Metaproteomics of Colonic Microbiota Unveils Discrete Protein Functions among Colitic Mice and Control Groups

Dr. Clara Moon,
Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA

Dr. Gregory S. Stupp,
Department of Integrative Structural and Computational Biology The Scripps Research Institute La Jolla, CA, USA

Prof. Andrew I. Su, and
Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA.
Department of Integrative Structural and Computational Biology The Scripps Research Institute La Jolla, CA, USA

Prof. Dennis W. Wolan
Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA

Abstract

Metaproteomics can greatly assist established high-throughput sequencing methodologies to provide systems biological insights into the alterations of microbial protein functionalities correlated with disease-associated dysbiosis of the intestinal microbiota. Here, the authors utilize the well-characterized murine T cell transfer model of colitis to find specific changes within the intestinal luminal proteome associated with inflammation. MS proteomic analysis of colonic samples permitted the identification of ≈10 000–12 000 unique peptides that corresponded to 5610 protein clusters identified across three groups, including the colitic $Rag1^{-/-}$ T cell recipients, isogenic $Rag1^{-/-}$ controls, and wild-type mice. The authors demonstrate that the colitic mice exhibited a significant increase in *Proteobacteria* and *Verrucomicrobia* and show that such alterations in the microbial communities contributed to the enrichment of specific proteins with transcription and translation gene ontology terms. In combination with 16S sequencing, the authors' metaproteomics-based microbiome studies provide a foundation for assessing alterations in intestinal luminal protein functionalities in a robust and well-characterized mouse model of colitis, and set the stage for future studies to further explore the functional mechanisms of altered protein functionalities associated with dysbiosis and inflammation.

Keywords

colitis; ComPIL; metaproteomics; microbiome

1. Introduction

Inflammatory bowel disease (IBD) is a complex disorder caused by many variables, including host genetics, environmental factors, and the intestinal microbiota.[1–6] Numerous mouse models of colitis suggest an important role for the intestinal bacteria in disease propagation. For example, treatment of colitic mice with antibiotics or raising such inflammation-prone mice in a germ-free environment result in the amelioration and/or prevention of disease.[7,8] These results readily translate into the clinic, as human IBD patients are frequently prescribed antibiotics, such as metronidazole (flagyl) and ciprofloxacin.[9] Accordingly, alterations in the composition of the normal commensal gut microbiota, known as dysbiosis, is a major contributing factor to disease.[10,11] Metagenomics-based approaches have demonstrated that decreases in the *Bacteroidetes* and *Firmicutes* phyla, and increases in *Proteobacteria* and *Actinobacteria* are globally altered in IBD.[12] However, specific species with biological relevance to driving the disease phenotype have yet to be identified due to the inter-/intraindividual variation in microbial diversity.

We and others have developed technological advances in metaproteomics to interrogate and quantitate differences in microbial proteins, as protein functions are well conserved across commensal organisms and their hosts despite the tremendous ecological diversity within gut microbiomes.[13,14] Such key studies employing MS have laid the foundation for the field of microbiome metaproteomics.[15,16] Recently, comparisons between healthy and IBD human patient samples have been published. For example, Erickson et al. compared fecal samples from six twin pairs that included healthy, ileal Crohn's Disease (CD), and colonic CD patients. Metaproteomic data collected by LC–MS/MS identified 700–1250 or 1900–3000 protein clusters within the fecal samples when searched against the individual's matched metagenome or 51 human microbial isolate reference genomes, respectively.[17,18] The integration of metagenomics sequencing and reference genome sequences of common gut bacteria to assess metaproteomic data permitted this study to identify significant alterations in protein functionalities differentially represented in the healthy versus ileal CD patients (e.g., proteogenomics).

In this study, we expand upon the current metaproteomics techniques and employed multidimensional protein identification technology (MudPIT)[19] and the recently described Comprehensive Protein Identification Library (ComPIL)[20] to generate a compendium of host and microbial protein functions that are significantly correlated with intestinal inflammation. Using these methods, we found 10 000–12 000 unique peptides per treatment group that corresponded to a total of 5610 protein clusters. Bioinformatic analyses of colonic contents isolated from colitic mice and comparison of the metaproteomic composition to two different control groups delineated many unique host and microbial functionalities among all three groups, as well as interesting commonalities. Our results

demonstrate and support that a combination of variables, including host genetics, the intestinal gut microbiota, and inflammation importantly contribute to the alterations observed in the intestinal luminal proteome.

2. Experimental Section

2.1. Mice

Animal protocols were approved by The Institutional Animal Care and Use Committee at The Scripps Research Institute; The Institutional Animal Care and Use Committee protocol #:16-0023. 6-week old B6.129S7-$Rag1^{tm1mom}$/J ($Rag1^{-/-}$) mice (Rag) and wild-type C57BL/6J (WT) mice were cohoused for two weeks to normalize the microbiota (see Supporting Information, S1 for details). Briefly, CD3$^+$CD4$^+$CD8$^-$CD25$^-$Foxp3$^-$ naïve T cells from the spleens of donor $Foxp3$-$EGFP$ reporter mice were transferred retro-orbitally to five Rag mice to generate the T cell recipient mice group ("RT" mice), as previously described.[21,22] Control Rag ($n = 2$) and WT ($n = 5$) mice were injected retro-orbitally with a similar volume of sterile PBS (Supporting Information, Table S1) and all mice were separated by treatment group for the remainder of the experiment. Mice were euthanized on day 56 post T cell transfer (one mouse sacrificed prior to day 56 upon reaching 80% of initial weight). Colonic contents were collected by flushing with 5 mL of sterile PBS. Samples were filtered, aliquoted, centrifuged at $10\,000 \times g$ for 10 min, aspirated, and the resulting bacterial pellets were frozen at $-80\ °C$ until needed. Of note, our preparation methods may reduce accessibility to microbes adhered to food particulates and mucus due to the filtration of colonic samples and result in a bias towards the identification of proteins from luminal microbes. Segments of small intestinal and colonic tissues were fixed in 10% formalin for histology and crypt heights and the number of crypts per field were measured in a blinded manner (Supporting Information, Figure S1).

