AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Host-Microbe Biology

Evaluating Metagenomic Prediction of the Metaproteome in a 4.5-Year Study of a Patient with Crohn's Disease

Robert H. Mills,[a,b,c,g] Yoshiki Vázquez-Baeza,[c] Qiyun Zhu,[c] Lingjing Jiang,[c,d] James Gaffney,[c] Greg Humphrey,[c] Larry Smarr,[e,f,g] Rob Knight,[c,g] David J. Gonzalez[a,b,g]

[a]Department of Pharmacology, University of California, San Diego, San Diego, California, USA
[b]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, San Diego, California, USA
[c]Department of Pediatrics, and Department of Computer Science and Engineering, University of California, San Diego, San Diego, California, USA
[d]Department of Family Medicine and Public Health, University of California, San Diego, San Diego, California, USA
[e]Department of Computer Science and Engineering, University of California, San Diego, San Diego, California, USA
[f]California Institute for Telecommunications and Information Technology, University of California, San Diego, San Diego, California, USA
[g]Center for Microbiome Innovation, University of California, San Diego, San Diego, California, USA

ABSTRACT Although genetic approaches are the standard in microbiome analysis, proteome-level information is largely absent. This discrepancy warrants a better understanding of the relationship between gene copy number and protein abundance, as this is crucial information for inferring protein-level changes from metagenomic data. As it remains unknown how metaproteomic systems evolve during dynamic disease states, we leveraged a 4.5-year fecal time series using samples from a single patient with colonic Crohn's disease. Utilizing multiplexed quantitative proteomics and shotgun metagenomic sequencing of eight time points in technical triplicate, we quantified over 29,000 protein groups and 110,000 genes and compared them to five protein biomarkers of disease activity. Broad-scale observations were consistent between data types, including overall clustering by principal-coordinate analysis and fluctuations in Gene Ontology terms related to Crohn's disease. Through linear regression, we determined genes and proteins fluctuating in conjunction with inflammatory metrics. We discovered conserved taxonomic differences relevant to Crohn's disease, including a negative association of *Faecalibacterium* and a positive association of *Escherichia* with calprotectin. Despite concordant associations of genera, the specific genes correlated with these metrics were drastically different between metagenomic and metaproteomic data sets. This resulted in the generation of unique functional interpretations dependent on the data type, with metaproteome evidence for previously investigated mechanisms of dysbiosis. An example of one such mechanism was a connection between urease enzymes, amino acid metabolism, and the local inflammation state within the patient. This proof-of-concept approach prompts further investigation of the metaproteome and its relationship with the metagenome in biologically complex systems such as the microbiome.

IMPORTANCE A majority of current microbiome research relies heavily on DNA analysis. However, as the field moves toward understanding the microbial functions related to healthy and disease states, it is critical to evaluate how changes in DNA relate to changes in proteins, which are functional units of the genome. This study tracked the abundance of genes and proteins as they fluctuated during various inflammatory states in a 4.5-year study of a patient with colonic Crohn's disease. Our results indicate that despite a low level of correlation, taxonomic associations were consistent in the two data types. While there was overlap of the data types, several associations were uniquely discovered by analyzing the metaproteome component. This case study provides unique and important insights into the fundamental rela-

Citation Mills RH, Vázquez-Baeza Y, Zhu Q, Jiang L, Gaffney J, Humphrey G, Smarr L, Knight R, Gonzalez DJ. 2019. Evaluating metagenomic prediction of the metaproteome in a 4.5-year study of a patient with Crohn's disease. mSystems 4:e00337-18. https://doi.org/10.1128/mSystems.00337-18.

Editor Marcus J. Claesson, University College Cork

Copyright © 2019 Mills et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Rob Knight, robknight@ucsd.edu, or David J. Gonzalez, djgonzalez@ucsd.edu.

By integrating metagenomic and metaproteomic data over a 4.5-year Crohn's time series, researchers were able to determine fundamental relationships between microbial systems while in flux.

Received 19 December 2018
Accepted 17 January 2019
Published 12 February 2019

tionship between the genes and proteins of a single individual's fecal microbiome associated with clinical consequences.

