerative colitis

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The mass spectrometry proteomics data are available in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008675 and PXD007819.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (31900070, ZX Yan), Guangdong Basic and Applied Basic Research Foundation (2019A1515011771, ZX Yan), and the Science and Technology Development Fund of Macau SAR (0098/2019/A2, R Yan).

Authors' contributions

ZY conceived and designed the study, performed the data analysis and drafted the manuscript. XL, LL, and JZ offered guidance on IBD disease subtyping. FX and RY helped to interpret the results. LC and FH downloaded the raw data. RY, HS, HH, DL, and YW helped in computing resources. ZY, HS, and XL edited the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

References

1. Mahmoud SA, Chien P. Regulated Proteolysis in Bacteria. *Annu Rev Biochem.* 2018;87:677–96.
2. Konovalova A, Søgaard-Andersen L, Kroos L. Regulated proteolysis in bacterial development. *FEMS Microbiol Rev.* 2014;38(3):493–522.
3. Deng CY, Zhang H, Wu Y, Ding LL, Pan Y, Sun ST, et al. Proteolysis of histidine kinase VgrS inhibits its autophosphorylation and promotes osmotic stress resistance in *Xanthomonas campestris*. *Nat*

- Commun. 2018;9(1):4791.
4. Kuroda A, Nomura K, Ohtomo R, Kato J, Ikeda T, Takiguchi N, et al. Role of inorganic polyphosphate in promoting ribosomal protein degradation by the Lon protease in *E. coli*. *Science*. 2001;293(5530):705–8.
 5. Christensen SK, Pedersen K, Hansen FG, Gerdes K. Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J Mol Biol*. 2003;332(4):809–19.
 6. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4(2):293–305.
 7. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569(7758):655–62.
 8. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol*. 2017;2:17004.
 9. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol*. 2018;3(3):337–46.
 10. Li F, Hitch TCA, Chen Y, Creevey CJ, Guan LL. Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle. *Microbiome*. 2019;7(1):6.
 11. Hettich RL, Pan C, Chourey K, Giannone RJ. Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal Chem*. 2013;85(9):4203–14.
 12. Blakeley-Ruiz JA, Erickson AR, Cantarel BL, Xiong W, Adams R, Jansson JK, et al. 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*. 2019;7(1):18.
 13. Xiong W, Brown CT, Morowitz MJ, Banfield JF, Hettich RL. Genome-resolved metaproteomic characterization of preterm infant gut microbiota development reveals species-specific metabolic shifts and variabilities during early life. *Microbiome*. 2017;5(1):72.
 14. Zhang X, Ning Z, Mayne J, Moore JI, Li J, Butcher J, et al. MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota. *Microbiome*. 2016;4(1):31.
 15. Zhang X, Deeke SA, Ning Z, Starr AE, Butcher J, Li J, et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat Commun*. 2018;9(1):2873.
 16. Kim JS, Monroe ME, Camp DG, Smith RD, Qian WJ. In-source fragmentation and the sources of partially tryptic peptides in shotgun proteomics. *J Proteome Res*. 2013;12(2):910–6.
 17. Shao W, Guo T, Toussaint NC, Xue P, Wagner U, Li L, et al. Comparative analysis of mRNA and protein degradation in prostate tissues indicates high stability of proteins. *Nat Commun*. 2019;10(1):2524.

18. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol.* 2014;32(8):834–41.
19. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature.* 2016;533(7604):543–46.
20. Zou Y, Xue W, Luo G, Deng Z, Qin P, Guo R, et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol.* 2019;37(2):179–85.
21. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010;464(7285):59–65.
22. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature.* 2017;550(7674):61–6.
23. Edgar RC. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460-1.
24. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics.* 2012;11(4):M111.010587.
25. 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.
26. Wang LH, Li DQ, Fu Y, Wang HP, Zhang JF, Yuan ZF, et al. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2007;21(18):2985–91.
27. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. *J Proteome Res.* 2011;10(4):1794–805.
28. Krokhin OV, Spicer V. Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem.* 2009;81(22):9522–30.
29. Mesuere B, Devreese B, Debyser G, Aerts M, Vandamme P, Dawyndt P. Unipept: tryptic peptide-based biodiversity analysis of metaproteome samples. *J Proteome Res.* 2012;11(12):5773–80.
30. Gurdeep Singh R, Tanca A, Palomba A, Van der Jeugt F, Verschaffelt P, Uzzau S, et al. Unipept 4.0: Functional Analysis of Metaproteome Data. *J Proteome Res.* 2019;18(2):606–15.
31. Schechter I, Berger A. (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 1967;27:157 – 62.
32. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Res.* 2004;14(6):1188–90.
33. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 2018;46(W1):W486–94.
34. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 2017;45(W1):W180–8.