2.2. Sample Preparation for Unenriched Proteomic Analyses

After all RT mice were confirmed by histology to have inflammation in the ileum and colon (Supporting Information Table S1, Supporting Information, Figure S1), colonic content bacterial samples were thawed on ice, pooled by treatment group and were used immediately for proteomic sample preparation: bacterial pellet samples were resuspended in 500 μL of lysis buffer (Roche complete protease inhibitor tablet in PBS). Samples were lysed via sonication at 4 °C for 10 min, insoluble cellular material was removed via centrifugation ($10\,000 \times g$ for 5 min), and the remaining soluble protein concentration was measured (Pierce BCA Protein Assay Kit). Aliquots (100 μg) were subjected to trypsin digestion to generate peptides for MudPIT shotgun proteomics analysis, as previously described and subsequently stored at $-20\ °C$ until LC–MS/MS analysis (see Supporting Information, S1 for details).[19]

2.3. Liquid Chromatography–Tandem Mass Spectrometry

Trypsin-digested peptides were loaded onto a biphasic MudPIT column with C18 resin followed by strong cation exchange resin (see Supporting Information, S1 for details). Standard MudPIT MS/MS was performed using a Thermo LTQ Orbitrap XL mass spectrometer. Briefly, peptides were eluted at 0.2 mL/min using an 11-step MudPIT program, as previously described.[19] Precursor ions were recorded by scanning in the range

of *m/z* 400.00–1800.00 with the FTMS analyzer and a resolution of 60 000. The top eight peaks were selected for fragmentation using CID with normalized collision energy set to 35.0. Dynamic exclusion was enabled with a repeat count of 1, repeat duration of 30.00 s, exclusion list size of 100, and exclusion duration set to 60.00 s.

2.4. Peptide Identification using ComPIL

Precursor and fragmentation ion data were extracted from the Xcalibur RAW files via rawXtract 1.9.9.2 (http://fields.scripps.edu/yates/wp/?page_id=17) in the MS1 and MS2 file formats. The MS2 spectra were scored with Blazmass 0.9993 against peptides of the ComPIL database, containing over 80 million proteins from multiple microbial database sources as well as human, mouse, and plant proteins.[20] Both Blazmass and ComPIL source code are open source (https://github.com/sandipchatterjee/blazmass_compil). Settings for peptide scoring included: (1) a variable modification of oxidized methionine (+15.9949 Da), (2) a static modification for alkylated cysteine residues (+57.02146 Da), (3) a precursor mass tolerance of 10 and 50 ppm tolerance for fragmentation ions, and (4) two missed tryptic cleavages. Filtering was performed using DTASelect 2.1.3 (http://fields.scripps.edu/yates/wp/?page_id=17), requiring two peptides per protein and a false discovery rate (FDR) of 1% with respect to peptides.[23,24] FDR was assessed using the target-decoy approach with protein sequences reversed and concatenated with their original protein records.[20,25] The following parameters were used for filtering when run from the command line: "–quiet – brief –trypstat –modstat -y 2 -DM 10 –extra –dm –sfp 0.01 -p 2".

2.5. LC–MS/MS Data Analysis

The source code for this analysis is available online (https://github.com/stuppie/CM7_CM1E2d56col_unenr123_rawextract_2017/). Protein clustering, cluster taxonomy, and gene ontology (GO) term annotations were performed as previously described.[20,23] Briefly, protein loci were mapped to protein clusters using a pre-clustered version of ComPIL with a sequence identity threshold of 70%.[20] A protein cluster was annotated with all GO terms associated with any domain for all possible proteins within that cluster, while removing any GO terms that were parents ("is a" or "part of" relationships) of other GO terms in that protein cluster, using annotations generated from InterProScan v5 (version 5.8-49.0).[23] Protein cluster differential analysis was performed using DESeq2 1.14.1. The DESeq2 statistical analysis package allows for testing of differential expression in count data, and provides methods designed for dealing with overdispersion, features with low counts, and experiments with low numbers of biological replicates.[26] The method is briefly described as follows: spectral count data is modeled using the negative binomial distribution, which allows determination of the mean–variance relationship. Variance is estimated using an information sharing approach whereby a single feature's variance is estimated by taking into account information about variances of other similar features measured in the same experiment. Feature significance calling and ranking is then performed by estimating effect sizes, accounting for the logarithmic fold change (LFC) for a feature between treatment and control, and the noisiness of the LFC estimate. The coefficient of variation (CV) for the variance-stabilized transformed spectral counts was calculated by dividing the SD by the mean of spectral counts within each group of technical replicates and the median of the collective CV values is reported as the CV.

2.6. Gene Set Enrichment Analysis

Gene set enrichment analysis was performed using gseapy 0.7.6 (https://github.com/BioNinja/gseapy), a Python implementation of the Broad Institute's Gene Set Enrichment Analysis algorithm.[27] GO gene sets were generated from all identified protein clusters. Terms were subsequently filtered according to the MSigDB guidelines: (1) large sets, defined as those containing more than half the total number of protein clusters identified, were removed; (2) sets with less than five members were removed; (3) child terms with the exact same protein cluster members as their parent were removed; and (4) sibling terms with the exact same protein cluster members as other siblings were removed to generate one sibling. Significantly altered protein clusters (at a default p-value of <0.20) were ranked by DESeq2-determined shrunken LFC and analyzed. Significantly altered gene sets were called at an adjusted p-value of <0.05.

2.7. Taxonomy Analysis from Proteomic Data

Peptide spectral counts were normalized across all samples by a normalization factor of the total number of counts for one experiment divided by the median across all LC–MS/MS experiments, as previously described.[23] Briefly, the peptide taxonomy search space was restricted to the proteins identifiable in a given sample and analysis was performed at the phylum level. Each peptide was traced back to the phylum and if uniquely classifiable, the peptide was classified with a weight of normalized counts. Peptides without a discernible phylum (e.g., could belong more than one) were discarded from analysis. The normalized counts were then used to determine an approximate fractional taxonomic makeup of the sample.

