HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2017 May 23.

Published in final edited form as:

J Proteome Res. 2017 February 03; 16(2): 1014–1026. doi:10.1021/acs.jproteome.6b00938.

Quantitative metaproteomics and activity-based probe enrichment reveals significant alterations in protein expression from a mouse model of inflammatory bowel disease

Michael D. Mayers[†], **Clara Moon**[†], **Gregory S. Stupp**[†], **Andrew I. Su**[†,‡,*], and **Dennis W. Wolan**[†,§,*]

[†]Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550, North Torrey Pines Road, La Jolla, CA 92037, United States

[‡]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550, North Torrey Pines Road, La Jolla, CA 92037, United States

[§]Department of Chemical Physiology, The Scripps Research Institute, 10550, North Torrey Pines Road, La Jolla, CA 92037, United States

Abstract

Tandem mass spectrometry based shotgun proteomics of distal gut microbiomes is exceedingly difficult due to the inherent complexity and taxonomic diversity of the samples. We introduce two new methodologies to improve metaproteomic studies of microbiome samples. These methods include the stable isotope labeling in mammals to permit protein quantitation across two mouse cohorts, as well as the application of activity-based probes to enrich and analyze both host and microbial proteins with specific functionalities. We used these technologies to study the microbiota from the adoptive T cell transfer mouse model of inflammatory bowel disease (IBD) and compare these samples to an isogenic control; thereby, limiting genetic and environmental variables that influence microbiome composition. The data and results generated highlight quantitative alterations in both host and microbial proteins due to intestinal inflammation and corroborates the observed phylogenetic changes in bacteria that accompany IBD in humans and mouse models. The combination of isotope labeling with shotgun proteomics resulted in the total identification of 4434 protein clusters expressed in the microbial proteomic environment, 276 of which demonstrated differential abundance between control and IBD mice. Notably, application of a novel cysteine-reactive probe uncovered several microbial proteases and hydrolases overrepresented in the IBD mice. Implementation of these methods demonstrates that substantial insights into the identity and dysregulation of host and microbial proteins altered in IBD can be accomplished and can be used in the interrogation of other microbiome-related diseases.

[*]Corresponding Authors: DWW (wolan@scripps.edu) and AIS (asu@scripps.edu).

Author Contributions

DWW and AIS conceived of the project. MDM performed wet lab experiments and collected tandem MS data. CM helped generate the IBD mice and advised on sample and data collection. GSS designed and maintained ComPIL. MDM and GSS analyzed tandem LC-MS/MS data. AIS, MDM, and GSS contributed to peptide mapping and functional analysis. All authors contributed to the preparation and editing of the manuscript.

Notes

The authors declare that they have no competing interests.

Keywords

metaproteomics; SILAM; activity-based probes; microbiome; MudPIT; quantitative proteomics; ComPIL; go-term enrichment

INTRODUCTION

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is a powerful technique now being employed by researchers to determine the functional makeup of the highly complex proteomic contents of intestinal microbiota.[1,2] Such studies are providing key information into the kinds of proteins most abundantly expressed by the gut microbiome. For example, non-targeted shotgun metaproteomics on samples prepared from healthy human distal gut microbiota identified many microbial proteins primarily involved housekeeping functions, including translation, carbohydrate metabolism, and energy production.[1] More recently, integration of metagenomics with proteomics was performed to elucidate the phylogenetic alterations and accompanying functional changes in microbial proteins from the gut microbiota of patients suffering from Crohn's disease.[2] These results and others demonstrate that the application of proteomics to the study of highly complex microbial proteomes yields compelling insights into proteins expressed and functional characterization.[3–5]

Despite the increasing application of LC-MS/MS analyses on gut microbiomes, two key limitations in metaproteomics remain. First, rigorous quantitation of microbiome proteomics data obtained through shotgun-based methods has not been performed to date. Differential protein expression has been measured exclusively by spectral counting, but this methodology is only semi-quantitative due to limitations in data-dependent acquisition and low numbers of spectral counts for low-abundance proteins. Furthermore, significant differences in MS1 retention times and intensities for candidate peptides, used to perform spectral-based quantitation, are typically observed across samples and are due to variability in liquid chromatography.[6–8] Metabolic labeling techniques help to overcome the many issues associated with spectral-based quantitation and limit systematic errors in sample preparation and LC-MS/MS proteomic data collection.

Quantitative proteomics between two biological states is readily accomplished in tissue-cultured cells and whole animals via isotope labeling.[9] For example, stable isotopic labeling in mammals (SILAM) is a reliable method to accomplish measuring and quantitating proteomic differences in both mice and rats, whereby animals are restricted to a diet of isotopically ^{15}N-labeled spirulina as the only source of nitrogen.[10–14] Importantly, SILAM has shown utility in labeling and quantifying the whole organism, including long-lived proteins of the brain.[15] Despite these previous validations of SILAM, the isotope incorporation efficiency and quantification capabilities of microbiome constituents has yet to be determined. We wanted to build on the SILAM methodology and measure isotope incorporation in fecal samples of animals on a ^{15}N diet by metaproteomics. Our primary goal was to determine if limiting environmental and genetic variables associated with gut

microbial diversity would increase the likelihood of the identification and quantitation of differentially expressed proteins between isogenic control and diseased murine cohorts.

The second limitation in metaproteomics profiling is the high degree of complexity in microbiome samples. Recent proteomics-based advances have attempted to address the issue of the microbiome complexity by improving methods in sample preparation[16,17] database search strategies and algorithms,[18,19] and implementation of proteogenomics.[20,21] Despite these extensive improvements to metaproteomics, the sensitivity of LC-MS/MS instrumentation constrains data collection to a small percentage of the most abundant peptides in a proteomic sample. We and others have begun to apply pre-fractionation "enrichment" steps to circumvent instrument sensitivity issues and access low-abundant biologically important proteins within a complex sample.

One of the most efficient methods of "enrichment" includes the application of activity-based probes (ABP) that target specific protein families. Addition of small-molecule ABPs permits systematic quantitation of individual classes of proteins potentially lower in abundance than the detectable limit and therefore missed by LC-MS/MS of whole microbiome-derived proteomes. Such chemical probes have already been designed to target more than a dozen protein classes (*i.e.,* hydrolases, proteases, kinases, phosphatases, and glycosidases[22]) and have been successfully employed in identification of dysregulated proteins in cancerous tumors[23], parasitic infections[24], and fatty livers[25]. To address the limitations in MS-based identification of low-abundance proteins in highly complex and concentrated proteomic samples, we applied an ABP that covalently labels proteins with nucleophilic cysteine residues.

Here, we combine these two strategies, SILAM-based quantitation and application of ABP probes, to address these two key bottlenecks in metaproteomic analysis. We applied this integrated system to the study of a mouse model of IBD. Using this model, we could identify and quantitate differences in host and microbiome protein functionalities between mice with intestinal inflammation and isogenic controls fed identical diets. Our data show that both host and microbiome proteins can be identified and quantitated between two cohorts. These results also demonstrate new techniques that can be extrapolated to the metaproteomic interrogation of other animal model and human microbiomes. These findings provide a deeper perspective of the microbiome proteome and of the proteins that are altered in expression between the control and IBD mouse groups.

MATERIALS AND METHODS

Murine adoptive T cell transfer chronic IBD model and fecal sample collection

Animal protocols were approved by The Institutional Animal Care and Use Committee (IACUC) at The Scripps Research Institute (TSRI). All mice were purchased from Jackson Laboratories and maintained in a specific pathogen-free barrier facility at TSRI for the duration of the study.

We employed the well-established T cell transfer model of colitis to induce intestinal inflammation, as this method promoted rapid and reproducible intestinal inflammatory

pathophysiology within an approximate 8-week timeframe[26]. Twenty 10-week old female B6.129S7-*Rag1*^{tm1mom}/*J* (Rag1$^{-/-}$) mice were purchased from Jackson Laboratories (stock #002216), separated into 4 cages (5 per cage), and co-housed for one week to normalize the microbiota. $CD3^+CD4^+CD8^-CD25^-Foxp3^-$ naïve T cells collected via fluorescence-activated cell sorting from the spleens of donor *Foxp3-EGFP* reporter mice (Jackson Laboratories, stock #006772) were transferred retro-orbitally to 10 *Rag1*$^{-/-}$ mice (approximately $5x10^5$ T cells per mouse) (hereafter referred to as "IBD mice"). The control *Rag1*$^{-/-}$ mice (hereafter referred to as "RAG$^{-/-}$ mice") were injected retro-orbitally with a similar volume of sterile PBS and subsequently separated for the remainder of the time course from the IBD cohort. Fourteen days post T cell transfer, the mice were switched from standard chow (Harlan Teklad) to a spirulina-based diet, with the RAG$^{-/-}$ mice receiving ^{15}N labeled spirulina and the IBD mice receiving a ^{14}N control spirulina diet (Cambridge Isotope Laboratories) (Figure 1A). The "heavy" ^{15}N-labeled diet has no effect on food consumption, weight gain, normal physiology, reproduction, or development.[10,11,13,27] Mouse fecal samples were collected daily, pooled by cage and week, and immediately frozen at −80 °C for further use (Supporting Information, Table S1). Five IBD mice were sacrificed prior to the day 56 endpoint due to severe signs of disease and a decrease in 20% body weight with respect to day 0 weight (Supporting Information, Figure S1). The remaining 5 IBD mice exhibited signs of disease, including reduced weight gain (Supporting Information, Figure S1), diminished grooming habits, and hunched posture. All IBD mice were confirmed to have intestinal inflammation by histology (Supporting Information, Figure S2) and were included in the experiment.