35. Garault P, Letort C, Juillard V, Monnet V. Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Appl Environ Microbiol.* 2000;66(12):5128–33.
36. Nichols DS, Presser KA, Olley J, Ross T, McMeekin TA. Variation of branched-chain fatty acids marks the normal physiological range for growth in *Listeria monocytogenes*. *Appl Environ Microbiol.* 2000;68:2809–13.
37. Santiago B, MacGilvray M, Faustoferri RC, Quivey RG Jr. The branched-chain amino acid aminotransferase encoded by *ilvE* is involved in acid tolerance in *Streptococcus mutans*. *J Bacteriol.* 2012;194(8):2010–9.
38. Tran HQ, Ley RE, Gewirtz AT, Chassaing B. Flagellin-elicited adaptive immunity suppresses flagellated microbiota and vaccinates against chronic inflammatory diseases. *Nat Commun.* 2019;10(1):5650.
39. Gomollón F, Gargallo CJ, Muñoz JF, Vicente R, Lue A, Mir A, et al. Oral Cyanocobalamin is effective in the treatment of vitamin B12 deficiency in Crohn's Disease. *Nutrients.* 2017;9(3):308.
40. Cockburn DW, Koropatkin NM. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J Mol Biol.* 2016;428(16):3230–52.
41. Salo DC, Pacifici RE, Lin SW, Giulivi C, Davies KJ. Superoxide dismutase undergoes proteolysis and fragmentation following oxidative modification and inactivation. *J Biol Chem.* 1990;265(20):11919–27.
42. Caprilli R. Why does Crohn's disease usually occur in terminal ileum? *J Crohns Colitis.* 2008;2(4):352–56.
43. Ungaro R, Mehandru S, Allen PB, PeyrinBiroulet L, Colombel JF. Ulcerative colitis. *Lancet.* 2017;389(10080):1756–70.
44. Vergnolle N. Protease inhibition as new therapeutic strategy for GI diseases. *Gut.* 2016;65(7):1215–24.
45. Steck N, Mueller K, Schemann M, Haller D. Bacterial proteases in IBD and IBS. *Gut.* 2012;61:1610–18.
46. Midtvedt T, Zabarovsky E, Norin E, Bark J, Gizatullin R, Kashuba V, et al. Increase of faecal tryptic activity relates to changes in the intestinal microbiome: analysis of Crohn's disease with a multidisciplinary platform. *PLoS One.* 2013;8(6):e66074.
47. Petersen C, Bell R, Klag KA, Lee SH, Soto R, Ghazaryan A, et al. T cell-mediated regulation of the microbiota protects against obesity. *Science.* 2019;365(6451). pii: eaat9351.
48. Bunker JJ, Erickson SA, Flynn TM, Henry C, Koval JC, Meisel M, et al. Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science.* 2017;358(6361). pii: eaan6619.
49. Armstrong H, Alipour M, Valcheva R, Bording-Jorgensen M, Jovel J, Zaidi D, et al. Host immunoglobulin G selectively identifies pathobionts in pediatric inflammatory bowel diseases. *Microbiome.* 2019;7(1):1.
50. van der Post S, Jabbar KS, Birchenough G, Arike L, Akhtar N, Sjøvall H, et al. Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. *Gut.* 2019;68:2142–51.

51. Viladomiu M, Kivolowitz C, Abdulhamid A, Dogan B, Victorio D, Castellanos JG, et al. IgA-coated *E. coli* enriched in Crohn's disease spondyloarthritis promote TH17-dependent inflammation. *Sci Transl Med.* 2017;9(376):eaaf9655.

