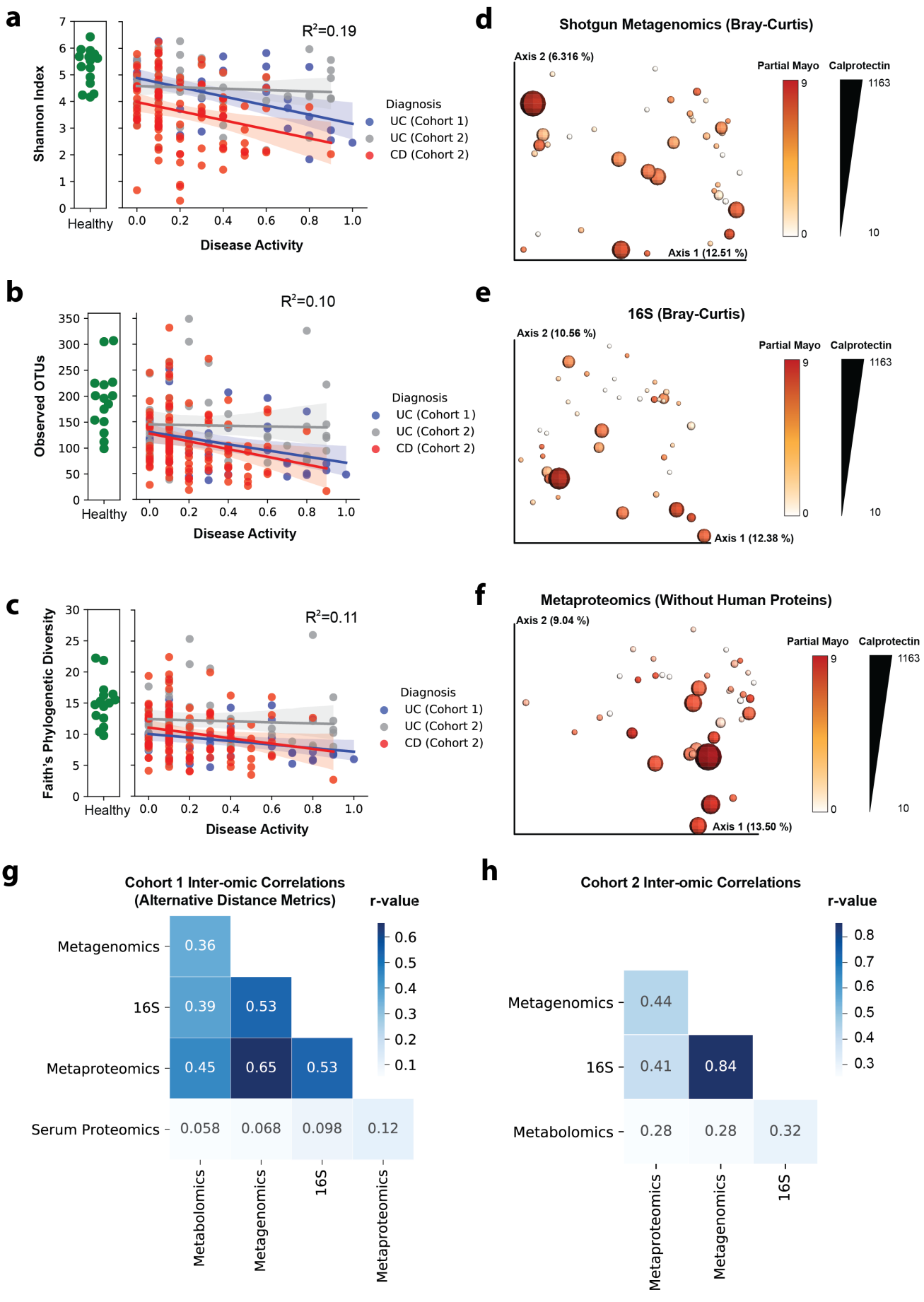
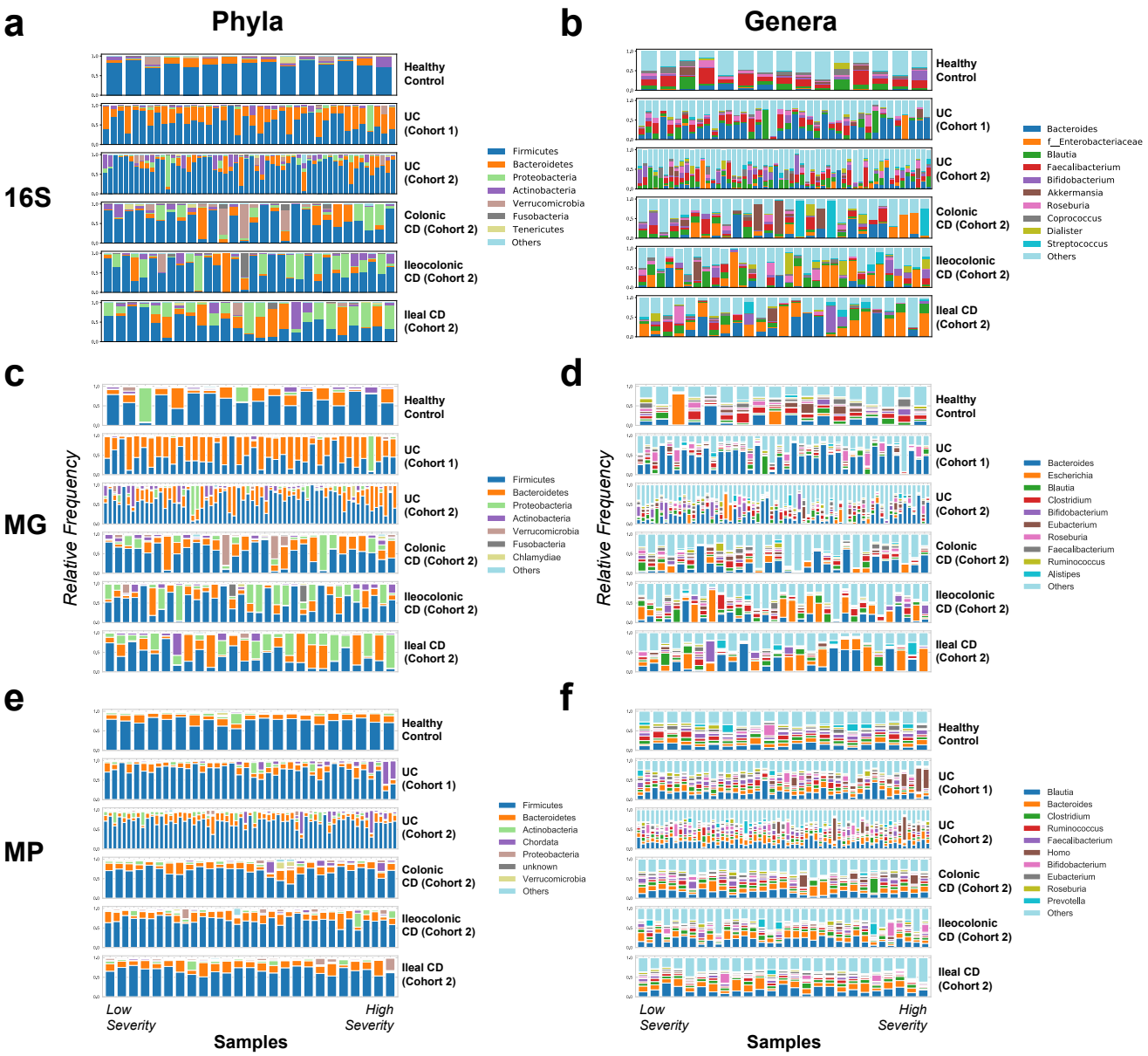
Supplementary Table 1: IBD patient characteristics

Baseline Characteristic	UC patients Cohort 1 (n=40)	UC patients Cohort 2 (n=73)	P value (UC Cohort 1 vs 2)	CD patients Cohort 2 (n=117)
Age, median (IQR)	38.5 (28.5 – 52.5)	47 (33.0-56.0)	0.51	35 (26-47)
Female gender, n (%)	11 (28%)	32 (44%)	<0.01	53 (34%)
Caucasian race, n (%)	30 (75%)	48 (66%)	0.40	94 (80%)
Smoker, n (%)	3 (7.5%)	10 (13.6%)	0.38	23 (20%)
Disease duration, median (IQR)	8 (4.5 – 19)	12 (6-20)	0.46	10 (5-17)
Historic extent pancolitis, n (%)	18 (45%)	28 (38%)	0.55	-
Stricturing/Penetrating Crohn's, n (%)	-	-	-	37 (32%)