2.8. 16S rDNA Illumina Sequencing and Analysis

Colonic content bacterial samples from mice were collected as described above. Genomic DNA was extracted using the Zymo Research Fecal DNA Mini Prep Kit (#11-322) according to manufacturer's instructions. Genomic DNA samples were submitted to the Scripps Research Next Generation Sequencing Core Facility for preparation of multiplexed amplicon libraries using the NEXTflex 16S V4 Amplicon-Seq Kit 2.0 (#4203-03), and sequencing using the 2×300-base pair protocol with the Illumina MiSeq platform. All analysis was performed using Quantitative Insights Into Microbial Ecology (QIIME, version 1.9.1).[28] Paired-end reads were assembled and quality screened, and had primers removed using PANDAseq 2.10.[29] Only sequences with length >200 bp were saved. The reads were clustered into operational taxonomic units using the open reference protocol using Greengenes 13.8.[30] Alpha and beta diversity analyses were conducted on data rarefied to 10 000 sequences per sample. Samples were clustered using the unweighted pair group method with arithmetic mean (UPGMA) method with weighted UniFrac as the distance metric.[31]

2.9. Availability of Data and Materials

The datasets generated and/or analyzed during the current study are available in this published article and its supplementary information files, as well as online. Analysis Code can be found at: https://github.com/stuppie/

CM7_CM1E2d56col_unenr123_rawextract_2017/. The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[32] with the project accession identifier PXD006384.

3. Results

3.1. Mouse Model of Colitis

We used the well-established murine T cell transfermodel of colitis for our studies.[21,22] Over the course of 6–8 weeks, the RT mice developed robust inflammation along the length of the intestinal tract. Because these mice were not littermates, we initially cohoused the *Rag1*[−/−] mice with WT C57BL/6 mice for 2 weeks prior to T cell transfer (or PBS control) to account for any microbiota differences in breeding colonies (Figure 1A). In addition, we designated WT mice and Rag mice as "healthy" control groups for all our proteomic comparisons (Supporting Information, Table S1). Upon T cell injection, we separated the mice by treatment group to permit the disease to progress in the RT mice unabated without the potential transfer of "healthy" microbes from the control groups (and vice versa).

3.2. Sample Preparation and LC–MS/MS Data Collection

We isolated and prepared intestinal bacteria from colonic content samples for LC–MS/MS proteomic analysis at the 8-week endpoint (Figure 1B). For each treatment group, bacterial pellets were combined and subsequently lysed by sonication, with five mice pooled for the WT and RT groups and two mice for the Rag group (Supporting Information, Table S1). Our goal for this initial study was to focus our metaproteomics data collection and analysis on technical replicates of pooled cohort samples and identify the core changes observed with inflammatory disease. Pooling the mouse samples minimized variability expected in the RT group, where differences in disease severity could lead to large variability between mice. However, aberrant proteins in individual mice may be missed due to dilution upon combining samples.

LC–MS/MS metaproteomics were performed on three technical replicates for each pooled treatment group and were highly reproducible, as demonstrated by a CV of 4.1% using the DESeq2 variance-stabilized transformed spectral counts.[33] The resulting proteomic data was searched against the ComPIL database[20,23] and principal component analysis based on the peptide composition among the LC–MS/MS datasets verified that each replicate for the individual treatment groups strongly clustered together (Figure 1C). A hierarchical clustering dendrogram based on the correlation distance calculated from the presence or absence and abundance of all peptides in a sample further demonstrated that the technical replicates were similar for each treatment group (Supporting Information, Figure S2). Jaccard clustering distances were also calculated and as expected were reduced across cohorts relative to the correlation distances due to the limited likelihood of identifying identical peptides across highly complex metaproteomic samples (Supporting Information, Figure S2). In total, we identified 10 676 unique peptides in the WT replicates, 12 073 unique peptides in the Rag replicates, and 10 852 unique peptides in the RT replicates (Figure 1D, Supporting Information, Figure S3). A total of 3863 unique peptides were

shared across all three cohorts, while each of the treatment groups consisted of ≈4000 unique and exclusive group-specific peptides (Figure 1D). Our proteomic data clearly delineated similarities and differences in the intestinal luminal proteome across the three groups. Importantly, the altered peptides were not exclusively variable among the "healthy" and colitic mice, but also between the WT and Rag controls.

3.3. Protein Clusters Differentially Expressed in the Rag and WT Controls

ComPIL was employed to facilitate functional identifications of proteins in each sample group and DESeq2 was used for differential analysis. We first investigated the differences between the Rag and WT control groups and identified a number of protein clusters that were significantly altered (Figures 2A and **3**A, Supporting Information, Tables S2 and S3.). Out of the combined 5610 protein clusters found among all samples, 22 murine and 680 microbial protein clusters were significantly altered with an FDR ‹0.05 and the absolute value of the $\log_2$ fold-change ($|\log_2FC|$) ›1. With respect to the altered murine protein clusters, 19 were increased in the WT samples (or missing in Rag mice), while three protein clusters were increased in the Rag samples (Figure 2A, Supporting Information, Table S2). Many of the murine proteins identified in these two groups were secreted proteins of either epithelial or immune cell origin. For example, components of the secretory IgA system are increased in the WT samples and absent in the Rag samples, including: (1) the immunoglobulin A heavy chain (IGHA); (2) the J chain (IgJ) which joins the IgA dimers; and (3) the epithelial polymeric immunoglobulin receptor (pIgR) which is cleaved and released along with the dimeric IgA as the secretory component. Additionally, several immunoglobulin light chains as well as the immunoglobulin G heavy chain are also detected only in proteomic samples generated from WT mice. This is expected as the Rag mice lack B cells (and therefore immunoglobulins), thus providing proof-of-principle that MudPIT proteomics of highly complex microbiome proteomes can distinguish differences in host proteins.