Figures

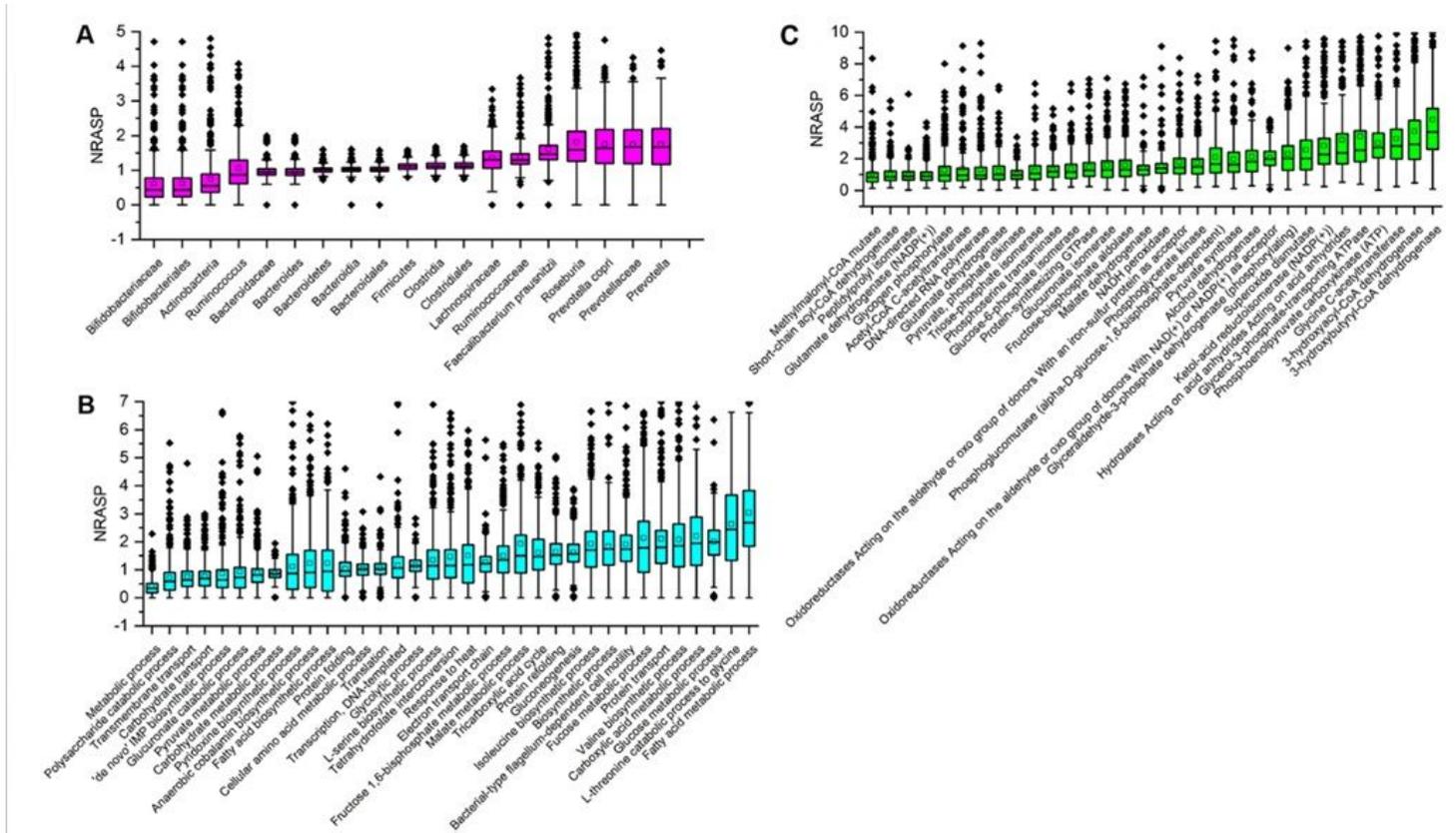


Figure 1

Normalized relative abundance of semi-tryptic peptides (NRASP, semi-tryptic peptide abundance/fully tryptic peptide abundance) derived from major microbial groups and biological processes in 447 fecal metaproteomics samples. Features are ranked in ascending order for different groups of bacteria (a), biological processes (b), and enzymes (c). Box plots represent the median (the line in the middle of the box), 25th and 75th percentiles, whiskers represent 1.5 times the interquartile range (IQR), and outliers are shown as dots.