Fecal microbiome sample preparation for LC-MS/MS analysis

Pooled fecal samples were thawed in 3 mL of cold PBS and vigorously vortexed to break up fecal pellets. Suspended samples were subjected to low-speed centrifugation (100xg, 2 min, 4 °C) to separate insoluble fecal matter. The fecal supernatant containing suspended bacterial cells was aliquoted into 1 mL fractions. Fractions were pelleted (6,500xg, 15 min, 4 °C), washed twice with cold PBS, and suspended into 450 μL of lysis buffer consisting of 0.1% Triton-X 100 in PBS and Roche cOmplete protease inhibitor cocktail. Samples were lysed via sonication in a Qsonica Q700 sonicator with Cup Horn attachment at 4 °C for 15 min. Insoluble cellular material was removed via centrifugation (10,000xg, 5 min, 4 °C) and the remaining soluble protein concentration was measured (Pierce BCA Protein Assay Kit). Prepared soluble microbiome samples from RAG$^{-/-}$ control and IBD mice were mixed 1:1 in aliquots containing 50 μg each ^{15}N and ^{14}N labeled proteome (Figure 1B). The proteomic content was precipitated with 30% w/v of trichloroacetic acid (TCA) and stored at 4 °C overnight. The microbiome proteome was pelleted (21,000xg, 20 min, 4 °C) and washed twice with −80 °C acetone prior to trypsin digestion to generate peptides for LC-MS/MS data collection (Figure 1B).

Fecal microbiome preparation with BioGlyCMK probe labeling for "enriched" samples

50 mg of RAG$^{-/-}$ (^{15}N) and IBD (^{14}N) mouse bacterial pellets derived from fecal matter (isolated and washed as described above) were resuspended in PBS and mixed in a 1:1 ratio in a total volume of 500 μL. The combined samples were subjected to labeling with 100 μM of a biotinylated glycine-chloromethyl ketone (BioGlyCMK) probe (1% DMSO for

unlabeled controls) and incubated overnight at 4 °C under light agitation (see the Supporting Information for BioGlyCMK synthesis). Bacteria were pelleted (6,500xg, 15 min, 4 °C), washed 2x with 1 mL PBS to remove unreacted probe, and resuspended in 450 μL of PBS with 0.1% SDS. Samples were lysed via sonication in a Qsonica Q700 sonicator with Cup Horn attachment at 4 °C for 15 min. All lysed microbial samples were assessed for BioGlyCMK labeling by streptavidin blot and concentrations determined with colorimetric BCA assays. Proteins were denatured via dilution with 500 μL of 2% SDS in PBS followed by heating at 95 °C for 15 min. Denatured samples were diluted with 4 mL PBS (0.2% SDS) and incubated with 100 μL of high capacity streptavidin agarose beads (Pierce) overnight at room temperature. Beads were pelleted (500xg, 2 min) and subjected to extensive washing with 0.2% SDS in PBS (1x), PBS (3x), and ultrapure water (3x) prior to trypsin digestion.

Microbiome protein trypsin digestion

Both “unenriched” and “enriched” BioGlyCMK probe-labeled washed lysate pellets were subjected to trypsin digestion to generate peptides for MudPIT shotgun proteomics analysis.[28] Proteins were resuspended and denatured in 60 μL of 8 M urea, 100 mM TrisHCl, pH 8.0, introduced to 1 μL of 300 mM TCEP to reduce all disulfide bonds, and agitated for 20 min at 25 °C. The reduced thiols were then alkylated via addition of 7 μL 500 mM 2-chloroacetamide and incubated with gentle agitation for 15 min at 25 °C while protected from light. Following alkylation, samples were diluted with 180 μL of 100 mM TrisHCl, pH 8.0 to reduce the urea concentration to less than 2 M. General proteomic digestion was performed by addition of 2.5 μg of trypsin to each sample in the presence of 1 mM $CaCl_2$ and incubation overnight at 37 °C. “Unenriched” samples were quenched with 13 μL of formic acid, centrifuged (21,000xg, 20 min) and the supernatant was stored at −20 °C until LC-MS/MS analysis. BioGlyCMK-enriched samples were filtered via spin column (100xg, 1 min) (Pierce) and the beads were washed twice with 50 μL ultrapure H_2O. All filtrates were combined and acidified with 17 μL of formic acid, centrifuged, and the supernatant was stored at −20°C until LC-MS/MS analysis.

Preparation of MudPIT LC column

Trypsin-digested peptides were loaded onto a biphasic MudPIT column (250 μm fused silica (Agilent), paced with 2.5 cm of 5 μm Aqua C18 resin followed by 2.5 cm of Partisphere strong cation exchange resin (SCX)). An analytical column was prepared from 100 μm fused silica pulled to a 5 μm tip by a micropipette puller (Sutter Instrument Company, Model P-2000). This column was then pressure loaded with 12 cm of 3 μm Aqua C18 resin.

LC-MS/MS MudPIT data collection

Standard MudPIT tandem mass spectrometry was performed using a Thermo LTQ-Orbitrap XL mass spectrometer. The sample and analytical columns were joined by a zero dead volume union (Waters). Peptides were eluted at 300 nL/min using an 11-step MudPIT program. Each step began with 1 minute of 100% Buffer A (95% H_2O, 5% acetonitrile, 0.1% formic acid), a 4-min salt pulse with *x*% buffer C (500 mM ammonium acetate, 95% H_2O, 5% acetonitrile, 0.1% formic acid), then 5 min 100% buffer A, followed by a 105 min gradient from 5–65% buffer B (20% H_2O, 80% acetonitrile, 0.1% formic acid) and finally 5 min of 100% buffer A. The 4-minute buffer C salt pulses (*x*) were as follows: 10%, 15%,

20%, 25%, 30%, 40%, 50%, 60%, 80%, 100% with the final pulse consisting of 90% buffer C and 10% buffer B. 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 8 peaks were selected for fragmentation using HCD with normalized collision energy set to 35.0. Dynamic exclusion was enabled with exclusion duration set to 60.0 seconds.

Peptide identification with 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 Comprehensive Protein Identification Library (ComPIL) database, containing over 80 million proteins from multiple microbial database sources as well as human, mouse, and plant proteins.[18] 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), and 3) a precursor mass tolerance of 10 ppm and 50 ppm tolerance for fragmentation ions. Filtering was performed using DTASelect 2.1.3 (http://fields.scripps.edu/yates/wp/?page_id=17), requiring 2 peptides per protein and a false discovery rate (FDR) of 1% with respect to proteins. The following parameters were used for filtering when run from the command line: “--quiet --brief --trypstat --modstat -y 2 -DM 10 --extra --dm --pfp 0.01 -p 2”. Samples containing a mixture of ^{14}N and ^{15}N peptides for ratio quantification were searched against a ComPIL database consisting of ^{14}N and then ^{15}N-labeled peptides. DTASelect filtering was performed on each search individually as well as the combined results, producing 3 outputs for several downstream bioinformatic analyses.

Peptide and protein quantification

The $^{14}N/^{15}N$ ratio of each peptide was quantified using the program Census (available on the Integrated Proteomics Pipeline (IP2), http://goldfish.scripps.edu/). Census uses the results of DTASelect filtering along with the extracted MS1 spectral data to both determine the isotopic enrichment within a sample and to calculate the $^{14}N/^{15}N$ isotopic ratio for each identified peptide within one LC-MS/MS data collection experiment. Enrichment calculations were performed on filtered results searched only against the ^{14}N database as to not bias the results toward ^{15}N peptides.[29] Census quantification on the combined $^{14}N/^{15}N$ filtering results file was performed with the default settings for ^{15}N ratio quantification in high-resolution, including a 15 ppm tolerance for isotope extraction. Census filtering parameters were modified to disable iterative outlier analysis and exclude outliers with a p-value >0.05. The median of the peptide ratios was centered to 1 via division by the inverse log of the median of the log-transformed ratios. Protein loci ratios were determined by computing a weighted ratio of the associated peptides, weighted by the regression factor (r), a value determined by Census to describe the confidence in ratio quantification.

LC-MS/MS data analysis

Code from the Microbiome Metaproteomics package (available at https://bitbucket.org/sulab/metaproteomics) was modified to incorporate DTASelect results for the individual

filtering of the ^{14}N and ^{15}N database searches, as well as the combined filter result and Census ratio quantification into each sample. The source code for this process, as well as all analyses is available in Python 3.5.2 (https://github.com/mmayers12/n15_mice/). Protein clustering, cluster taxonomy, and gene ontology (GO) term annotations were stored within ComPIL and assigned as previously described.[18] Samples were grouped together by type to determine the average fold change for each protein locus. Significance values were determined by a Student's t-test, adjusted by the Benjamini-Hochberg correction.[30]

GO term enrichment analysis

For determination of GO term enrichment between the sample preparation conditions, BioGlyCMK-enriched and "unenriched," a Fisher's Exact Test was performed. This was dependent on the number of annotation occurrences of a given term in the set of protein clusters (*i.e.,* proteins with 70% sequence identity and common functionality),[18] unified across experimental replicates for the sample preparation condition. To compare GO terms enriched in the microbial proteins between different biological conditions, IBD and $RAG^{-/-}$, gseapy 0.7.0 (https://github.com/BioNinja/gseapy), a Python implementation of the Broad Institute's Gene Set Enrichment Analysis (GSEA) algorithm was used.[31] GO gene sets were generated from all identified protein clusters in a given sample type (BioGlyCMK-enriched or "unenriched"). 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 such that only 1 sibling remained.