Colonic involvement Crohn's, n (%)	-	-	-	79 (68%)
Prior bowel surgery, n (%)	-	-	-	46 (39%)
Prior 5-ASA exposure, n (%)	36 (90%)	65 (89%)	1.00	-
Current 5-ASA exposure, n (%)	20 (50%)	41 (56%)	0.56	-
Prior steroid exposure, n (%)	29 (73%)	50 (68%)	0.83	99 (85%)
Current steroid exposure, n (%)	6 (15%)	15 (21%)	0.61	70 (60%)
Prior IS exposure, n (%)	16 (40%)	37 (51%)	0.33	94 (80%)
Current IS exposure, n (%)	7 (18%)	18 (25%)	0.48	76 (65%)
Biologic exposure, n (%)	19 (48%)	40 (55%)	0.55	88 (75%)
Partial Mayo score, median (IQR)	3 (1 – 5.75)	1 (1 – 4)	0.36	-
Moderate-Severe endoscopic activity, n (%)	19 (48%)	22 (30%)	0.10	-
Simple Endoscopic Score Crohn's disease, median (IQR)	-	-	-	3 (0-7)

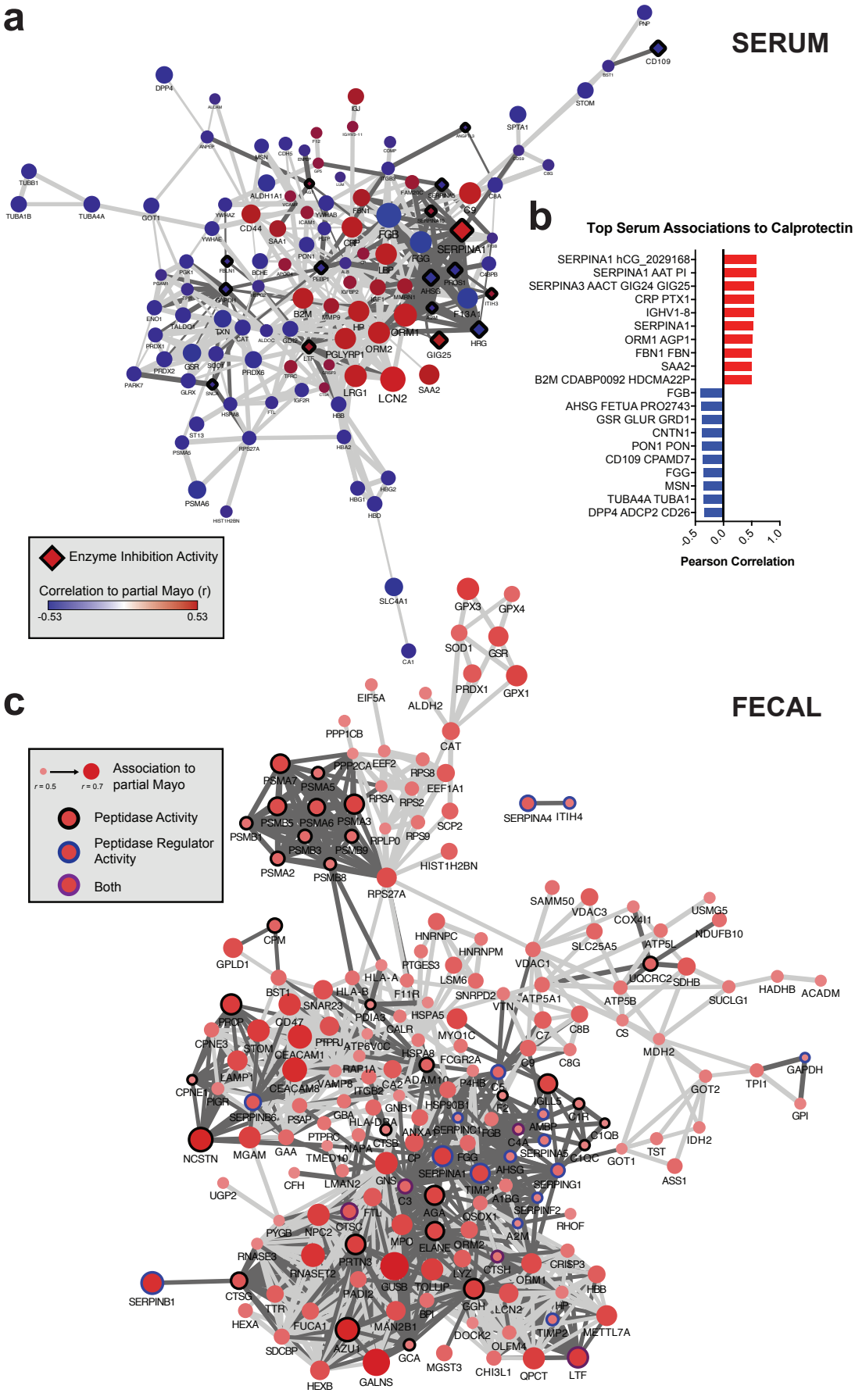
All baseline characteristics are at time of sample collection. Endoscopic scoring done blinded to clinical data or biomarker data, and histologic scoring done blinded to clinical, biomarker, and endoscopic data. Endoscopic scoring was performed by a physician with expertise and advanced training in IBD, and histologic scoring was performed by a pathologist with expertise and advanced training in GI pathology. 5-ASA – 5 aminosalicylate therapy. IS: azathioprine, 6-mercaptopurine, methotrexate. Biologics included TNF-antagonists, Vedolizumab, Ustekinumab, tofacitinib.