KEYWORDS colonic Crohn's disease, inflammatory bowel disease, metagenomics, metaproteomics, microbiome, multiomics, tandem mass tags, time series, gut inflammation, proteomics

Due to the growing evidence for a connection between microbial communities and human health, exploration of the microbiome has rapidly expanded in the past decade. To date, the primary avenue for studying the microbiome has been through genomic technologies (1–3). These techniques help provide an understanding of what and how abundant the microbial constituents are and can define their associated metabolic potential. However, gene copy numbers are not representative of protein levels due to the complex systems governing when and how much of a given protein should be present (4). Further, RNA expression has been well documented to have limited correlation to protein abundance within many eukaryotes and bacteria (5). These relationships have not been thoroughly investigated in the context of the complex communities inhabiting the human gut microbiome, thus limiting the utility of DNA-based (or even RNA-based) analyses for understanding microbiome function.

Metaproteomics is an emerging technique that directly characterizes proteins from multispecies matrices. There has been over a decade of development of the field (6–9), though most studies have been limited in scope due in part to complex technical hurdles, including a lack of proteome coverage (8), sample sizes typically below 20 samples (9), limited reference database selection (10–12), and peptide assignment to proteins of similar identity (10). The introduction of new methods and instruments for use in mass spectrometry (MS) has dramatically increased the number of quantifiable peptides and proteins, allowing a greater-than-20-fold increased coverage of the metaproteome in the past few years (8, 13). Here, we leveraged tandem mass tag (TMT) technology, allowing higher throughput by combining up to 11 samples within one MS experiment, without the necessity of culturing (14). In addition, TMT workflows utilize synchronous precursor selection (SPS) and liquid chromatography-tandem mass spectrometry/triple-stage MS ($LCMS^2/MS^3$)-based quantitation workflow to increase accuracy and reduce the sparsity associated with label-free proteomics (15). This combination has enabled unprecedentedly deep characterization of proteomes at large scales (16–18). In comparison to current metagenomic technology, the metaproteome field is still limited in depth of coverage and throughput. Nevertheless, performing direct protein-level analysis through advances in MS may provide new insights into complex biological systems.

Here we utilized these technical advances to better understand the relationship between fluctuations in microbiome protein expression and fluctuations in microbiome gene content. Crohn's disease (CD), a subtype of inflammatory bowel disease (IBD), represents a chronic autoimmune condition associated with large fluctuations in the microbiome (19–22). A study published in 2012 was the first to integrate the metagenome and metaproteome in the context of IBD (23). The results indicated that in six Crohn's disease patients, ileal Crohn's disease (ICD) had a unique metaproteome distinct from that associated with colonic Crohn's disease (CCD) (23). Subsequently, a meta-analysis of human single nucleotide polymorphisms from 30,000 IBD patients corroborated the split between ICD and CCD (24). While further metaproteome studies have been conducted on the human gut microbiome of IBD (13, 25, 26), few have integrated and compared results from metagenome and metaproteome data.

A distinguishable aspect of our study is a shift from contrasting IBD cohorts and healthy subjects to exploring a time series perspective from a single patient. Previous studies investigated metaproteome stability in the context of healthy subjects (27, 28); however, those studies were limited to time periods at or below 1 year. Here, we tracked the disease activity of our patient through the abundances of several subcomponents of the immune system which form the basis of several clinical tests used to

TABLE 1 Roles of immunological proteins of interest[a]