Taxonomy analysis

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. Peptide taxonomy search space was restricted to the proteins identifiable in a given sample. Analysis was performed at the phylum level. Each peptide was traced back to a phylum and if uniquely classifiable, the peptide was classified with a weight of normalized counts. Peptides without a discernable phylum (*e.g.,* could belong to more than one) were discarded from analysis. The normalized counts were then used to determine an approximate fractional taxonomic makeup of the sample.

RESULTS AND DISCUSSION

Isotope-labeled SILAM mice rapidly incorporate ^{15}N into their microbiomes

We first established the rate and extent of ^{15}N isotope incorporation within the microbiomes of both $RAG^{-/-}$ and IBD mice. Mice were placed on an isotopically enriched spirulina-based chow 14 days post T cell transfer, whereby the $RAG^{-/-}$ control group (10 mice) received the ^{15}N isotopically labeled spirulina and the IBD group (10 mice) were fed the corresponding unlabeled (natural abundance) ^{14}N spirulina.[10,11] LC-MS/MS MudPIT data collection and quantitative analysis were performed on the unenriched fecal bacterial samples within the first 24 hr on the spirulina-based chow. Data collection and analysis of

the 24-hr time point demonstrated that the murine microbiomes incorporated a relatively high level of ^{15}N with a median peptide isotope enrichment of approximately 84% (Figure 1C). However, this value is an estimate, as previous studies suggested a minimum of 500 quantifiable peptides for accurate assessment of enrichment. We identified 181 and 80 quantifiable peptides from the ^{14}N IBD and ^{15}N $RAG^{-/-}$ cohorts, respectively. Despite the initial rapid rate of ^{15}N incorporation, the rate slowed over the first week with 86% enrichment four days after the dietary change and 88% at the end of 7 days, as based on 569 and 923 peptides, respectively. Importantly, the acceptable level of enrichment required for quantification of 95% was attained within 4 weeks on the ^{15}N spirulina diet. By the day 56 endpoint (or 6 weeks on the isotopically labeled spirulina chow), the mice reached 96% enrichment.

Previous metaproteomic-based microbiome studies have detected the presence of dietary proteins in fecal samples; however, these diet-related proteins are usually found in low abundance.[3] For our experiment, the source of dietary protein originates entirely from a single prokaryotic organism (*e.g.,* spirulina) and we needed to verify that dietary proteins would remain in low abundance in the fecal samples. As such, our initial concern was that dietary proteins may be in such abundance as to dominate LC-MS/MS collection and significantly limit detection of any host- and microbiome-derived peptides. Mass spectrometry data obtained from the ^{14}N IBD mice demonstrated these fecal samples to have a relatively low number of spirulina proteins with respect to all host and bacterial proteins (Supporting Information, Figure S3). The attributable presence of spirulina proteins in the IBD mice samples reached a 1.3% maximum of all identifiable proteins. Conversely, the abundance of spirulina proteins as a fraction of spectral counts in the ^{15}N labeled $RAG^{-/-}$ control samples were significantly more prominent than the IBD group. Day 1 samples reached as high as 72% of all LC-MS/MS measured spectral counts, and reduced to approximately 20% throughout the remainder of the experiment (Figure 1E). Despite the high spectral count signals derived from spirulina proteins, the fraction of identifiable protein loci originating from spirulina samples remained fairly low, with the day 1 sample containing 18% spirulina loci and the remainder of collected fecal samples between 3 and 7% with respect to all host and microbial proteins (Figure 1D). We expected spirulina protein signals to dissipate to levels demonstrated for other dietary experiments allowing for detectable levels of microbial proteins. Despite these persistent proteins from the SILAM spirulina diet, we were able to quantitate many host and microbial proteins that are differentially present between control and IBD mice.

IBD mice overexpress host protease inhibitors and inflammatory proteins

We focused metaproteomic data collection and analysis on week-8 endpoint samples when inflammation was most severe, and the differences in protein expression between the $RAG^{-/-}$ control and IBD mice would be greatest. Samples were pooled and prepared for experimental replicates as described Supporting Information, Table S1. After MudPIT LC-MS/MS and data analysis with ComPIL, the average replicate contained 1989±180 peptides with 1040±136 corresponding matched proteins, and 1424±86 peptides matching 800±62 proteins from the IBD and $RAG^{-/-}$ samples, respectively. This amounted to 3277 protein clusters identified across all replicates and conditions (Supporting Information, Table S2).

To assess reproducibility among experimental replicates, samples were visualized by a hierarchical clustering dendrogram based on the Jaccard distance calculated from presence or absence of peptides in a sample (Supporting Information, Figure S4). The samples clustered well with some evident batch effects. Close clustering was observed among the $RAG^{-/-}$ control experimental replicates as well as for the IBD mouse samples suggesting differences in the proteomic makeup between control and IBD mice (Supporting Information, Figure S4).

Despite the conservation of the overall number of detected peptides in the MS data, quantitative differences in peptide abundance were evident between the control and IBD mice. Utilizing the protein ratio values from Census and applying a student's t-test combined with a Benjamini–Hochberg adjustment to correct for multiple testing, we identified 89 proteins significantly more abundant in IBD mice and 112 proteins significantly reduced in IBD in comparison to the $RAG^{-/-}$ control group (Figures 2A, B). Of the proteins with greater abundance in the IBD mice, 33 were murine, including Serpinc1, Mug1, Pzp, and Serpina3c. These proteins are serine protease inhibitors previously established to be present in greater quantities during gut inflammation.[32–35] Significant increases were also observed for host-derived c-type lectins, such as Reg1 (lithostathine) and Reg2 (lithostathine 2) that are involved in the proliferation and differentiation of various cell types and upregulated in diabetes and gastrointestinal cancers.[36,37] Finally, calcium-binding proteins S100A8 and S100A9 that form calprotectin, an ion sequestering antimicrobial protein produced by neutrophils, were tremendously increased in colitis and represent known host inflammatory markers (adjusted p-value 0.0005, Figure 2A).[38,39]

The majority of microbial peptides measured to be overabundant in the IBD mice were derived from bacterial proteins that are required for general metabolic processes and likely represent some of the most abundant proteins expressed by bacteria. Such proteins included glyceraldehyde-3-phosphate dehydrogenases (GAPDH), outer membrane transporter proteins, triosephosphate isomerases, pyruvate kinases, and ribosomal proteins primarily from *Akkermansia sp.* and *Lactobacillus sp*. Notably, several proteins have no known annotated function and/or the protein cluster may only include peptide matches to genomic sequences lacking annotated taxa (Supporting Information, Table S3).

Our increased detection of *Akkermansia sp.* and *Lactobacillus sp.* proteins in IBD microbiota samples correlates with several previous studies on humans and mice. Unfortunately, there are conflicting reports as to the prominence and importance of both genuses with respect to microbiomes of healthy and inflamed states. *Akkermansia muciniphila* has been determined to decrease in abundance in IBD patients, as assessed by 16S rRNA sequencing;[40] however, the bacterium was demonstrated to promote gut inflammation in mice infected with *Salmonella enterica*.[41] Similarly, despite the use of members of the *Lactobacillus sp.* as probiotics to fight intestinal inflammation, increases in the genus as well as *Bifidobacterium* were measured by 16S rRNA gene rtPCR on the biopsies of Crohn's disease patients with active inflammation.[42] Despite the inconsistencies among reports, alterations in the microbiome phylogeny accompany the changes in the epithelial and mucosal lining of the distal gut during inflammation. Those bacteria that degrade mucus as a food source, such as *Akkermansia sp.* may have improved ability to

survive over food-dependent luminal bacteria in stressful gastrointestinal conditions.[43,44] Our results demonstrate that application of quantitative metaproteomics to microbiome studies will complement sequencing efforts and help shed light on the phylogenetic alterations associated with microbiome-associated intestinal diseases.