a
Shannon Index
R²=0.19
Healthy
Disease Activity
Diagnosis
UC (Cohort 1)
UC (Cohort 2)
CD (Cohort 2)
b
Observed OTUs
R²=0.10
c
Faith's Phylogenetic Diversity
R²=0.11
d
Shotgun Metagenomics (Bray-Curtis)
Axis 2 (6.316 %)
Axis 1 (12.51 %)
Partial Mayo
Calprotectin
e
16S (Bray-Curtis)
Axis 2 (10.56 %)
Axis 1 (12.38 %)
f
Metaproteomics (Without Human Proteins)
Axis 2 (9.04 %)
Axis 1 (13.50 %)
g
Cohort 1 Inter-omic Correlations
(Alternative Distance Metrics)
r-value
Metagenomics
16S
Metaproteomics
Serum Proteomics
Metabolomics
h
Cohort 2 Inter-omic Correlations
Metabolomics

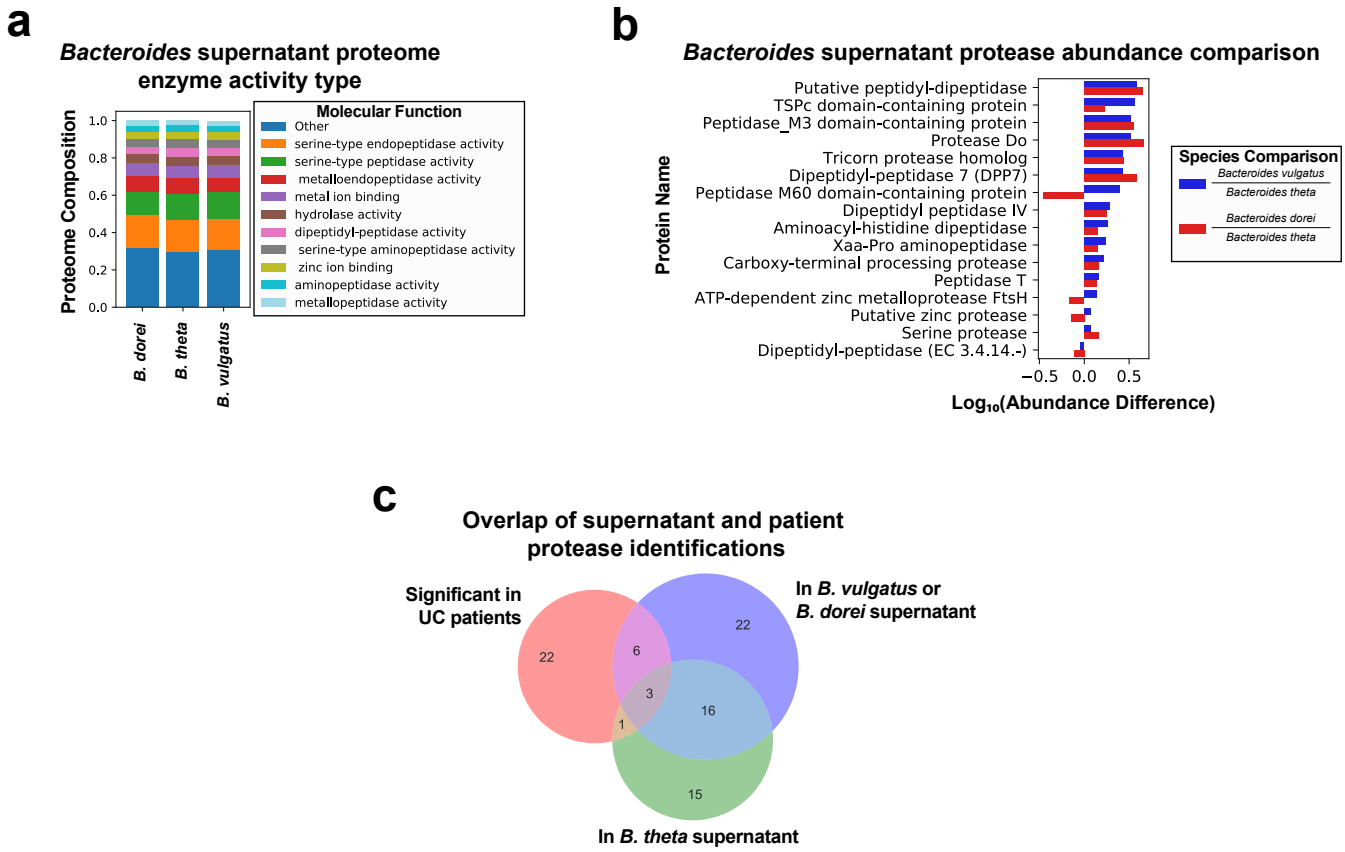
Supplementary Figure 1. Evaluating alpha and beta-diversity relationships to disease activity and data type comparisons. In addition to Pielou's evenness metric for alpha-diversity reported in Figure 1b, other alpha-diversity metrics were tested for their relationship to disease activity in 16S data. These metrics included Shannon's index (**a**), Observed OTUs (**b**), and Faith's PD (**c**). Best-fit linear regression lines with 95% confidence intervals are shown and R^2 statistics are reported from an ordinary least-squares regression using the formula (Disease Activity + Diagnosis + Disease Activity:Diagnosis). **d,** PCoA of shotgun metagenomic data from UC cohort 1. Bray-Curtis distance metric was used and the first two principal coordinates are displayed. Samples are sized by fecal calprotectin abundance and colored