The 680 significantly altered microbial protein clusters were equally divided between the WT and Rag groups, with 350 and 330 differential protein clusters, respectively (Figure 3A, Supporting Information, Table S3). Gene set enrichment analysis on the microbial protein clusters reveals that the major biological processes (BP) enriched in the Rag mice are locomotion (GO:0040011), movement of cell/subcellular component (GO:0006928), and cell motility (GO:0048870) (Figure 3A, Supporting Information, Table S4), in agreement with the observance of numerous flagellin proteins significantly increased in our Rag samples. There was a significant enrichment of protein clusters associated with localization as well, in which the majority of the protein clusters associated with this GO term were transporters. Accordingly, molecular function (MF) GO terms for Rag samples showed a higher level of transporter activity (GO:0005215). The majority of the BP GO terms enriched in the WT samples are associated with RNA metabolic processes (including RNA processing (GO:0006396), RNA biosynthetic process (GO:0032774), mRNA metabolic process (GO:0016071), and ncRNA metabolic process (GO:0034660)) and translation (GO:0006412). Analysis of the MF GO terms changed in WT samples show a similar signature, with terms including, but not limited to, RNA polymerase activity (GO:0097747), DNA binding (GO:0003677), RNA binding (GO:0003723), and structural constituent of ribosome

(GO:0003735). Overall, the Rag gut microbial proteome shows a vast increase in motility/flagellin proteins and transporter activity, whereas there is an enrichment in RNA and protein synthesis in the WT gut microbial proteome.

3.4. Protein Clusters Significantly Altered between Colitic and Healthy Mice

To distinguish potentially important proteins that are involved in intestinal inflammation and not due to the *Rag1*$^{-/-}$ background (i.e., baseline differences between the WT vs Rag host proteins and/or gut microbiota), we focused our efforts on the identification of aberrant protein clusters in RT samples that were significantly altered between this treatment group and the combination of WT and Rag samples (hereafter called "controls") (Figures 2B and 3B, Supporting Information, Tables S5 and S6). A total of 66 murine and 309 microbial protein clusters were significantly altered (FDR < 0.05 and |log$_2$FC| > 1). Of the significantly altered murine protein clusters, six were increased in the control samples, and 54 were increased in the RT samples (Figure 2B, Supporting Information, Table S5). Interestingly, many of the murine proteins significantly increased in the RT samples are immune response-related genes, as would be expected with inflammation. Some examples include myeloperoxidase, granzyme A, lactoferrin, and the pro-inflammatory proteins S100A8 and S100A9.[34] Two of the most highly increased proteins are the two S100A proteins, which are abundantly found in neutrophils and other myeloid cells, and together form the damage-associated molecular pattern molecule calprotectin. This heterodimer has a wide variety of functions, including immune cell activation via binding and activation of toll like-receptor 4 and the receptor for advanced glycation end products, calcium and zinc binding, and antimicrobial properties.[34–36] Elevated levels of calprotectin have been implicated in a variety of diseases such as cancer and IBD, and (along with myeloperoxidase and lactoferrin) has been used as a fecal marker of intestinal inflammation in IBD.[37]

In addition, we observe an increase in many protease inhibitors, including several serpins (Serpina3k, Serpinc1, Serpina3f, Serping1, Serpina1a), MUG1, and PZP. This suggests an elevation of proteolytic activity that is usually associated with inflammation, and a host response attempting to suppress the elevated proteolytic activity. Furthermore, Serpina1 (a1-antitrypsin) has long been used as a fecal marker of intestinal inflammation in IBD.[37,38]

It is well described that intestinal inflammation also results in epithelial responses. Various cytokines have been shown to increase intestinal epithelial cell production/secretion of antimicrobial peptides and mucus.[39,40] Accordingly, we observe an enrichment of several antimicrobial REG peptides (REG1, REG2, REG3A, REG3B) as well as MUC2 in the RT proteomic datasets. Together, this data shows that our metaproteomics methodology permits the identification and relative quantitation of many secreted host proteins known to be associated with intestinal inflammation.

Of the 309 significantly altered microbial protein clusters, 174 were increased in the control samples, and 135 were increased in the RT samples (Figure 3B, Supporting Information, Table S6). GO term analysis on the microbial proteins shows that the MF GO terms enriched in control samples include hydrolase activity acting on *O*-glycosyl bonds (GO:0004553), coenzyme binding (GO:0050662), and electron carrier activity (GO:0009055) (Supporting Information, Table S7). In contrast, the MF GO terms in higher abundance in RT mice

include RNA binding (GO:0003723), phosphoglycerate kinase activity (GO:0004618), and substrate-specific transporter activity (GO:0022892). Accordingly, we observe many phosphoglycerate kinases involved in glycolysis, as well as maltose transporters enriched in the RT samples.

Many of the significantly altered microbial proteins and GO terms enriched in the RT samples overlapped with functional signatures identified in either the Rag or WT samples. Therefore, to further interrogate these similarities and differentiate the functions unique to RT among the three cohorts, we performed additional comparisons between RT versus WT (Figures 2C and 3C, Supporting Information, Tables S8–S10) and RT versus Rag (Figures 2D and 3D, Supporting Information, Tables S11–S13). These analyses uncovered protein functionalities unique with respect to the RT samples, including the MF GO terms phosphoglycerate kinase (GO:0004618) and substrate-specific transporter activity (GO: 0022892) as mentioned above, along with the BP GO terms monosaccharide metabolic process (GO:0005996) and single-organism localization (GO:1902578), among others. Notwithstanding, the RT mice shared enriched GO term functions with the Rag mice that were diminished and/or missing in WT (including locomotion (GO:0040011), movement of cell/subcellular component (GO:0006928), and cell motility (GO:0048870)), as well as numerous microbial protein functionalities matching the WT mice that were absent and/or reduced in the Rag group (namely those involved in transcription and translation).