Proteins increased in control mice are primarily from microbes

The majority of significantly increased proteins in the $RAG^{-/-}$ mice (or depleted in the IBD mice) are of microbial origin. We observed several enriched peptides that correspond to housekeeping proteins from *Firmicutes*, such as lipases, transferases, flagellar proteins, and acyl transferases (Supporting Information, Table S4). Surprisingly, the majority of the peptides identified to be more prevalent in the $RAG^{-/-}$ originate from proteins with unknown functions (Figure 2B, Supporting Information, Table S4). One genus with an overrepresentation of proteins in $RAG^{-/-}$ includes the *Arthrospira sp.* and likely represents an artefactual enrichment due to peptides from this source having a higher level of ^{15}N incorporation than those from mouse and microbiome sources (Figure 2B, Supporting Information, Table S4). Our proteomic data also identified host proteins decreased in diseased mice, including murine pentraxin (Mptx1), intelectin (Itln1), and alpha defensin (Defa7) (Figure 2A). Importantly, these proteins have been previously identified to be altered in gut inflammation, and further validates the utility of our MS-based microbiome interrogation. Murine pentraxin was previously found to have diminished expression during intestinal oxidative stress.[45–47] Similarly, the glycolipid barrier protection protein Itln1 and the antimicrobial peptide Defa7 are secreted by specialized intestinal goblet and Paneth cells and have been identified to be downregulated in inflammatory states.[48–51]

GO term analysis shows alterations in microbial protein functionalities of IBD mice

Our LC-MS/MS data collection and ComPIL analysis of the SILAM samples identified a total of 2893 unique microbiome protein clusters among both groups of mice that were not produced by the host or spirulina diet. Out of a total of 3277 protein clusters in our SILAM datasets (contributed by host, diet, microbes), 201 were found to significantly change between healthy and disease samples (Figures 2A, B). From the identified clusters, 83.6% of the 3277 clusters have at least one GO term annotation (*i.e.,* molecular function, biological process, cellular component) in ComPIL (Supporting Information, Tables S5–S7). To take advantage of this high level of annotation coverage, we applied a Gene Set Enrichment Analysis (GSEA) algorithm to uncover any statistically significant functional alterations between the control and IBD mice. In terms of the annotated molecular function GO terms, the upregulated microbial proteins in IBD mice were dominated by oxidoreductase and lyase activities (Supporting Information, Table S5). Those GO term functions that describe biological processes revealed an upregulation in IBD mice of many metabolic processes, including carbohydrate catabolic processes (Supporting Information, Table S6). Conversely, our metaproteomic results suggested that $RAG^{-/-}$ controls have excess housekeeping functionalities, including RNA Pol activity and many biosynthetic processes (Supporting Information, Tables S5, S6). While the $RAG^{-/-}$-associated functions are indicative of normal metabolic pathways, the depletion of these microbial proteins in IBD mice may suggest a limited abundance of resources for biosynthesis.

Differences in taxonomic composition between control and IBD mice

Aside from the identification and quantitation of microbial proteins, we wanted to evaluate the ability of our metaproteomics data to estimate the microbial composition of the samples analyzed. All peptides from the SILAM datasets were traced to the lowest common bacterial ancestor from which they were uniquely derived and used to generate insights into the phylogenetic composition of the control and IBD microbiome samples. The approximate bacterial composition was determined by weighting all peptides by normalized spectral counts. With respect to phylum-level composition, our metaproteomics data are in strong agreement with published 16S sequencing whereby *Firmicutes* and *Bacteroidetes* dominate the overall population of the microbial content and *Bacteroidetes* decreases in an inflammatory state (Figures 3A, B).[52–54] Importantly, our data demonstrated a statically significant increase in *Proteobacteria* and *Verrucomicrobia* (*e.g., Akkermansia sp.*) in the IBD mice and these results are strongly correlated with the phylogenetic composition as assessed by metagenomics sequencing (Figures 3C, D). The overabundance of both phyla have been observed in humans with ulcerative colitis as well as Crohn's disease. While *Proteobacteria* has more commonly been observed to bloom in IBD,[55–59] several publications have demonstrated *Verrucomicrobia* also increase in number.[41,60,61] The correlation of genomic sequencing and our proteomics data suggest that our microbiome sample preparation, LC-MS/MS data collection, and analytic methods are not biased towards any particular microbial components with respect to phylum level and can be used to corroborate sequencing results.

Labeling of reactive cysteines allows for interrogation of a new subset of proteins

Our results verified the overabundance/expression of host anti-proteolytic proteins already observed in IBD mice. However, we did not detect any host and/or microbial proteases also previously established to accompany intestinal inflammation, as measured by proteolytic activity found in fecal content.[62–64] The extreme complexity of the proteome derived from intestinal contents in combination with current limitations in tandem mass spectrometer sensitivity ensures that a significant number of host and microbial proteins will be missed with MudPIT shotgun metaproteomics. We therefore wanted to determine if an additional "enrichment" step that would target a subset of proteins, including proteases, from within our fecal samples would provide further insight into the presence of host and microbial proteases. To accomplish this, we applied an additional ABP enrichment step to the SILAM control and IBD fecal samples in order to isolate proteins with cysteine nucleophilic reactivity. Using an activity-based probe enrichment step to label specific functionalities of interest may be able to reduce the complexity of the proteomic environment while simultaneously highlighting new differences between two biological groups. Here, our goal was to determine if the enrichment process resulted in the LC-MS/MS identification of a different set of host and microbial proteins in comparison to the unenriched datasets, and if any biologically relevant information on aberrant protease functionalities can be quantitated between the control and IBD mice.

Based on the successes of previous probe-based research in identifying proteases from the lysates of animal tissues, we synthesized a biotinylated glycine containing a C-terminal chloromethyl ketone warhead to generate a general cysteine-reactive molecule termed

BioGlyCMK (Figure 4A, Supporting Information for synthesis). Attack by nucleophilic cysteine residues on BioGlyCMK results in irreversible biotinylation of proteins that then permits the probe-reactive subset of proteins to be enriched and isolated from the general proteome by avidin-coated beads. Here, our fecal sample preparation was modified to accommodate the additional probe incubation step (Figure 4A). Our proof-of-principle study with labeling introduced the probe after the 1:1 mix of $^{14}N/^{15}N$ SILAM fecal samples in order to equally enrich the samples with the BioGlyCMK probe.

Enrichment with BioGlyCMK resulted in identification of an average of 1837±293 peptides corresponding to 1088±160 proteins, and 968±128 peptides with 895±114 protein matches for the ^{14}N IBD and ^{15}N control samples, respectively. Despite the overall number of measured peptides and proteins matching the unenriched LC-MS/MS data, only 51% of the identified protein loci were common to both the BioGlyCMK-enriched and unenriched SILAM datasets (Figure 4B). Importantly, 49% of the identified protein loci were unique to the BioGlyCMK data, supporting our hypothesis that probe-based enrichment isolates a unique subset of proteins in comparison to the unenriched datasets (Figure 4B, Supporting Information, Table S8). Microbial proteins with reactive active-site cysteines specifically identified in the BioGlyCMK-enriched data include, but are not limited to, protein clusters with GO terms related to peptidases, alcohol dehydrogenases, and acetaldehyde dehydrogenases (Table 1). These findings suggest that use of our probe reduced the presence of housekeeping functionalities lacking a reactive cysteine (Table 1, Supporting Information, Table S8). Of note, a Fisher's Exact test used to compare the overall representation of GO terms annotated in the unenriched versus BioGlyCMK enriched samples demonstrated that the most strongly enriched molecular function GO Term was cysteine-type peptidase activity (GO:0008234) with an odds ratio of 14.9 (p-value = $3.0 \cdot 10^{-9}$; Table 1). This supports the idea that an activity-based probe system can be used to further interrogate the complexity of the distal gut microbiome proteome.

Use of an ABP likely reduces the complexity of the microbiota sample in comparison to unenriched sample preparation methods, as the probe will preferentially isolate only those proteins with selective reactivity. Here, the number of identified proteins is larger in the BioGlyCMK dataset than for the collection of proteins found in the unenriched dataset. These results suggest that the general promiscuity of our BioGlyCMK probe for proteins with nucleophilic cysteine residues results in enriched samples that are still greater in diversity and complexity than the LC-MS/MS detection capabilities. Preparation of the samples in an aerobic environment may result in a fraction of the nucleophilic cysteine-containing proteins to oxidize and limit the ability to bind BioGlyCMK. Use of ABPs with more specific spectrums of target functionalities will likely generate enriched datasets that fall within the detection limit of tandem mass spectrometers and would further improve the reproducibility of both biological and technical proteomic data replicates.

Application of BioGlyCMK probe highlights additional differences in control and IBD mice

We focused our analysis to quantitate differences among the BioGlyCMK-enriched $RAG^{-/-}$ control and IBD proteomic datasets to those peptides and proteins uniquely identified in comparison to the unenriched data sets. Most significant is the overall reduction in

ubiquitous and highly conserved proteins, including ribonuclease activity, RNA polymerase activity, and DNA binding proteins. GO term analysis with the GSEA algorithm shows a statistically significant increase in peptides from microbial peptidases and hydrolases from the IBD microbiome samples, and proteolysis ranked as the most highly enriched biological process (Table 2). While our probe only targeted the subset of cysteine-based proteases, it is clear from these results as well as the unenriched data set where host-produced protease inhibitors are tremendously increased, that proteolytic activity is a critical component of the IBD mouse model. Use of probes that target serine, metallo-, and aspartate proteases will likely provide additional insights and help generate a compendium of potential microbially produced proteins that can be further assessed for their importance in propagating IBD.