Protein	Role
CRP	An acute-phase response protein produced by the liver upon stimulation by IL-6, TNF-α, and IL-1-β and a common clinical marker of general inflammation (32); it is found in both human blood serum and stool
Lysozyme	A glycoside hydrolase used in the innate immune system for hydrolysis of cell walls of Gram-positive bacteria (84); measurements of lysozyme in the stool of patients with IBD have shown some correlation to disease activity in colonic IBD (84)
Secretory IgA	The most abundant antibody in the human colon; helps tightly control the relationship between commensal microbes and the host by delaying or abolishing the ability of microbes to adhere to the epithelium (49)
Calprotectin	An antimicrobial protein that sequesters manganese to prevent the growth of pathogenic microbes that require these metals (85); consisting of two subunits, S100A8 and S100A9, calprotectin is a molecule that is important to the innate immune system, constituting 40% of the cytoplasmic proteins in neutrophils; fecal calprotectin levels have been described as a stronger indicator of endoscopic activity than CRP levels, and its presence has potential for identifying endoscopic remission (29, 31, 50)
Lactoferrin	An antimicrobial glycoprotein and a major component of the secondary granules of neutrophils (50), the antimicrobial properties of lactoferrin represent the result of iron sequestration and have potential for both discriminatory and activity tests in the clinic (31, 50)

[a]IL-6, interleukin-6; TNF-α, tumor necrosis factor alpha.

monitor IBD disease activity (29–32). These proteins include C-reactive protein (CRP), lysozyme, secretory immunoglobulin A (S-IgA), calprotectin, and lactoferrin (Table 1). Our experimental design includes one patient and eight time points, with a focus on the comparisons between metagenomic and metaproteomic data. By tracking IBD episodic dynamics through the metagenome and metaproteome, we identified a set of bacterial taxa and a set of functional groups that were found to be time-correlated with immunological biomarkers in our patient. Further, we evaluated metagenomic prediction of the metaproteome and identified unique aspects of function accessible through metaproteomics.

RESULTS

Patient information. The $n = 1$ patient was a nonsmoker male. He was diagnosed in 2011, at age 63, with CCD by William J. Sandborn at the University of California Health System. The inflamed region of the colon was determined, via colonoscopy and abdominal magnetic resonance imaging (MRI) analysis, to be confined to 6″ to 8″ of the sigmoid colon. Specifically, a 2012 colonoscopy revealed that this region had extensive diverticulosis and inflammatory focal ulceration, inflammatory pseudopolyps, and patchy friability not associated with the diverticular orifices. During the time interval covered in this work (28 December 2011 to 22 May 2016), the patient had one period of antibiotic therapy, which consisted of ciprofloxacin 500 mg administered twice daily and metronidazole 250 mg administered three times daily for 1 month starting 31 January 2012. During that period, the patient was also taking 40 mg prednisone daily. In another 4-month period from August through November 2013, the patient had simultaneous courses of mesalamine (Lialda; anti-inflammatory) and budesonide (Uceris) administered at 9 mg daily. During the reported period, the patient had episodic symptoms of rectal bleeding, abdominal cramps, bloating, and malaise. Lastly, there was no surgery performed on the patient during the time period covered by this work.

Selection of immunological proteins of interest. The immunological proteins fecal C-reactive protein (CRP), lysozyme, S-IgA, calprotectin, and lactoferrin were selected for their unique properties and clinical applications in IBD. We observed similar expression patterns over time for calprotectin, lactoferrin, and S-IgA (Fig. 1a). Lactoferrin and S-IgA abundances were the most strongly correlated to calprotectin (Pearson $r = 0.96$ and 0.50, respectively), which led to overlapping results in downstream analysis. Because calprotectin is more widely used for the assessment of IBD (29), we focused primarily on the relationships found with calprotectin rather than on those found with lactoferrin and S-IgA.

Technical comparisons between –omic types and protein database methodology. As discussed above, eight fecal samples from our patient representing a wide range of disease activity were collected over a time period from 2011 to 2016. Samples