by the associated partial Mayo severity scores. **e,** PCoA of 16S data using the Bray-Curtis distance metric. Samples were sized and colored as described for (d). **f,** PCoA of fecal metaproteomics data after host proteins were removed using the Bray-Curtis distance metric. Samples were sized and colored as described for (d). Host proteins displayed a strong influence in the metaproteome's relationship to disease activity, and the microbial contribution can be seen here. **g,** Data type correlations. Pearson correlations between data types are displayed in a heat map. The Bray-Curtis distance metric was used for all data types other than the metagenome and 16S, which used unweighted UniFrac. Correlations were performed on distance matrices from cohort 1 through Mantel test. **h,** Data type correlations for cohort 2. As described in (g) using all samples processed for every –omic from cohort 2. Correlations were performed on Bray-Curtis distance matrices.



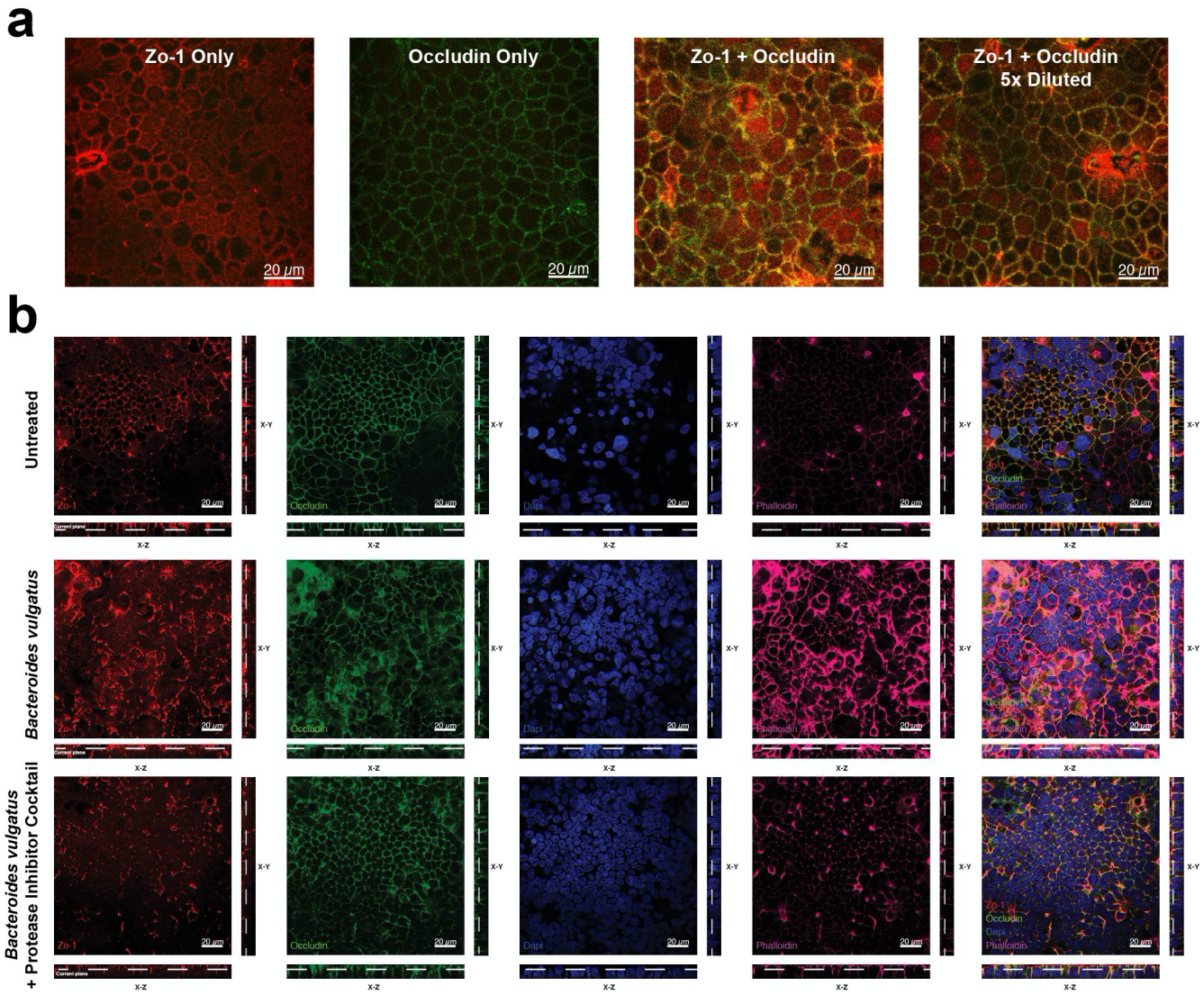
Supplementary Figure 2. Taxonomic composition plots by data type. Phyla composition of the 16S (**a**), MG (**b**), and MP (**c**) data are shown for each fecal sample analyzed. Samples are ordered by the relative disease activity the patient is currently experiencing. Patients are split by independently processed cohort for UC patients and diagnosis. The top 7 most abundant phyla are displayed with all other phyla grouped in "Others". Genera composition plots for each patient is also shown (**b**-**f**). For genera plots, the top 10 genera are displayed with all other phyla grouped in "Others"