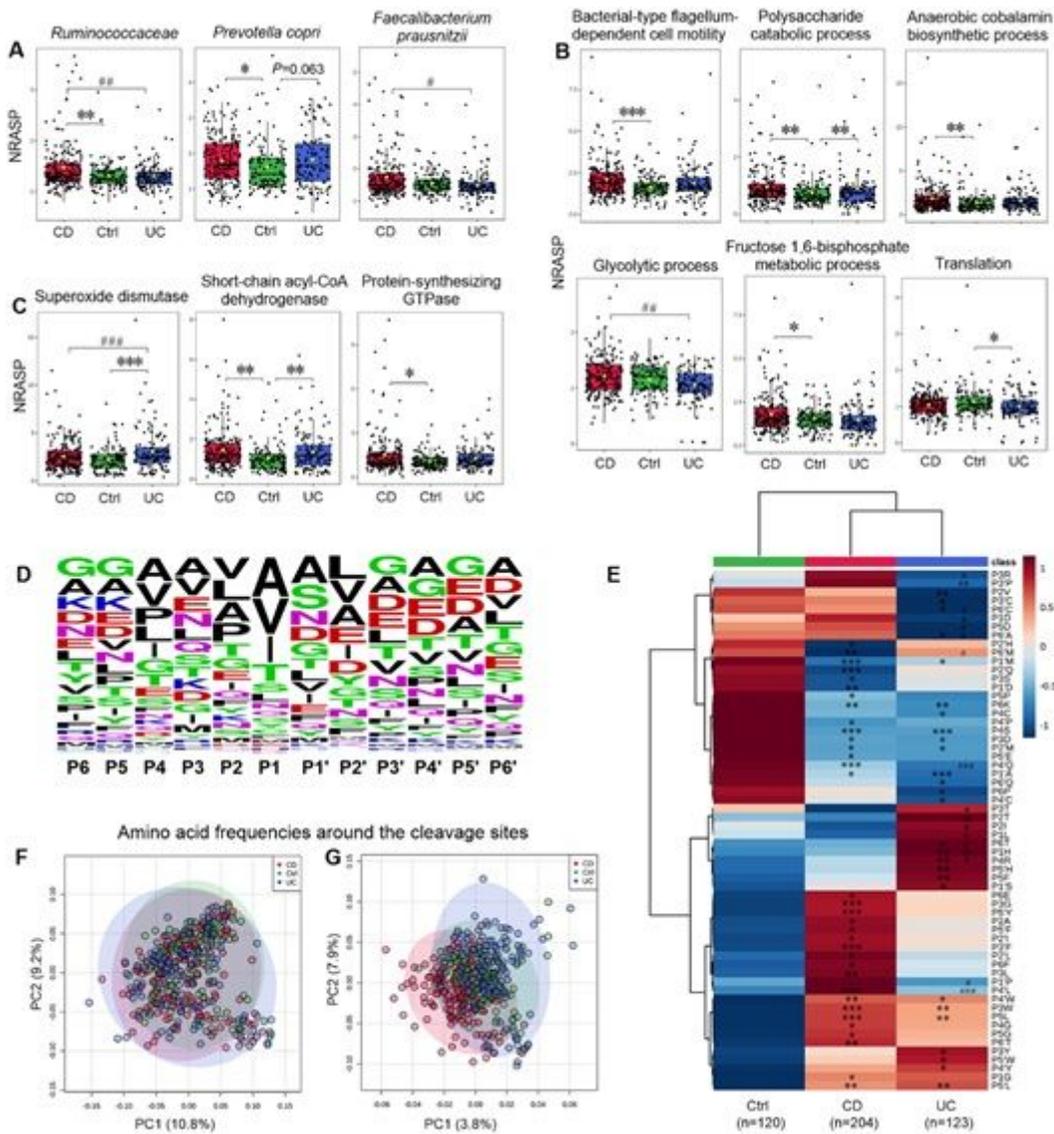


Figure 2

Signatures of altered microbial proteolysis at different levels in the fecal metaproteome of IBD. Normalized relative abundance of semi-tryptic peptides (NRASP) reveals altered microbial proteolysis at different taxonomic levels (a) as well as in different biological processes (b) and enzyme sub-classes (c). d WebLogo of proteolytic cleavage site depicting the amino acid conservation and frequency at each position. e Hierarchical clustering analysis of altered amino acid frequencies around the cleavage sites of microbial proteins in IBD. Principal component analysis (PCA, f) and partial least squares discriminant analysis (PLS-DA, g) of amino acid frequencies around the cleavage sites. Dunn-Bonferroni post-hoc analysis following Kruskal-Wallis test was employed to detect significant difference among three groups (CD, Ctrl, and UC). *P < 0.05 versus Ctrl; **P < 0.01 versus Ctrl; ***P < 0.001 versus Ctrl; # P < 0.05 (CD versus UC); ##P < 0.01 (CD versus UC); ### P < 0.001 (CD versus UC).