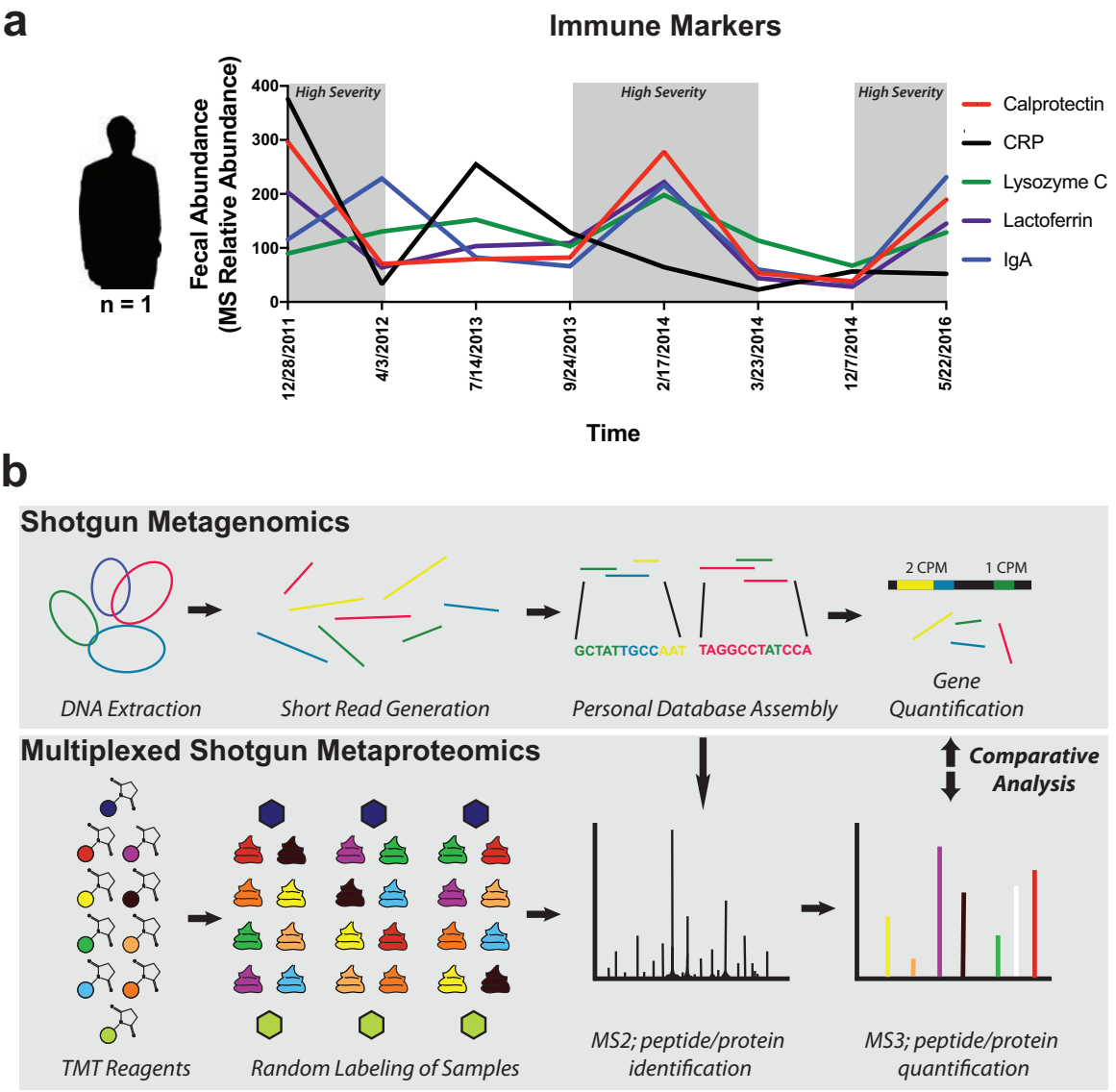


FIG 1 Study design. (a) Immune markers associated with samples. Mass-spectrometry-based relative abundances of fecal calprotectin, CRP, lysozyme, lactoferrin, and secretory IgA are plotted as indicated on the left *y* axis for each of the eight time points in this study. (b) Workflow schematic describing omic methods. Shotgun sequencing and metaproteomic methods were performed in parallel for the analysis of eight selected samples. Both methods were performed in technical triplicate for evaluation of technical variability. Tandem mass tag (TMT) labeling of tryptic peptides was performed for three mass spectrometry experiments. Green and dark blue hexagons represent composite samples used as controls, while other colors represent the random labeling of samples using the remaining TMT reagents. Shotgun sequencing reads were combined and assembled into a shared reference database (Personal Database Assembly) for assigning gene counts (in counts per million [cpm]) and protein abundances. Data corresponding to MS1, which was used for precursor selection, are not depicted.

were processed in technical triplicate through the use of shotgun metagenomic sequencing and a proteomic workflow using TMT-mediated liquid chromatography triple-stage MS (LC-MS^3) (Fig. 1b).