3.5. Taxonomic Comparison with 16S rDNA Sequencing and Metaproteomic Mapping

We performed 16S rDNA sequencing on the same samples employed for proteomic analysis (Figure 4A, Supporting Information, Table S14). In agreement with the field, we observed that *Firmicutes* and *Bacteroidetes* comprised the majority of the commensal gut microbiota in the mice. Furthermore, we observed an altered ratio of these phyla in the RT mice, with an increase in *Proteobacteria* and *Verrucomicrobia*, which has been described in other mouse models of colitis as well as in IBD patients.[3,12,41,42] Importantly, these sequencing results correlate exceedingly well with the microbial composition analyses generated from our LC–MS/MS proteomics data at the phylum level (Figure 4B). Importantly, the measurable increase in *Proteobacteria* and *Verrucomicrobia* in the RT samples was detected by both sequencing and proteomics methods. These results strongly support that the assessment of bacterial composition with proteomic data correlates well with the standard in the field of 16S sequencing.

4. Discussion

For our T cell transfer colitis studies, we utilized two control groups to compare with the RT mice and help identify proteins important in colitis: (1) isogenic *Rag1*$^{-/-}$ mice that did not receive T cells and (2) C57BL6/J WT mice that have a commensal gut microbiota not altered by the lack of adaptive immune cells. Both PCA analysis and hierarchical clustering demonstrated distinct proteomic differences across these two control groups (Figure 1C, Supporting Information, Table S2) and analysis at the peptide level shows ≈50% overlap between Rag and WT samples (Figure 1D). Among the most striking differences between the two control groups was the enrichment in motility/flagellar proteins in the Rag gut

microbial proteome, compared to a robust enrichment in transcription- and translation-related proteins in the WT gut microbial proteome (Supporting Information, Table S3). The majority of the significantly altered proteins that could be traced back to a specific bacterial phylum were mapped to *Firmicutes* and *Bacteroidetes*, similar to what we observed for the total taxonomic make-up by 16S sequencing and proteomic analyses (Figure 4). These results show that while there are moderate changes in the microbial composition between the WT and immunodeficient Rag mice, there are tremendous alterations in microbial proteins and microbial protein functionalities represented in these mice, even in the absence of inflammation.

Similar to the striking locomotion/motility protein signature enriched in the Rag mice compared to WT, the RT mice also showed an abundance of flagellar proteins compared to WT mice (Supporting Information, Tables S9 and S10). Previous studies using *Tlr5*[−/−] mice demonstrated that the absence of TLR5 resulted in the lack of flagellin-specific immunoglobulin, which allowed for the over-expression of flagellin by commensals.[43] Conversely, the presence of the anti-flagellin antibodies inhibited microbial motility, leading to the downregulation of flagellar proteins by the bacteria. A similar mechanism may be occurring in the Rag and RT mice in our studies, with the lack of B cells (and therefore immunoglobulins) leading to an increase of the normally low flagellin levels. Analysis of the murine proteins identified by our proteomic analyses show a B cell-deficient signature as expected, with a significant and notable absence of immunoglobulin proteins in the Rag and RT samples compared to WT. Interestingly, bacterial flagellins have been associated with pathogenicity, where pathogenic bacteria produce flagella to promote colonization and invasion of the mucosa.[44] The increased expression of bacterial flagellins in these mice may be another indication of a change in the commensal microbial population towards a more dysbiotic state. Furthermore, an increase in antiflagellar antibodies have been associated with IBD, specifically CD,[45] demonstrating a host response to this immune-reactive protein. Future studies utilizing a cotransfer of B cells in this T cell transfer colitis model can help distinguish the contributions of intestinal inflammation versus genetic background in the expansion of bacterial flagellins in these mice.

Analysis of microbial protein-associated GO terms also revealed prominent similarities between the RT and WT samples, particularly with respect to GO terms pertaining to RNA and protein synthesis (Supporting Information, Tables S4, S7, and S13). We posit that the GO terms shared among the RT and WT mice are contributable to bacterial growth; however, the parental bacteria synthesizing the proteins are vastly different between the two groups. Both metaproteomics and 16S sequencing revealed that the inflammatory state of RT mice significantly alters the microbial phylogenetic composition, with a significantly increased footprint of *Proteobacteria* in the RT samples (Figure 4). This phylum is known to expand in other intestinal inflammation models.[3,12] Therefore, the overabundance of microbial proteins involved in bacterial replication and protein synthesis likely represent the ability of *Proteobacteria* to thrive in the inflammatory state of RT mice, as previously shown. [46] Conversely, these same protein functionalities that dominate the microbial proteome in WT mice during homeostasis are attributable to the commensal organisms belonging to the *Firmicutes* and *Bacteroidetes* phyla (Figure 3). In support of this hypothesis, we find that the majority of the bacterial proteins associated with these GO terms in the RT samples can be

mapped back to *Proteobacteria*, while the WT sample proteins associated with these GO terms can be mapped to *Firmicutes* and *Bacteroidetes* (Supporting Information, Tables S3, S6, and S12). We do not observe an increase in transcriptional/translational proteins from the Rag microbiota relative to WT and RT mice, and this finding may be due to the significant differences in flagellar proteins that supersede the identification of those proteins associated with growth and division.

Analysis of the multiple comparisons identified protein functionalities unique to the RT group as well. As would be expected, many murine proteins associated with inflammation were significantly elevated in the RT group compared to either of the control groups, including both subunits of calprotectin, several protease inhibitors, and a variety of antimicrobial peptides (Supporting Information, Tables 5, 8, 11). In addition, microbial protein functionalities were found specifically enriched in the RT group as well, including the MF GO term substrate-specific transporter activity (Supporting Information, Tables 7, 10, 13). One class of proteins that corresponded to this enriched GO term included the maltose transporters (Supporting Information, Tables 6, 9, 12). Studies elucidating the maltose transport system using *E. coli* have identified several components involved in the transport and utilization of maltose and maltodextrins, which is ultimately broken down to glucose by several enzymes.[47,48] Furthermore, glucose starvation leads to elevated expression of the maltose system genes, whereas bacterial cultures grown in high glucose concentrations leads to the block in expression of the *mal* genes.[47] The enrichment of maltose transporters and others in the RT associated bacteria may be an indicator of an increased need for or utilization of maltose and/or glucose in the inflamed state.