a
SERUM
Enzyme Inhibition Activity
Correlation to partial Mayo (r)
-0.53
0.53
b
Top Serum Associations to Calprotectin
SERPINA1 hCG_2029168
SERPINA1 AAT PI
SERPINA3 AACT GIG24 GIG25
CRP PTX1
IGHV1-8
SERPINA1
ORM1 AGP1
FBN1 FBN
SAA2
B2M CDABP0092 HDCMA22P
FGB
AHSG FETUA PRO2743
GSR GLUR GRD1
CNTN1
PON1 PON
CD109 CPAMD7
FGG
MSN
TUBA4A TUBA1
DPP4 ADCP2 CD26
-0.5
0.0
0.5
1.0
Pearson Correlation
c
FECAL
Association to partial Mayo
r = 0.5
r = 0.7
Peptidase Activity
Peptidase Regulator Activity
Both

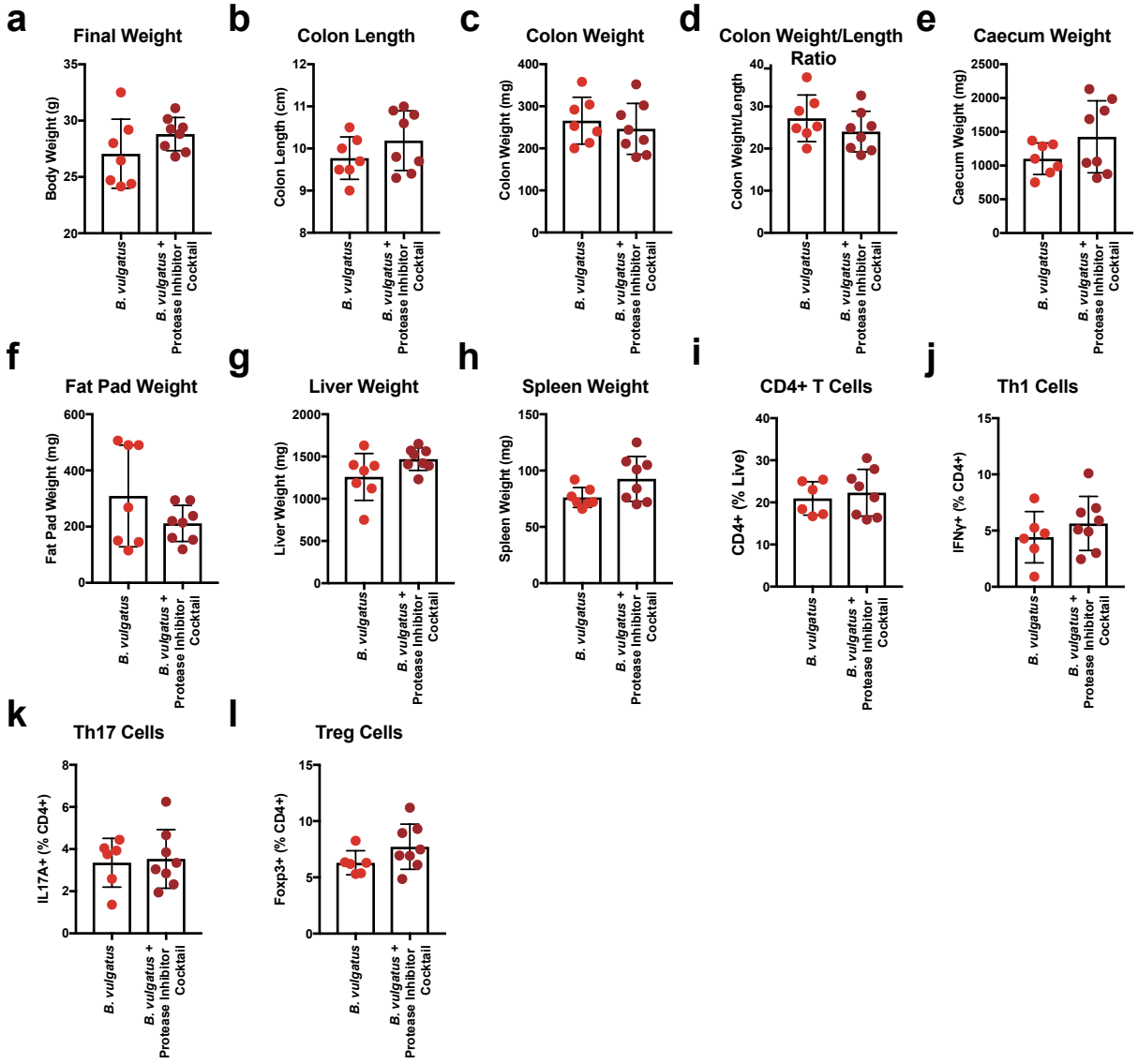
Supplementary Figure 3. Host protein networks highlight that host enzyme activity and regulation is associated with UC severity. a, Enzyme inhibition is enriched within serum proteins related to disease activity**.** A network is displayed showing connections of serum proteins correlated to partial Mayo disease activity ($|r| > 0.2$), which were determined through String-DB. Edges are sized by combined confidence in the interaction, with only high confidence connections displayed (> 0.7; theoretical max confidence = 1). Nodes are sized and colored depending on the correlation to the partial Mayo disease activity score. A functional enrichment for enzyme inhibition activity was found and proteins annotated with this function are indicated by a diamond shape and a black border. The connecting edges of enzyme inhibitor proteins are colored dark gray while all other connections are colored light gray. **b**, Serum proteins associated to Calprotectin. Pearson correlations were performed comparing fecal calprotectin relative abundances to serum proteome relative abundances. The top 10 positively and negatively correlated serum proteins are plotted by their Pearson correlation coefficient (r). Positively associated proteins are shown in red while negatively associated proteins are shown in blue. **c**, Peptidase related proteins are highly connected within fecal exosome proteins correlated to disease activity. Exosome proteins were highly enriched among the human fecal proteins correlated to Ulcerative Colitis disease activity. Displayed is a network of exosome proteins highly correlated with the partial Mayo severity score ($r > 0.5$). Edges are sized by combined confidence in the interaction, with only high confidence connections displayed (> 0.7; theoretical max confidence = 1). Nodes are sized and colored depending on the correlation to the partial Mayo disease activity score. A functional enrichment for peptidase activity and peptidase regulator activity was found and a thickened border indicates proteins annotated with these functions. Edges connecting to proteins with peptidase related activity are colored dark gray while all other connections are colored light gray.