To address the lack of a standardized database methodology (10, 11), two different protein reference database approaches were used for analysis of LC-MS^3 data. Our first approach utilized the shotgun metagenomic reads generated within the study to create a personalized database (pDB) containing 1.3 million protein-coding regions (23). Through alignment of our protein-coding regions to taxonomic and functional databases, the pDB provided genus-level annotations for 80% of the genes and functional annotations to KEGG orthologous (KO) groups for 15% of genes. The pDB approach was crucial for comparison between metagenomic and metaproteomic data

as it provided a shared reference for gene and protein abundances. For comparison, we separately performed a two-step method (12) for searches of the MS data using a public database of gut microbial genes (the Integrated Gene Catalog [IGC]) (33). Our methods resulted in 123,806 predicted open reading frames (ORFs) from the pDB with DNA quantification and 29,370 with protein quantification (see Fig. S1a in the supplemental material). A search through both databases yielded similar numbers of peptides and proteins, with a total of 113,373 unique peptides and 72.5% of peptides shared between the pDB and the IGC database methodology (Fig. S1b). The degree of overlap in peptides was consistent with previous findings (12).

Notably, a lack of sequences shared between samples is a known trait of microbiome studies (34). We observed that the TMT-based metaproteomic methods provided quantification measurements within all samples for larger percentages of proteins (52% of proteins identified from the pDB and 65% of proteins identified from the IGC) than the metagenomic techniques provided for gene quantifications (4%) (Fig. S1a). This increased overlap was likely a result of the TMT multiplexing methods, which are known to reduce sparsity in comparison to label-free MS (35). Our methods also enabled parallel quantifications of nearly 1,000 human proteins (Fig. S1a). Human protein quantification is an important advantage of metaproteomics, especially in light of recent results showing the ability of human proteins to distinguish IBD patients from controls (13). Note that the use of different databases for protein assignment can result in different functional annotations. For example, we observed that the IGC approach identified 83% more unique KEGG orthologous (KO) groups than the pDB approach (Fig. S1c). This discrepancy in peptide matching is an ongoing area of investigation in computational biology (10–12, 36).

The technical and biological variability within each data set was assessed through principal-coordinate analysis (PCoA) using the Bray-Curtis distance metric (37). To overcome the problem of the presence of structural artifacts from the missing values within TMT experiments, only the proteins common to all samples were used in this analysis. After this adjustment, a comparison between our data sets was performed using Procrustes analysis and a Mantel test (Fig. 2a). The Procrustes analysis transforms two distance matrices from corresponding samples to compare distributions. These tests showed minimal technical variability and a strong association between the two data types (Mantel test $P < 0.001$). We also observed clustering based on the presence of a state of high or low inflammation (Fig. 2a). Group differences between high- and low-inflammation states were not statistically significant, likely a result of the small number of samples analyzed. Though the data were not significant, the metaproteome showed a stronger association with the inflammation state than the metagenome (pseudo-F = 1.54 for metaproteome, pseudo-F = 1.19 for metagenome) (Fig. S1d).

To investigate the relationship between gene-level and protein-level fluctuations, the data were subsetted to the 3,598 ORFs with quantitation in both the metagenome and metaproteome. Spearman correlations between the protein and gene abundances in each of the samples were assessed. Overall, the Spearman correlations were normally distributed around $\rho = 0.317$ (Fig. 2b). This limited correlation highlights the added value that a metaproteomic approach can present in cases such as CD, where disease severity is associated with fluctuations in the microbiome (19). We next investigated comparisons of data types from a functional perspective by summing abundances by Gene Ontology (GO) and KO annotations and performing Spearman analyses of correlations between the genes and the protein abundances. This analysis resulted in an approximately normal distribution near $\rho = 0.140$ for both annotation types (Fig. S1e to f). These weak correlations might have been expected given that our approach was based on comparing DNA to protein, as even RNA abundances are often weakly correlated to protein abundance (38).