Isotopic labeling with ^{15}N increases quantification reliability in enriched samples

Although isotopic labelling is considered to produce more reliable relative quantification in proteomic studies, many label-free methods of quantification, including those that rely heavily on spectral counts, have been developed and are currently in use.[2,17] To attempt to address this question and the utility of isotopic labeling in samples as complex as the microbiome, we compared the isotopic ratios generated by precursor ions, to the ratio of the normalized spectral counts between a ^{14}N and ^{15}N sample of a given mass spectrum. Examining the relationship between the precursor ion intensity ratio with that of the spectral counts for each identified peptide yielded a moderately strong correlation of 0.67 (Figure 4C). While there are many peptides that are discordant between these two measures, the accuracy of the two cannot readily be determined. One area where ^{15}N isotopic labeling may confer some advantage over spectral count quantification is in peptides of low abundance, that have 5 or fewer counts among the ^{14}N and ^{15}N samples. In the unenriched dataset, 578 peptides fall under this category, accounting for 9% of all peptides identified. Examining the MS data collection variance between replicates for these low abundance peptides by either SILAM ratios or relative spectral counts yielded almost identical distributions for the unenriched SILAM data (Figure 4D). Notwithstanding, performing this analysis on the BioGlyCMK-enriched samples shows much tighter variance when quantifying ratios via the ^{15}N method (Figure 4E). Our analyses demonstrate that isotopic ^{15}N labeling may confer a quantitative advantage over spectral count-based methods in metaproteomic studies when used in combination with an ABP enrichment step. We find that SILAM isotope quantitation on the highly complex unenriched microbial samples produces a similar level of variance to use of spectral count-based quantitation and use of SILAM should be considered as a potential approach for future metaproteomic studies.

CONCLUSIONS

Our study introduced SILAM and ABP enrichment to identify and quantitate differences in the highly complex protein mixture of host and microbial proteomes of RAG$^{-/-}$ control and IBD mice. MudPIT shotgun proteomics on SILAM isotopic labeled murine fecal samples in combination with our previously described ComPIL database permitted the identification of 4434 protein clusters in the microbial proteomic environment. Of these clusters, 276 were found to be in differential abundance between control and IBD mice, many of which are microbial proteins of unknown function. In addition, incorporation of an ABP enrichment

step into the sample preparation process allowed for a unique subset of the microbial proteins to be identified by LC-MS/MS relative to the unenriched sample collection. Together, SILAM and our cysteine-reactive BioGlyCMK ABP identified several peptidases and hydrolases to be overly abundant in IBD. Addition of these methodologies to metaproteomics of human and disease model systems will help to identify and measure microbial proteins dysregulated in disease and begin to create a novel list of drug discovery targets to combat microbiome-related intestinal diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources

This work was supported by The Scripps Research Institute and the US National Institutes of Health 1R21CA181027 (to DWW) and U54GM114833 (to AIS) and NIH Training Grant T32AI007244 (to CM).

We thank D. McClatchy for consultation on SILAM techniques; A. Wang for assistance with mass spectrometry and probe labeling protocols; S. Chatterjee for assistance and knowledge relating to ComPIL; O. Kirak for assistance with T cell transfer and mouse care; A. Solania for BioGlyCMK probe synthesis; R. Park for maintenance of Blazmass; and J. Moresco, J. Diedrich, and J. Yates for technical assistance with sample preparation and mass spectrometry instrumentation.

References

1. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, et al. Shotgun metaproteomics of the human distal gut microbiota. ISME J. 2008; 3(2):179–189. [PubMed: 18971961]
2. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, Shah M, Halfvarson J, Tysk C, Henrissat B, et al. Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. PLoS ONE. 2012; 7(11):e49138. [PubMed: 23209564]
3. Kolmeder CA, de Been M, Nikkilä J, Ritamo I, Mättö J, Valmu L, Salojärvi J, Palva A, Salonen A, de Vos WM. Comparative Metaproteomics and Diversity Analysis of Human Intestinal Microbiota Testifies for Its Temporal Stability and Expression of Core Functions. PLOS ONE. 2012; 7(1):e29913. [PubMed: 22279554]
4. Kolmeder CA, de Vos WM. Metaproteomics of our microbiome — Developing insight in function and activity in man and model systems. J Proteomics. 2014; 97:3–16. [PubMed: 23707234]
5. Xiong W, Abraham PE, Li Z, Pan C, Hettich RL. Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. PROTEOMICS. 2015; 15(20): 3424–3438. [PubMed: 25914197]
6. Prakash A, Mallick P, Whiteaker J, Zhang H, Paulovich A, Flory M, Lee H, Aebersold R, Schwikowski B. Signal Maps for Mass Spectrometry-based Comparative Proteomics. Mol Cell Proteomics. 2006; 5(3):423–432. [PubMed: 16269421]
7. Fischer B, Grossmann J, Roth V, Gruissem W, Baginsky S, Buhmann JM. Semi-supervised LC/MS alignment for differential proteomics. Bioinformatics. 2006; 22(14):e132–e140. [PubMed: 16873463]
8. Lee S, Kwon MS, Lee HJ, Paik YK, Tang H, Lee JK, Park T. Enhanced peptide quantification using spectral count clustering and cluster abundance. BMC Bioinformatics. 2011; 12:423. [PubMed: 22034872]
9. Ong SE, Mann M. Mass spectrometry–based proteomics turns quantitative. Nat Chem Biol. 2005; 1(5):252–262. [PubMed: 16408053]

10. McClatchy DB, Dong MQ, Wu CC, Venable JD, Yates JR. 15N Metabolic Labeling of Mammalian Tissue with Slow Protein Turnover. J Proteome Res. 2007; 6(5):2005–2010. [PubMed: 17375949]
11. McClatchy DB, Liao L, Park SK, Xu T, Lu B, Yates JR III. Differential Proteomic Analysis of Mammalian Tissues Using SILAM. PLoS ONE. 2011; 6(1):e16039. [PubMed: 21283754]
12. Rauniyar N, McClatchy DB, Yates JR. Stable isotope labeling of mammals (SILAM) for in vivo quantitative proteomic analysis. Methods. 2013; 61(3):260–268. [PubMed: 23523555]
13. Zanivan, S., Krueger, M., Mann, M. In Vivo Quantitative Proteomics: The SILAC Mouse. Integrin and Cell Adhesion Molecules. In: Shimaoka, M., editor. Methods in Molecular Biology. Humana Press; 2012. p. 435-450.
14. Frank E, Kessler MS, Filiou MD, Zhang Y, Maccarrone G, Reckow S, Bunck M, Heumann H, Turck CW, Landgraf R, et al. Stable Isotope Metabolic Labeling with a Novel 15N-Enriched Bacteria Diet for Improved Proteomic Analyses of Mouse Models for Psychopathologies. PLoS ONE. 2009; 4(11):e7821. [PubMed: 19915716]
15. Toyama BH, Savas JN, Park SK, Harris MS, Ingolia NT, Yates JR, Hetzer MW. Identification of Long-Lived Proteins Reveals Exceptional Stability of Essential Cellular Structures. Cell. 2013; 154(5):971–982. [PubMed: 23993091]
16. Xiong W, Giannone RJ, Morowitz MJ, Banfield JF, Hettich RL. Development of an Enhanced Metaproteomic Approach for Deepening the Microbiome Characterization of the Human Infant Gut. J Proteome Res. 2015; 14(1):133–141. [PubMed: 25350865]
17. Wu J, Zhu J, Yin H, Liu X, An M, Pudlo NA, Martens EC, Chen GY, Lubman DM. Development of an Integrated Pipeline for Profiling Microbial Proteins from Mouse Fecal Samples by LC–MS/MS. J Proteome Res. 2016; 15(10):3635–3642. [PubMed: 27559751]
18. Chatterjee S, Stupp GS, Park SKR, Ducom JC, Yates JR, Su AI, Wolan DW. A comprehensive and scalable database search system for metaproteomics. BMC Genomics. 2016; 17:642. [PubMed: 27528457]
19. Zhang X, Ning Z, Mayne J, Moore JI, Li J, Butcher J, Deeke SA, Chen R, Chiang CK, Wen M, et al. MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota. Microbiome. 2016; 4:31. [PubMed: 27343061]
20. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. Nat Methods. 2014; 11(11):1114–1125. [PubMed: 25357241]
21. Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, Blazewicz SJ, Harden J, Turetsky MR, McGuire AD, Shah MB, et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature. 2015; 521(7551):208–212. [PubMed: 25739499]
22. Cravatt BF, Wright AT, Kozarich JW. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. Annu Rev Biochem. 2008; 77(1):383–414. [PubMed: 18366325]
23. Jessani N, Humphrey M, McDonald WH, Niessen S, Masuda K, Gangadharan B, Yates JR, Mueller BM, Cravatt BF. Carcinoma and stromal enzyme activity profiles associated with breast tumor growth in vivo. Proc Natl Acad Sci U S A. 2004; 101(38):13756–13761. [PubMed: 15356343]
24. Greenbaum DC, Baruch A, Grainger M, Bozdech Z, Medzihradszky KF, Engel J, DeRisi J, Holder AA, Bogyo M. A Role for the Protease Falcipain 1 in Host Cell Invasion by the Human Malaria Parasite. Science. 2002; 298(5600):2002–2006. [PubMed: 12471262]
25. Barglow KT, Cravatt BF. Discovering Disease-Associated Enzymes by Proteome Reactivity Profiling. Chem Biol. 2004; 11(11):1523–1531. [PubMed: 15556003]
26. Ostanin DV, Bao J, Koboziev I, Gray L, Robinson-Jackson SA, Kosloski-Davidson M, Price VH, Grisham MB. T cell transfer model of chronic colitis: concepts, considerations, and tricks of the trade. Am J Physiol - Gastrointest Liver Physiol. 2009; 296(2):G135–G146. [PubMed: 19033538]
27. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR. Metabolic Labeling of Mammalian Organisms with Stable Isotopes for Quantitative Proteomic Analysis. Anal Chem. 2004; 76(17): 4951–4959. [PubMed: 15373428]
28. Wolters DA, Washburn MP, Yates JR. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001; 73(23):5683–5690. [PubMed: 11774908]