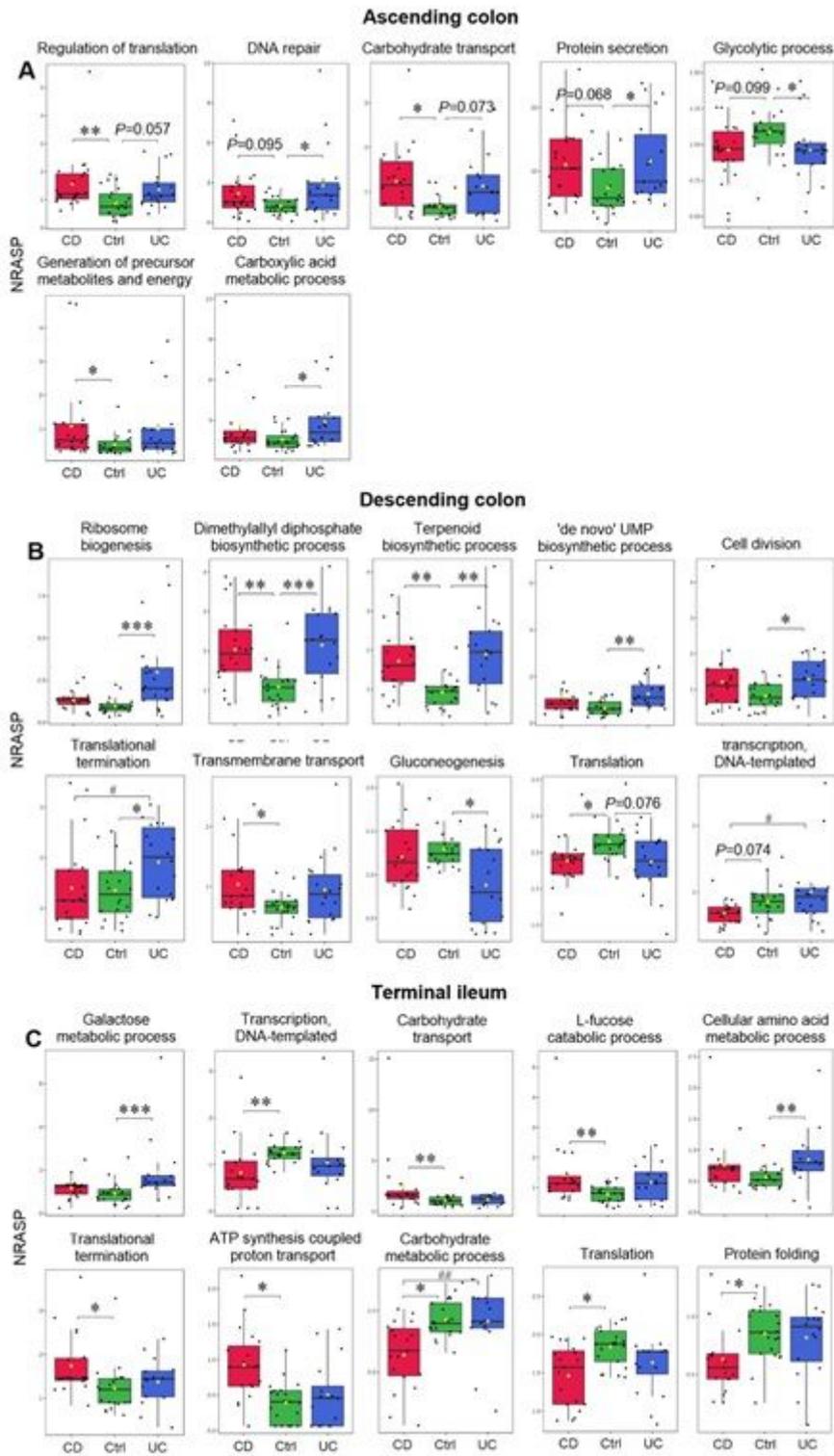


Figure 3

Signatures of altered microbial proteolysis at different intestinal locations in the mucosa-luminal interface metaproteome of IBD based on normalized relative abundance of semi-tryptic peptides (NRASP). a ascending colon, b descending colon, c terminal ileum. Dunn-Bonferroni post-hoc analysis following Kruskal-Wallis test was employed to detect significant difference among three groups (CD, Ctrl, and UC). * $P < 0.05$ versus Ctrl; ** $P < 0.01$ versus Ctrl; *** $P < 0.001$ versus Ctrl; # $P < 0.05$ (CD versus UC); ## $P < 0.01$ (CD versus UC); ### $P < 0.001$ (CD versus UC).