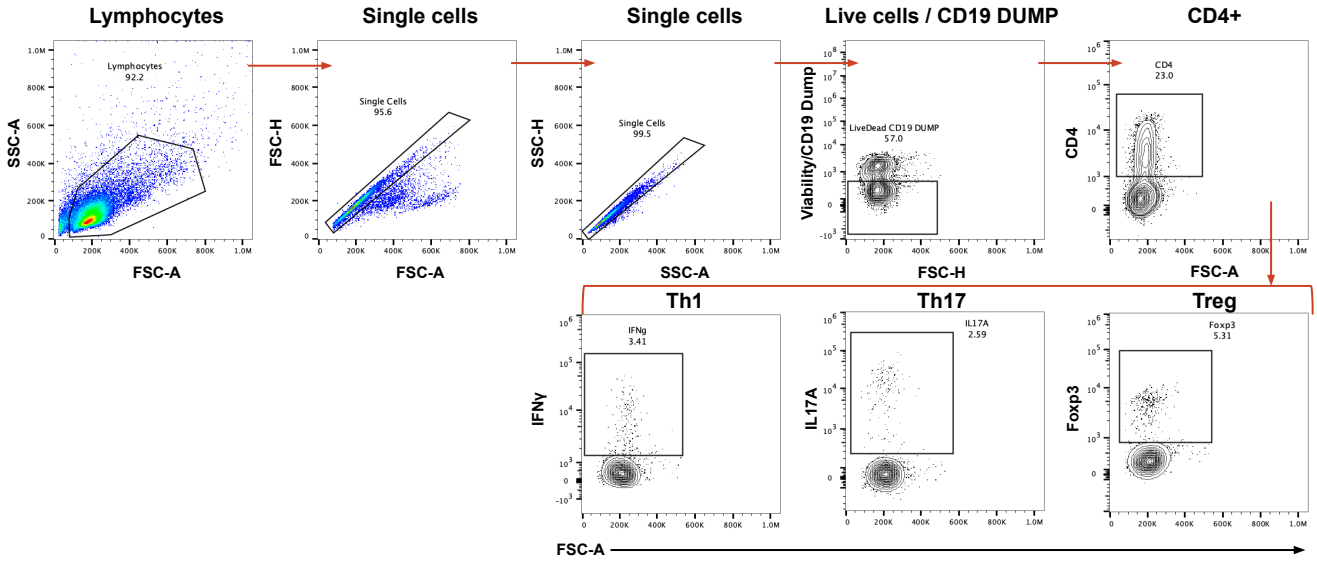
Supplementary Figure 4. Proteomic investigation of proteases in *Bacteroides* supernatant. a, Molecular function composition of enzymes and proteases in *Bacteroides* supernatant. Supernatant from overnight cultures of *B. dorei*, *B. theta*, and *B. vulgatus*, were analyzed by LC-MS^3 based proteomics. The average composition of GO Molecular Function for any protein annotated as an enzyme, peptidase or protease is displayed by species. **b,** Proteases specifically enriched within *B. vulgatus* supernatant. Proteases abundant within *B. vulgatus* supernatant were ranked and plotted by the relative abundance difference of each protease within *B. vulgatus* supernatant in comparison to *B. theta* supernatant. Comparisons between *B. dorei* and *B. theta* are also shown. **c**, Comparison of *B. vulgatus* proteases identified in UC patients and in supernatant proteomics. Protein names were compared for *B. vulgatus* or *B. dorei* proteases correlated to UC activity ($r > 0.3$) in either cohort, with the proteases identified in *B. vulgatus* or *B. dorei* supernatant, and the proteases identified in *B. theta* supernatant.



Supplementary Figure 5. Supplementary images from confocal microscopy of Caco-2 monolayer and *Bacteroides* co-culture experiments. a, Evaluating specificity and optimum concentration of primary antibodies. Representative images from untreated Caco-2 cells after imaging preparation with only Zo-1 primary antibodies added (far left), only occludin antibodies added (second from left), both primary antibodies at the recommended dilution (second from right), and both primary antibodies diluted 5x from recommended dilution. The antibodies were determined to be specific and the 5x dilution of primary antibodies were used for remaining imaging experiments. Images representative of n=2 experiments. **b,** Individual channels from imaging study displayed in Fig. 5d. Zo-1 is displayed in red in the far left, Occludin is shown second from the left, dapi is shown in the middle, phalloidin is shown second from the right, and the merge of all 4 channels is shown on the far right. Images are representative of n=2 transwells wells per group from n=2 independent experiments. Scale bars are 20 μm (**a**, **b**).



Supplementary Figure 6. Additional measurements from *B. vulgatus* monocolonization experiments. a-h Barplots showing the mean +/- SD of are shown for macroscopic organ measurements from *B. vulgatus* monocolonized IL10$^{-/-}$ mice with or without administration of a protease inhibitor cocktail in the drinking water over a 10-week colonization. Measurements include final weight of the mice (**a**), colon length (**b**), colon weight (**c**), ratios of the colon weight to length (**d**), caecum weight (**e**), fat pad weight (**f**), liver weight (**g**) and spleen weight (**h**). **i-l** Barplots showing the mean +/- SD for immune cells of the mesenteric lymph node generated through flow cytometry. Cell populations included the percent of live CD4+ T-cells (**i**), Th1 cells (**j**), Th17 cells (**k**), and Treg cells (**l**). Data from b-m represents n=6, 8 animals for *B. vulgatus* and *B. vulgatus* + Protease Inhibitor Cocktail groups examined over 2 independent experiments. The gating strategy for panels i-l is presented in Supplementary Figure 7.



Supplementary Figure 7. Flow cytometry gating strategy. Outlined is the gating strategy corresponding to data from Supplementary Figure 6i-l. A 7-color panel was used with markers for Foxp3 (FITC), IL10 (PE), IL17A (PE-Cy7), IFNg (APC-eFluor 780), CD25 (APC), CD4 (eFluor 450), Viability/CD19 DUMP (Aqua/BV510).