Due to the expected variability among technical replicates observed in LC–MS/MS-based proteomic approaches, as well as to minimize variability within a given treatment group, our goal for this initial study was to focus our metaproteomics data collection and analysis on technical replicates of pooled cohort samples. While there is a possibility that aberrant proteins in individual mice may be missed due to dilution upon sample pooling, the composite collection permits a general survey for distinct bacterial proteins and functions unique to an individual group that create a foundation on which we can begin to invest in and explore biological replicates. Altogether, our results demonstrate the important contributions of host genetics, the gut microbiota, and inflammation to the altered intestinal proteome observed in the RT mice, and support the combined use of 16S sequencing and metaproteomics methods incorporating an independent ComPIL protein database to microbiome studies.[20] These data provide a foundation to study the alterations in intestinal luminal protein functionalities in a robust and well-characterized mouse model of colitis, elucidate alterations in disease severity across biological replicates, and explore the functional mechanisms of some of these altered protein functionalities.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Abbreviations

BP	biological process
CD	Crohn's disease
ComPIL	Comprehensive Protein Identification Library
GSEA	gene set enrichment analysis
IBD	inflammatory bowel disease
MF	molecular function
Rag	Rag1$^{-/-}$ mice
RT	Rag1$^{-/-}$ T cell recipient mice
WT	wild-type mice

References

1. Cadwell K, Patel KK, Maloney NS, Liu TC, Ng A. Cell. 2010; 141:1135. [PubMed: 20602997]
2. Rogler G. Dig Diseases. 2011; 29:136.
3. Elson CO, Cong Y. Gut Microbes. 2012; 3:332. [PubMed: 22572873]
4. Leone V, Chang EB, Devkota S. J Gastroenterol. 2013; 48:315. [PubMed: 23475322]
5. Graham DB, Xavier RJ. Trends Immunol. 2013; 34:371. [PubMed: 23639549]
6. Xavier RJ, Podolsky DK. Nature. 2007; 448:427. [PubMed: 17653185]
7. Kang SS, Bloom SM, Norian LA, Geske MJ, Flavell RA, Stappenbeck TS, Allen PM. PLoS Med. 2008; 5:e41. [PubMed: 18318596]
8. Elson CO, Cong Y, McCracken VJ, Dimmitt RA, Lorenz RG, Weaver CT. Immunol Rev. 2005; 206:260. [PubMed: 16048554]
9. Sartor RB. Proc Nat Acad Sci USA. 2008; 105:16413. [PubMed: 18948599]
10. Dalal SR, Chang EB. J Clin Invest. 2014; 124:4190. [PubMed: 25083986]
11. Huttenhower C, Kostic AD, Xavier RJ. Immunity. 2014; 40:843. [PubMed: 24950204]
12. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Proc Nat Acad Sci USA. 2007; 104:13780. [PubMed: 17699621]
13. Human Microbiome Project Consortium. Nature. 2012; 486:207. [PubMed: 22699609]
14. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Nature. 2012; 489:220. [PubMed: 22972295]
15. Kolmeder CAC, de Been MM, Nikkilä JJ, Ritamo II, Mättö JJ, Valmu LL, Salojärvi JJ, Palva AA, Salonen AA, de Vos WMW. PLoS One. 2012; 7:e29913. [PubMed: 22279554]
16. Li XX, Leblanc JJ, Truong AA, Vuthoori RR, Chen SSS, Lustgarten JLJ, Roth BB, Allard JJ, Ippoliti AA, Presley LLL, Borneman JJ, Bigbee WLW, Gopalakrishnan VV, Graeber TGT, Elashoff DD, Braun JJ, Goodglick LL. PLoS One. 2011; 6:e26542. [PubMed: 22132074]

17. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK. ISME J. 2009; 3:179. [PubMed: 18971961]

18. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, Shah M, Halfvarson J, Tysk C, Henrissat B, Raes J, Verberkmoes NC, Fraser CM, Hettich RL, Jansson JK. PLoS One. 2012; 7:e49138. [PubMed: 23209564]

19. Wolters DAD, Washburn MPM, Yates JRJ. Anal Chem. 2001; 73:5683. [PubMed: 11774908]

20. Chatterjee S, Stupp GS, Park SKR, Ducom JC, Yates JR, Su AI, Wolan DW. BMC Genomics. 2016; 17:642. [PubMed: 27528457]

21. Powrie F, Leach MW, Mauze S, Caddie LB, Coffman RL. Int Immun. 1993; 5:1461. [PubMed: 7903159]

22. Ostanin DV, Bao J, Koboziev I, Gray L, Robinson-Jackson SA, Kosloski-Davidson M, Price VH, Grisham MB. Am J Physiol Gastrointest Liver Physiol. 2009; 296:G135. [PubMed: 19033538]

23. Mayers MD, Moon C, Stupp GS, Su AI, Wolan DW. J Proteome Res. 2017; 16:1014. [PubMed: 28052195]

24. Wu J, Zhu J, Yin H, Liu X, An M, Pudlo NA, Martens EC, Chen GY, Lubman DM. J Proteome Res. 2016; 15:3635. [PubMed: 27559751]

25. Elias JE, Gygi SP. Nat Methods. 2007; 4:207. [PubMed: 17327847]

26. Love MI, Huber W, Anders S. Genome Biol. 2014; 15:550. [PubMed: 25516281]

27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Proc Nat Acad Sci USA. 2005; 102:15545. [PubMed: 16199517]

28. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. Nat Chem Biol. 2010; 7:335.

29. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. BMC Bioinformatics. 2012; 13:31. [PubMed: 22333067]

30. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Appl Environ Microbiol. 2006; 72:5069. [PubMed: 16820507]

31. Lozupone C, Knight R. Appl Environ Microbiol. 2005; 71:8228. [PubMed: 16332807]

32. Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H. Nucleic Acids Res. 2016; 44:11033. [PubMed: 27683222]

33. Langley SR, Mayr M. J Proteomics. 2015; 129:83. [PubMed: 26193490]

34. Vogl T, Gharibyan AL, Morozova-Roche LA. Int J Mol Sci. 2012; 13:2893. [PubMed: 22489132]

35. Stríz I, Trebichavský I. Physiol Res. 2004; 53:245. [PubMed: 15209531]

36. Srikrishna G. J Innate Immun. 2012; 4:31. [PubMed: 21912088]

37. Foell D, Wittkowski H, Roth J. Gut. 2009; 58:859. [PubMed: 19136508]

38. Fischbach W, Becker W, Mössner J, Koch W, Reiners C. Gut. 1987; 28:386. [PubMed: 3495470]

39. Onyiah JC, Colgan SP. Cell Mol Life Sci. 2016; 73:4203. [PubMed: 27271753]

40. Mukherjee S, Hooper LV. Immunity. 2015; 42:28. [PubMed: 25607457]

41. Berry D, Kuzyk O, Rauch I, Heider S, Schwab C, Hainzl E, Decker T, Müller M, Strobl B, Schleper C, Urich T, Wagner M, Kenner L, Loy A. Front Microbiol. 2015; 6:1408. [PubMed: 26697002]

42. Berry D, Schwab C, Milinovich G, Reichert J, Ben Mahfoudh K, Decker T, Engel M, Hai B, Hainzl E, Heider S, Kenner L, Müller M, Rauch I, Strobl B, Wagner M, Schleper C, Urich T, Loy A. ISME J. 2012; 6:2091. [PubMed: 22572638]

43. Cullender TC, Chassaing B, Janzon A, Kumar K. Cell Host Microbe. 2013; 14:571. [PubMed: 24237702]

44. Ramos HC, Rumbo M, Sirard JC. Trends Microbiol. 2004; 12:509. [PubMed: 15488392]

45. Alexander KL, Targan SR, Elson CO. Immunol Rev. 2014; 260:206. [PubMed: 24942691]

46. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, Keestra AM, Laughlin RC, Gomez G, Wu J, Lawhon SD, Popova IE, Parikh SJ, Adams LG, Tsolis RM, Stewart VJ, Bäumler AJ. Science. 2013; 339:708. [PubMed: 23393266]