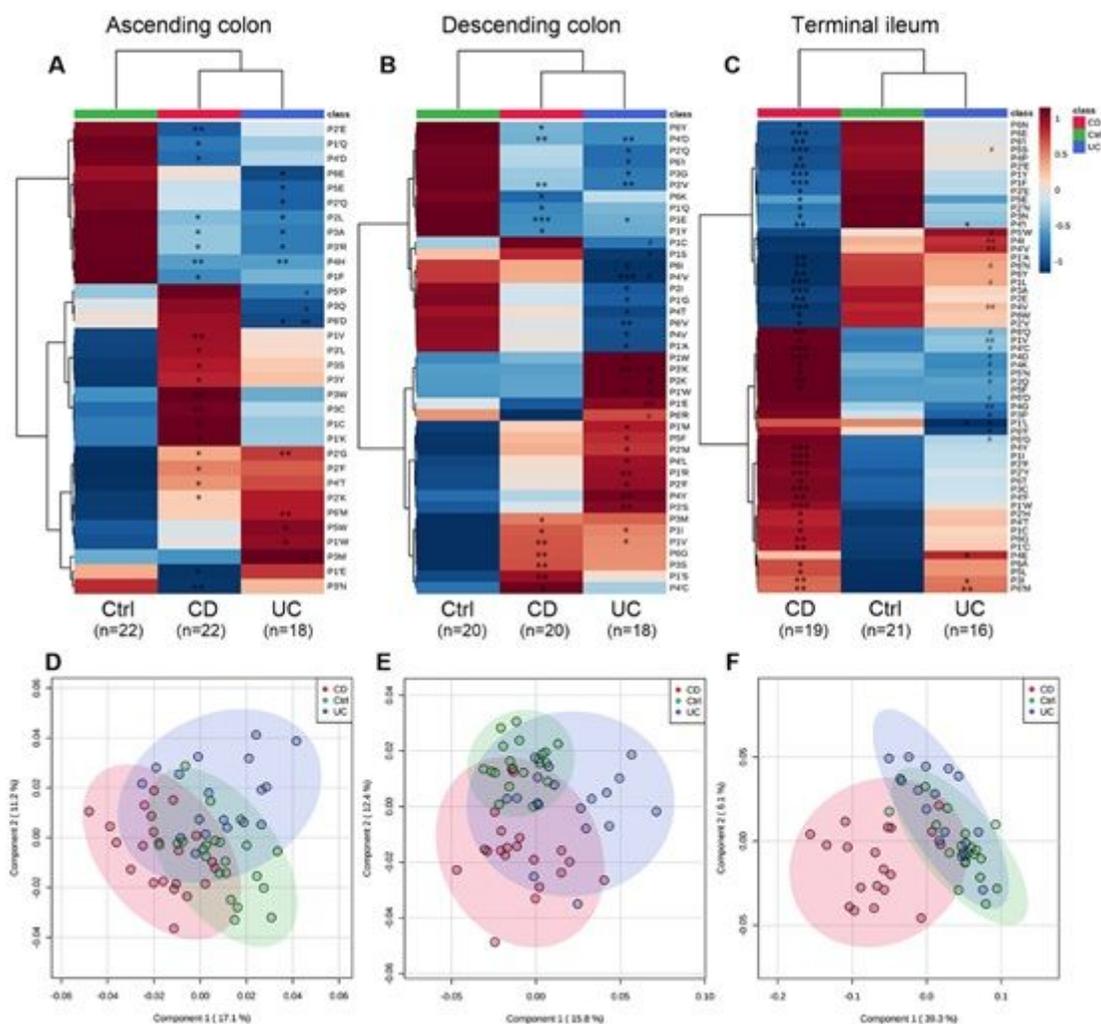


Figure 4

Location-specific alterations of gut microbial proteolytic motif in the MLI metaproteome of IBD. a and d ascending colon, b and e descending colon, c and f terminal ileum. Hierarchical clustering analysis (a-c) and partial least squares discriminant analysis (PLS-DA, d-f) of altered amino acid frequencies around the cleavage site in IBD. Group averages are shown in the heatmap. Dunn-Bonferroni post-hoc analysis following Kruskal-Wallis test was employed to detect significant difference among three groups (CD, Ctrl, and UC). *P < 0.05 versus Ctrl; **P < 0.01 versus Ctrl; ***P < 0.001 versus Ctrl; # P < 0.05 (CD versus UC); ##P < 0.01 (CD versus UC); ### P < 0.001 (CD versus UC).

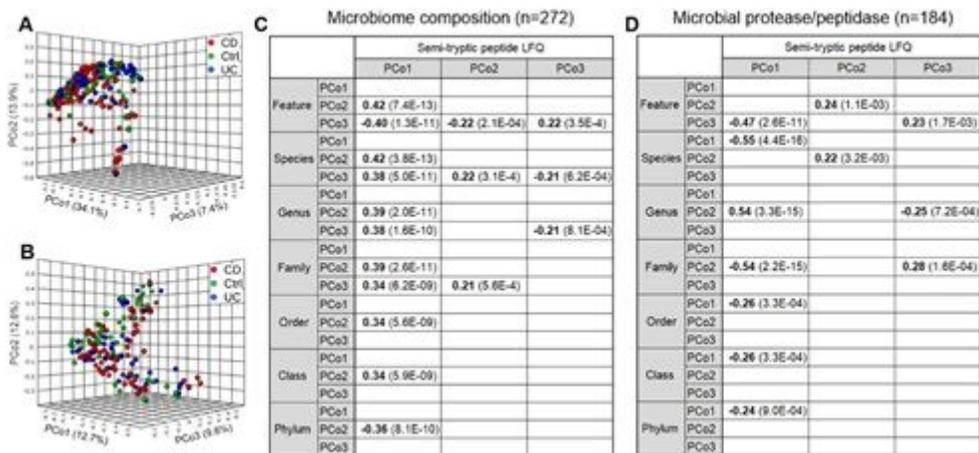


Figure 5

Correlation analysis between gut microbial proteolysis pattern, composition and protease expression. Principal coordinates analysis (PCoA) of metagenome (a) and microbial protease/peptidase transcriptome (b) is based on Bray-Curtis distance. Correlation analysis between microbial proteolytic signatures (PCo1-PCo3 of semi-tryptic peptide LFG), microbiome composition (n = 272, c) and protease/peptidase expression (n = 184, d) were calculated using Spearman's rank correlation based on PCo1-PCo3 of metagenome and metatranscriptome, respectively. Only significant correlations with coefficient >0.2 or <-0.2 (P < 0.05) are shown in c and d. The P value for each correlated pair is in parentheses.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1TableS1S34FigureS1S6.rar](#)