29. McClatchy, DB., Yates, JR, III. Stable Isotope Labeling in Mammals (SILAM). In: Martins-de-Souza, D., editor. Shotgun Proteomics. Vol. 1156. Springer New York; New York, NY: 2014. p. 133-146.Methods in Molecular Biology
30. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995; 57(1):289–300.
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005; 102(43): 15545–15550. [PubMed: 16199517]
32. Karbach U, Ewe K, Bodenstein H. Alpha 1-antitrypsin, a reliable endogenous marker for intestinal protein loss and its application in patients with Crohn's disease. Gut. 1983; 24(8):718–723. [PubMed: 6603392]
33. Fischbach W, Becker W, Mössner J, Koch W, Reiners C. Faecal alpha-1-antitrypsin and excretion of 111indium granulocytes in assessment of disease activity in chronic inflammatory bowel diseases. Gut. 1987; 28(4):386–393. [PubMed: 3495470]
34. Churg A, Dai J, Zay K, Karsan A, Hendricks R, Yee C, Martin R, MacKenzie R, Xie C, Zhang L, et al. Alpha-1-antitrypsin and a broad spectrum metalloprotease inhibitor, RS113456, have similar acute anti-inflammatory effects. Lab Investig J Tech Methods Pathol. 2001; 81(8):1119–1131.
35. Collins CB, Aherne CM, Ehrentraut SF, Gerich ME, McNamee EN, McManus MC, Lebsack MDP, Jedlicka P, Azam T, de Zoeten EF, et al. Alpha-1-antitrypsin therapy ameliorates acute colitis and chronic murine ileitis. Inflamm Bowel Dis. 2013; 19(9):1964–1973. [PubMed: 23835442]
36. Wu F, Chakravarti S. Differential Expression of Inflammatory and Fibrogenic Genes and Their Regulation by NF-κB Inhibition in a Mouse Model of Chronic Colitis. J Immunol. 2007; 179(10): 6988–7000. [PubMed: 17982090]
37. Parikh A, Stephan AF, Tzanakakis ES. Regenerating proteins and their expression, regulation and signaling. Biomol Concepts. 2012; 3(1):57–70. [PubMed: 22582090]
38. Ryckman C, Vandal K, Rouleau P, Talbot M, Tessier PA. Proinflammatory Activities of S100: Proteins S100A8, S100A9, and S100A8/A9 Induce Neutrophil Chemotaxis and Adhesion. J Immunol. 2003; 170(6):3233–3242. [PubMed: 12626582]
39. Koike A, Arai S, Yamada S, Nagae A, Saita N, Itoh H, Uemoto S, Totani M, Ikemoto M. Dynamic Mobility of Immunological Cells Expressing S100A8 and S100A9 in vivo: A Variety of Functional Roles of the two Proteins as Regulators in Acute Inflammatory Reaction. Inflammation. 2011; 35(2):409–419.
40. Png CW, Lindén SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin THJ. Mucolytic Bacteria With Increased Prevalence in IBD Mucosa Augment In Vitro Utilization of Mucin by Other Bacteria. Am J Gastroenterol. 2010; 105(11):2420–2428. [PubMed: 20648002]
41. Ganesh BP, Klopfleisch R, Loh G, Blaut M. Commensal Akkermansia muciniphila Exacerbates Gut Inflammation in Salmonella Typhimurium-Infected Gnotobiotic Mice. PLOS ONE. 2013; 8(9):e74963. [PubMed: 24040367]
42. Wang W, Chen L, Zhou R, Wang X, Song L, Huang S, Wang G, Xia B. Increased Proportions of Bifidobacterium and the Lactobacillus Group and Loss of Butyrate-Producing Bacteria in Inflammatory Bowel Disease. J Clin Microbiol. 2014; 52(2):398–406. [PubMed: 24478468]
43. Derrien M, van Passel MWJ, van de Bovenkamp JHB, Schipper R, de Vos W, Dekker J. Mucin-bacterial interactions in the human oral cavity and digestive tract. Gut Microbes. 2010; 1(4):254–268. [PubMed: 21327032]
44. Knights D, Lassen KG, Xavier RJ. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. Gut. 2013; 62(10):1505–1510. [PubMed: 24037875]
45. Drew JE, Farquharson AJ, Keijer J, Barrera B, LN. Complex regulation of mucosal pentraxin (Mptx) revealed by discrete micro-anatomical locations in colon. Biochim Biophys Acta BBA - Mol Basis Dis. 2006; 1762(9):844–848.
46. van der Meer-van Kraaij C, Siezen R, Kramer E, Reinders M, Blokzijl H, van der Meer R, Keijer J. Dietary modulation and structure prediction of rat mucosal pentraxin (Mptx) protein and loss of function in humans. Genes Nutr. 2007; 2(3):275–285. [PubMed: 18850182]

47. Fang K, Bruce M, Pattillo CB, Zhang S, Stone R, Clifford J, Kevil CG. Temporal genomewide expression profiling of DSS colitis reveals novel inflammatory and angiogenesis genes similar to ulcerative colitis. Physiol Genomics. 2011; 43(1):43–56. [PubMed: 20923862]
48. Wrackmeyer U, Hansen GH, Seya T, Danielsen EM. Intelectin: A Novel Lipid Raft-Associated Protein in the Enterocyte Brush Border. Biochemistry (Mosc). 2006; 45(30):9188–9197.
49. Zhang H, Massey D, Tremelling M, Parkes M. Genetics of inflammatory bowel disease: clues to pathogenesis. Br Med Bull. 2008; 87(1):17–30. [PubMed: 18753178]
50. Kelly P, Feakins R, Domizio P, Murphy J, Bevins C, Wilson J, Mcphail G, Poulsom R, Dhaliwal W. Paneth cell granule depletion in the human small intestine under infective and nutritional stress. Clin Exp Immunol. 2004; 135(2):303–309. [PubMed: 14738460]
51. Lisitsyn NA, Bukurova YA, Nikitina IG, Krasnov GS, Sykulev Y, Beresten SF. Enteric alpha defensins in norm and pathology. Ann Clin Microbiol Antimicrob. 2012; 11:1. [PubMed: 22236533]
52. Frank DN, Amand ALS, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci. 2007; 104(34):13780–13785. [PubMed: 17699621]
53. Eckburg PB. Diversity of the Human Intestinal Microbial Flora. Science. 2005; 308(5728):1635–1638. [PubMed: 15831718]
54. Arrieta MC, Finlay BB. The commensal microbiota drives immune homeostasis. Mol Innate Immun. 2012; 3:33.
55. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. Cell Host Microbe. 2007; 2(2):119–129. [PubMed: 18005726]
56. Mukhopadhya I, Hansen R, El-Omar EM, Hold GL. IBD—what role do Proteobacteria play? Nat Rev Gastroenterol Hepatol. 2012; 9(4):219–230. [PubMed: 22349170]
57. Rigottier-Gois L. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. ISME J. 2013; 7(7):1256–1261. [PubMed: 23677008]
58. Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, Jian R, Doré J. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut. 2003; 52(2):237–242. [PubMed: 12524406]
59. Kotlowski R, Bernstein CN, Sepehri S, Krause DO. High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. Gut. 2007; 56(5):669–675. [PubMed: 17028128]
60. Wills ES, Jonkers DMAE, Savelkoul PH, Masclee AA, Pierik MJ, Penders J. Fecal Microbial Composition of Ulcerative Colitis and Crohn's Disease Patients in Remission and Subsequent Exacerbation. PLOS ONE. 2014; 9(3):e90981. [PubMed: 24608638]
61. Bibiloni R, Mangold M, Madsen KL, Fedorak RN, Tannock GW. The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. J Med Microbiol. 2006; 55(8):1141–1149. [PubMed: 16849736]
62. Tripathi LP, Sowdhamini R. Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes. BMC Genomics. 2008; 9:549. [PubMed: 19019219]
63. Gibson SA, McFarlan C, Hay S, MacFarlane GT. Significance of microflora in proteolysis in the colon. Appl Environ Microbiol. 1989; 55(3):679–683. [PubMed: 2648991]
64. Carroll IM, Maharshak N. Enteric bacterial proteases in inflammatory bowel disease-pathophysiology and clinical implications. World J Gastroenterol. 2013; 19(43):7531–7543. [PubMed: 24431894]