We further investigated differences in data types by comparing the distributions of dynamic ranges and standard deviations. Ratios of maxima to minima showed that both data types demonstrated a normal distribution centered around 4.4 for proteins and 11 for gene copy numbers (Fig. 2c). The maximum-to-minimum ratios reached

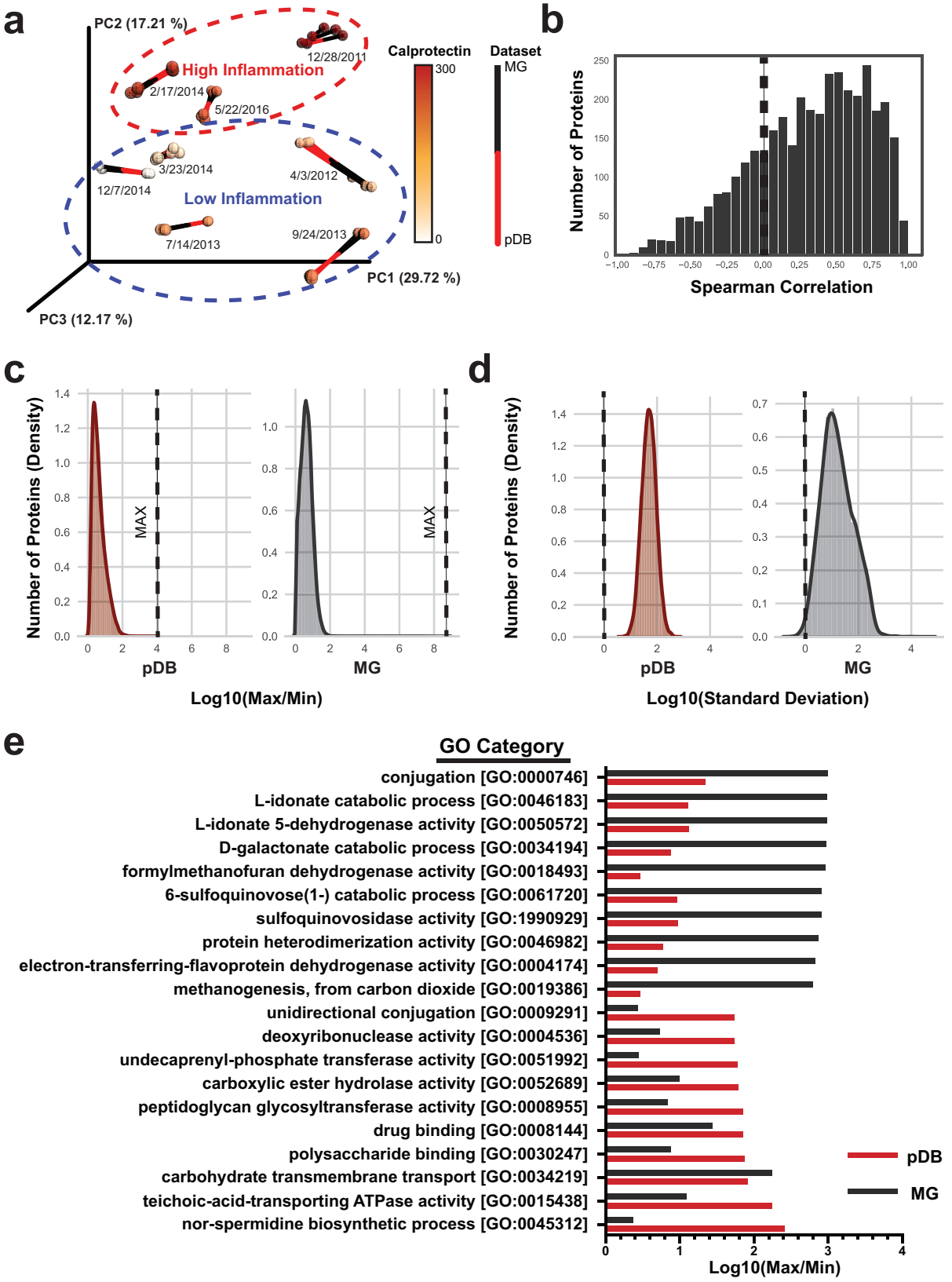


FIG 2 Broad-scale data type comparisons. (a) Procrustes analysis comparing clustering of the metaproteome to that of the metagenome. Bray-Curtis distance metric was used on both the metagenome and the metaproteome (only proteins common to all samples; pDB database) to assess technical and biological variability within and between data sets. Samples are colored according to calprotectin relative abundances. (b) Distribution of Spearman correlations comparing metagenomic and metaproteomic fluctuations. The *x* axis displays Spearman correlation (ρ)

(Continued on next page)

9,400 for proteins and 129 million for gene copy number (Fig. 2c), indicating a much greater dynamic range for the latter. These dynamic ranges may indicate the extent to which microbial genes and proteins can change over time within an individual. However, this result may be influenced by the differences in the depth of coverage, with the metagenome approaching more complete coverage than the metaproteome, and the less-abundant genes detected only by the metagenomic methods may have a greater dynamic range. The standard deviations of the genes and proteins were normally distributed but displayed differences in averages and variances (Fig. 2d). The metagenome had larger variance in the distribution of standard deviations, potentially indicating more variability within that platform (variances of 0.36 and 0.074 for the Microbial Genomes [MG] database and pDB). Still, this result may also be influenced by the differences in the depth of coverage. The values corresponding to maxima to minima for the GO and KO sums shared similar distributions between data types (Fig. S1g and h). The largest fluctuations in GO terms were greater than 100-fold for proteins and 1,000-fold for genes (Fig. 2e). Large changes were observed in categories of interest such as drug binding for proteins and methanogenesis (39) for genes. This was likely the result of the presence and then absence of two archaeal methanogens, *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* (40), whose genes were, on average, 15 times more abundant at the time point of the first collection (28 December 2011) than any other sample. These results give some indication of the fundamental dynamics of genes and proteins but were surely influenced by the techniques used in the study design.