47. Boos W, Shuman H. Microbiol Mol Biol Rev. 1998; 62:204. [PubMed: 9529892]

48. Nikaido H. FEBS Lett. 1994; 346:55. [PubMed: 8206159]

Significance of the study

The commensal gutmicrobiota is essential for maintaining health and has a primary role in digestion/metabolism, homeostasis, and protection from pathogenic bacteria. High-throughput sequencing has established *Bacteroidetes, Firmicutes*, *Proteobacteria*, and *Actinobacteria* as the four major bacterial phyla that comprise the ecological makeup of the intestinal microbiota. However, the tremendous inter-/intravariability in microbial composition across individuals, as well as along the length of the intestinal tract has made it difficult to definitively ascertain specific bacterial species associated with health or as drivers of disease states, including inflammatory bowel disease. In this study, we expanded upon the current metaproteomics techniques and use the robust and highly reproducible murine T cell transfermodel of colitis as well as a comprehensive database of mouse, human, plant, and all microbial genomes sequenced to date to elucidate alterations in both host and gut microbial proteins associated with intestinal inflammation. Our results show that host genetics, gutmicrobiota, and inflammation have tremendous influences on the intestinal luminal proteomic landscape.
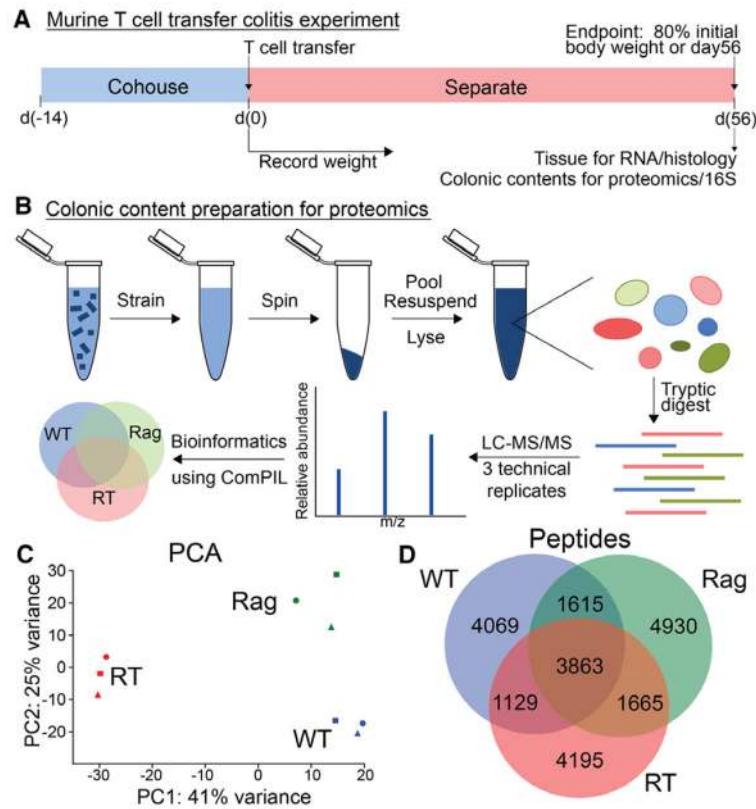
Figure 1.
Global proteomic survey of intestinal contents from a murine colitis model. A) Timeline of the T cell transfer colitis experiment. WT and Rag1$^{-/-}$ mice were initially cohoused for 2 weeks prior to injection of PBS or splenic T cells on day 0. Mice were then separated by treatment group for the duration of the experiment. Mice were euthanized on day 56, at which time intestinal tissue was taken for histology, and colonic contents were saved and prepared for proteomics. B) Schematic of colonic content preparation for proteomics. Colonic contents were prepared as described and the trypsin-digested bacterial proteome was subjected to LC–MS/MS for proteomic data analysis by ComPIL. C) Principal component analysis of LC–MS/MS samples. Different colors correspond to different treatment groups. Different shapes correspond to different technical replicates: circles = replicate 1, triangles = replicate 2, and squares = replicate 3. D) Venn diagram showing differences in peptides identified in each treatment group.
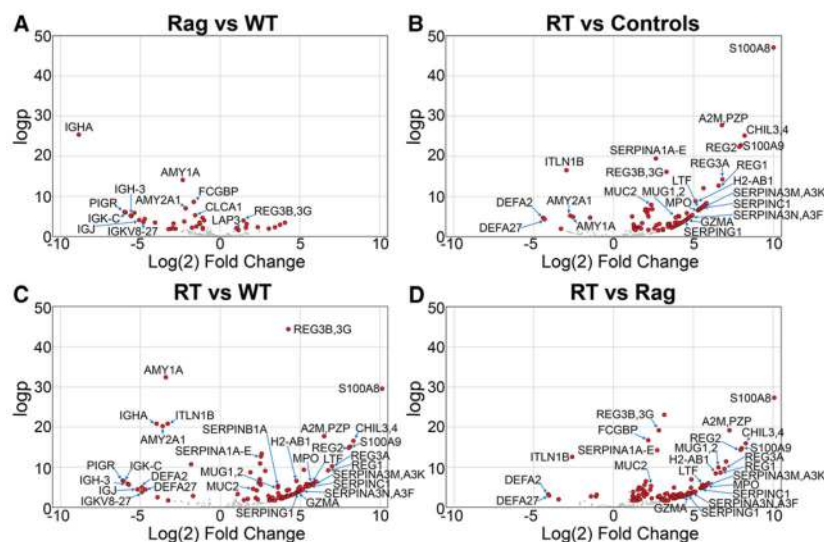
Figure 2.
Significantly altered murine protein clusters identified by LC–MS/MS. Volcano plots of murine protein clusters identified in the proteomic samples. The red dots represent protein clusters that are significantly different (FDR <0.05 (logp) and |Log$_2$FC| >1) in Rag compared to WT samples (A), RT compared to control samples (B), RT compared to WT samples (C), and RT compared to Rag samples (D). Full lists of the significant differentially expressed proteins and GO terms are available Supporting Information, Tables S2, S5, S8, and S11.
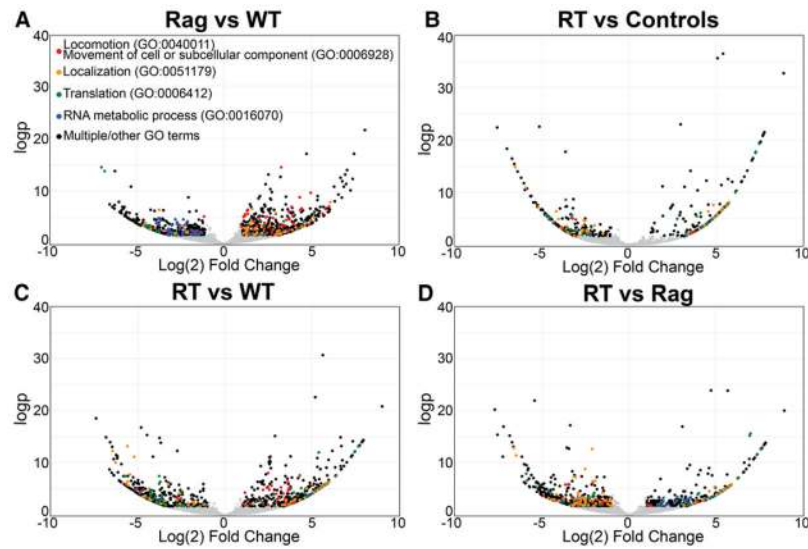
Figure 3.
Significantly altered microbial protein clusters identified by LC–MS/MS. Volcano plots of microbial protein clusters identified in the proteomic samples. The colored dots represent the protein clusters that fall under the parent GO terms (and associated child terms) as labeled in the key in A. The black dots represent protein clusters that are significantly different (FDR <0.05 and |Log_2FC| >1), but do not have assigned GO terms and/or are assigned to multiple/other GO terms. The data compared are as follows: Rag compared to WT samples (A), RT compared to control samples (B), RT compared to WT samples (C), and RT compared to Rag samples (D). Full lists of the significant differentially expressed proteins and GO terms are available Supporting Information, Tables S3, S4, S6, S7, S9, S10, S12, S13.
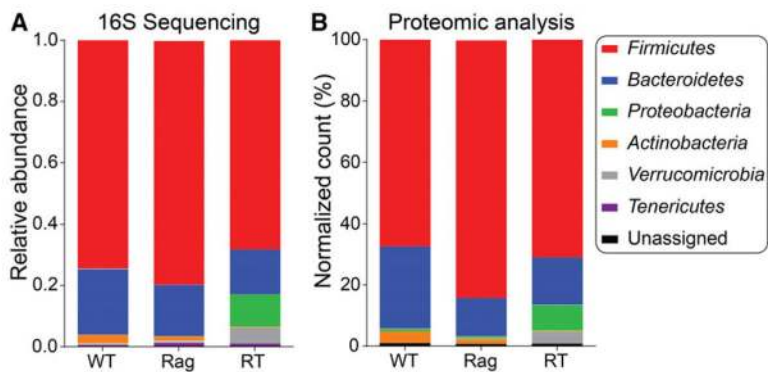
Figure 4.
Taxonomic analyses of colonic bacteria. A) Relative abundance of 16S rDNA sequence assignments of colonic bacterial samples at the phylum level. B) Averaged peptide spectral counts attributable to a single phylum obtained from LC–MS/MS-based metaproteomics.