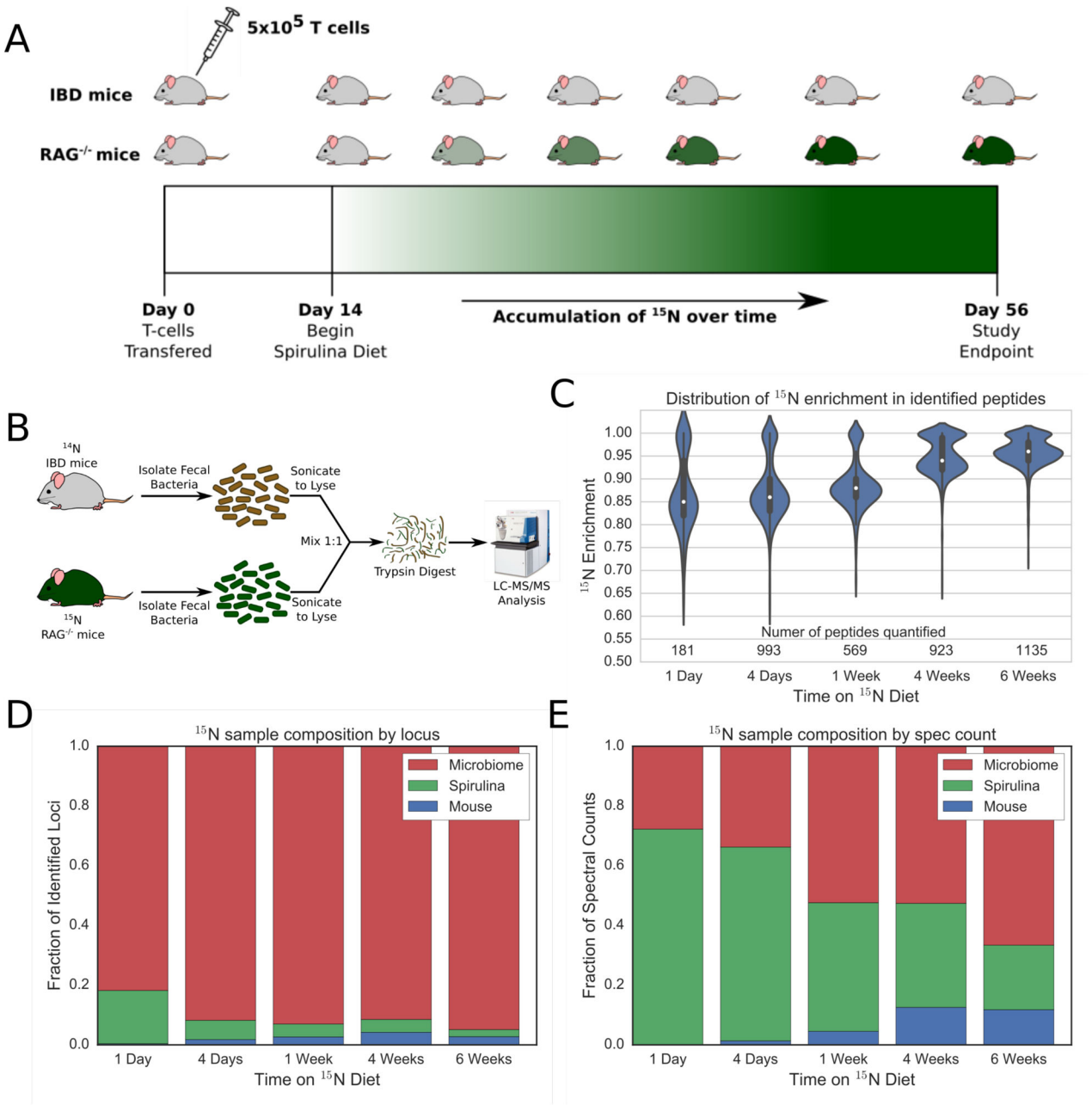

Figure 1.

Analysis of ^{15}N enrichment in the mouse fecal microbiome. (A) Experimental design schematic detailing the enrichment of ^{15}N over time in the mouse fecal microbiome. The eight week old, female *Rag1*$^{-/-}$ mice were divided into two cohorts. The first received a T cell transfer to induce colitis (IBD mice), while the second cohort served as a healthy control (RAG$^{-/-}$ mice). Starting 14 days post transfer, the IBD cohort was placed on an ^{14}N control spirulina diet, while the RAG$^{-/-}$ mice were fed ^{15}N spirulina. The mice were sacrificed 56 days post transfer, once the IBD cohort showed signs of disease and maximal ^{15}N labeling had been achieved. (B) Schematic detailing the unenriched sample preparation process.

Bacterial cells were isolated and lysed, before mixing proteome in a 1:1 ratio. Samples were then trypsinized and analyze via mass spectrometry. (C) Violin plot of ^{15}N enrichment in peptides at various time points after starting on the spirulina-based diet. Stacked bar plots showing the fraction of protein loci (D) or spectral counts (E) attributable to species sources.

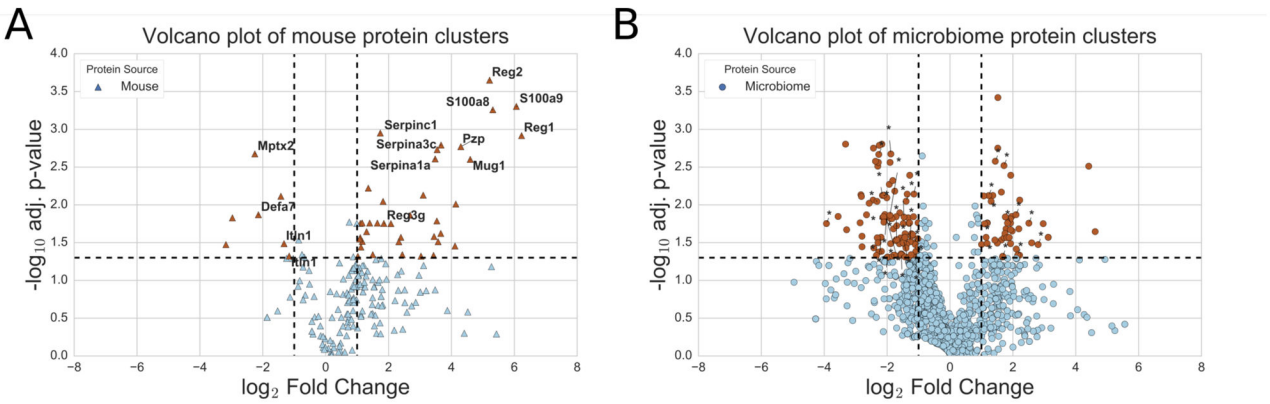

Figure 2.

Volcano plot of identified protein clusters from either (A) host or (B) microbiome sources. Protein clusters that show significantly different abundance between the $RAG^{-/-}$ and IBD mice are shown in red. Protein clusters with significantly increased abundance in IBD mice are in the upper right of the graph, while those more abundant $RAG^{-/-}$ mice are in the upper left. Proteins in panel A that have been highlighted in the text are labeled with their gene names. Proteins in panel B that have no known associated functional annotations are highlighted with an asterisk. A full list of the significantly differentially expressed proteins is available (Supporting Information, Tables S2, S3, and S4).

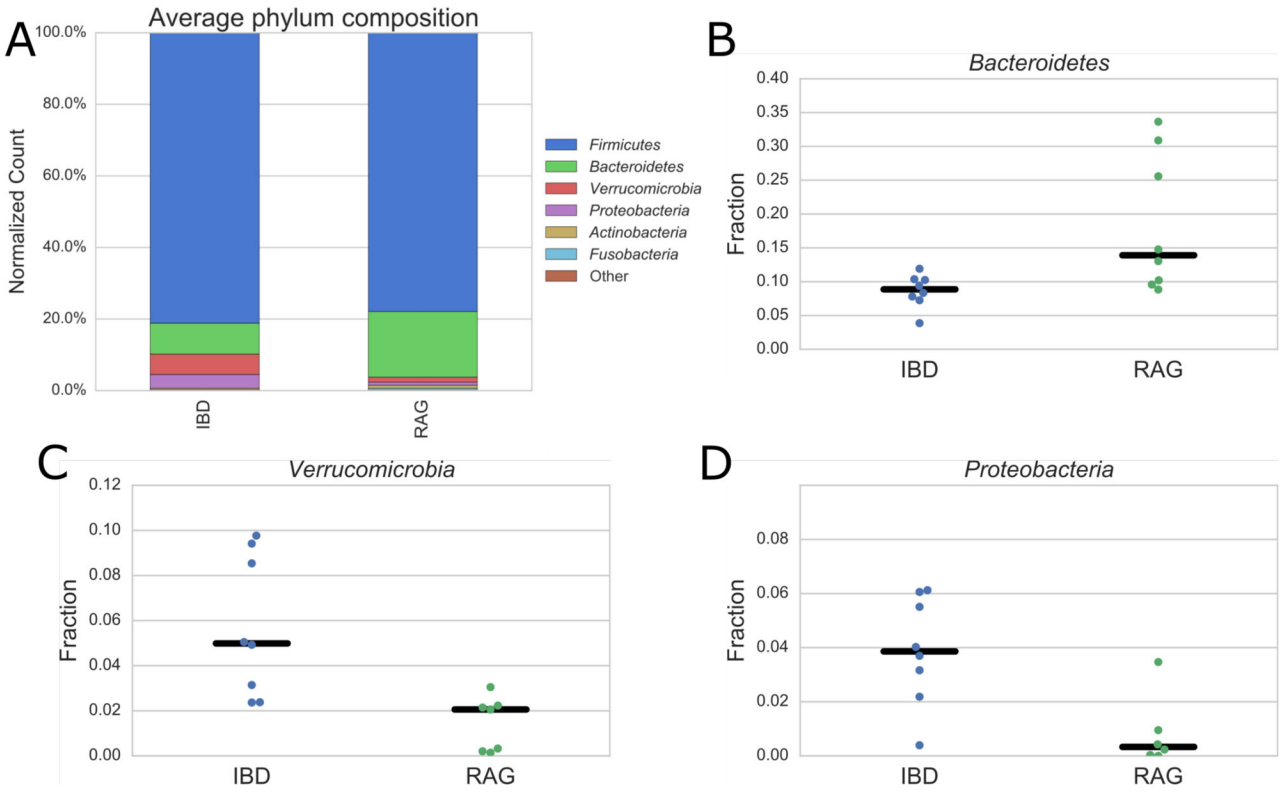

Figure 3.

Plots pertaining to taxonomic analysis. Bar plot (A) of averaged peptide spectral counts attributable to a single phylum obtained from LC-MS/MS-based metaproteomics of RAG$^{-/-}$ control versus IBD mice. (B–D) Plots showing the levels of spectral counts attributable to a single phylum in each individual experimental replicate. The phyla shown were found to be significantly different ($p < .05$) via t-test. The phyla *Verrucomicrobia* in panel C and *Proteobacteria* in panel D reached significance levels of ($p < .01$).

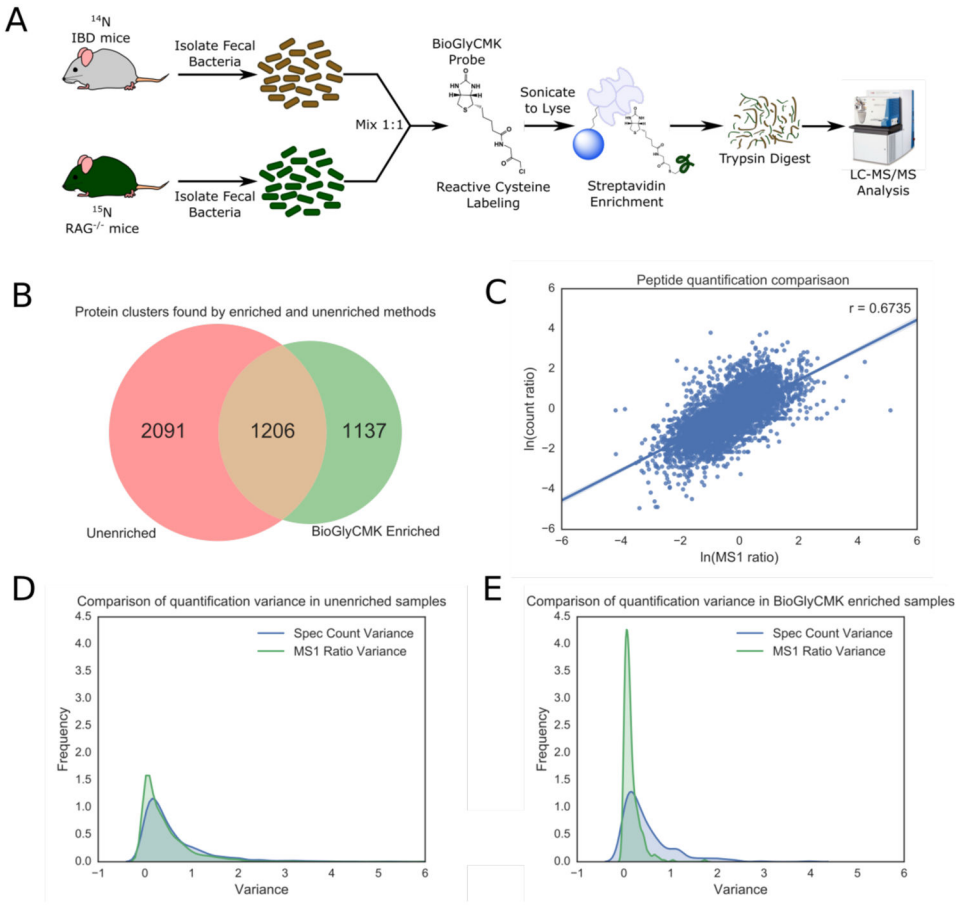

Figure 4.

Comparison of activity based probe enriched samples. (A) Schematic demonstrating the workflow for the ABP enriched samples. Bacterial cells were mixed in a 1:1 ratio before activity based labeling and lysed after the process. A BioGlyCMK probe was used to target reactive cysteines in the proteome. (B) Venn diagram showing the differences in protein clusters found via the unenriched protocol as compared to those found via BioGlyCMK enrichment. (C) Log-log plot of peptide ratios determined via corrected spectral count ratios vs ratios determined from MS1 precursor ion intensities. Run to run variance in the ratios of low abundance peptides in the unenriched (D) or BioGlyCMK enriched (E) samples. Ratios were generated either via spec counts or via MS1 precursor ion intensities.

Table 1

GO Terms enriched in BioGlyCMK labeled samples over the unenriched samples

GO Term	Description	Odds ratio	p-value	Num. Annotated	Num. in Group	Num. Expected
GO:0008234	Cys-type peptidase	14.9	$3.9 \cdot 10^{-09}$	25	21	12
GO:0016153	urocanate hydratase	14.8	$4.3 \cdot 10^{-05}$	13	11	8
GO:0004197	Cys-type endopeptidase	8.9	$4.4 \cdot 10^{-05}$	17	13	9
GO:0004022	alcohol dehydrogenase (NAD)	8.2	$1.3 \cdot 10^{-04}$	16	12	9
GO:0004825	methionine-tRNA ligase	7.9	$1.3 \cdot 10^{-03}$	12	9	7
GO:0008774	acetaldehyde dehydrogenase	7.4	$3.9 \cdot 10^{-04}$	15	11	8
GO:0016879	ligase, C-N bonds	3.8	$2.7 \cdot 10^{-04}$	33	19	15
GO:0016810	linear amide hydrolase, acting on C-N (but not peptide) bonds	3.4	$8.1 \cdot 10^{-05}$	46	25	20
GO:0004177	aminopeptidase	3.3	$9.3 \cdot 10^{-03}$	22	12	11
GO:0016620	oxidoreductase, acting on the aldehyde or oxo group of donors, NAD or NADP acceptor	3.2	$5.7 \cdot 10^{-11}$	140	72	49
GO:0016811	hydrolase, acting on C-N	3.2	$3.2 \cdot 10^{-03}$	28	15	13
GO:0016874	ligase	2.9	$1.3 \cdot 10^{-05}$	77	38	30
GO:0016638	oxidoreductase, acting on the CH-NH_2 group of donors	2.7	$3.0 \cdot 10^{-07}$	115	55	41
GO:0016639	oxidoreductase, acting on the CH-NH_2 group of donors, NAD or NADP acceptor	2.5	$7.4 \cdot 10^{-06}$	107	49	39
GO:0008233	peptidase	2.3	$9.1 \cdot 10^{-05}$	100	44	37
GO:0016746	acyl group transferase	2.1	$1.3 \cdot 10^{-03}$	91	38	34
GO:0016781	phosphotransferase, paired acceptors	2.0	$3.4 \cdot 10^{-05}$	173	69	59
GO:0016835	C-O lyase	1.9	$2.4 \cdot 10^{-03}$	110	43	40
GO:0016747	acyl group transferase activity, transferring acyl groups other than amino-acyl groups	1.9	$7.3 \cdot 10^{-03}$	84	33	32
GO:0016836	hydrolyase	1.8	$6.6 \cdot 10^{-03}$	108	41	39

Table 2

GSEA showing Molecular Function GO Terms IBD vs. $RAG^{-/-}$ control mice, enriched via BioGlyCMK probe

Term	Description	Enrichment Score (ES)	Norm. ES	p-value	FDR	Size	Num. Matched
GO:0008233	peptidase	0.76	2.28	0.0 *	$6.2 \cdot 10^{-04}$	70	12
GO:0008234	Cys-type peptidase	0.93	2.27	0.0	$3.1 \cdot 10^{-04}$	24	7
GO:0004197	Cys-type endopeptidase	0.93	2.21	0.0	$2.1 \cdot 10^{-04}$	16	6
GO:0004175	endopeptidase	0.83	2.16	0.0	$4.7 \cdot 10^{-04}$	29	8
GO:0016860	intramolecular oxidoreductase	0.80	2.00	$1.8 \cdot 10^{-03}$	$4.4 \cdot 10^{-03}$	32	7
GO:0016787	hydrolase	0.42	1.91	0.0	$1.0 \cdot 10^{-02}$	267	64
GO:0016810	hydrolase, acting on C-N (but not peptide) bonds	0.78	1.89	0.0	$1.0 \cdot 10^{-02}$	35	7
GO:0016798	hydrolase, acting on glycosyl bonds	0.57	1.77	$8.6 \cdot 10^{-03}$	$2.9 \cdot 10^{-02}$	33	14
GO:0004553	hydrolase, acting on O-glycosyl compounds	0.59	1.75	$1.8 \cdot 10^{-02}$	$3.3 \cdot 10^{-02}$	31	13
GO:0003676	nucleic acid binding	0.56	1.54	$4.6 \cdot 10^{-02}$	$1.3 \cdot 10^{-01}$	73	10
GO:0000287	magnesium ion binding	0.37	1.49	$4.4 \cdot 10^{-02}$	$1.5 \cdot 10^{-01}$	72	37

* indicates a nominal p-value of < 1/1000