Copy number prediction of protein abundances by functional categories. Because proteins have consistent roles (41), we expected that certain functional categories would show a stronger correlation between gene content and protein expression. We tested this hypothesis using several different functional databases for a comprehensive analysis. After removing human proteins and subdividing individual genes by functional category (evolutionary genealogy of genes: nonsupervised orthologous groups [eggNOG]), the distribution of the data representing gene-to-protein Spearman correlations was largely consistent with the overall mean ρ value of $\sim$0.3 (Fig. 3a). Categories with the largest number of features shared, such as "Energy production and conversion," "Carbohydrate transport and metabolism," and "Translation, ribosomal structure, and biogenesis," all had distributions centered near the Spearman ρ value of $\sim$0.3. Other categories with fewer features had more variability in their average correlation values. Less-abundant categories included "Cell cycle control," which had a lower average correlation, and "Inorganic ion transport and metabolism," which had a higher average correlation (Fig. 3a; see also Table S1 in the supplemental material). This indicates that there were no broad-scale functional group differences distinguishable from the overall low but positive correlation observed between all genes and proteins.

In addition to individual gene correlations, we also evaluated inter-omic relationships between the abundances of entire gene categories. We assessed these relationships through summing protein and gene abundances by GO annotation and performing Spearman correlations (Table S2). There was large variability ($\sigma = 0.445$) in the correlations of different functional groupings with an average Spearman ρ value of 0.135. Despite the low overall correlation, themes of GO categories with similar correlations were present. Several GO terms related to polysaccharide, formate, and anaerobic respiration all had strong positive correlations above $\rho = 0.6$ (Fig. 3b). Other categories had consistently low or even negative correlations below $\rho = 0.2$. Cell wall

FIG 2 Legend (Continued)
data, and the *y* axis displays the number of gene-protein pairs within a range of Spearman correlation values. (c) Dynamic range comparison. Histograms fitted with a Gaussian kernel density estimate are displayed at the gene and protein levels. The log 10 values representing the maximum value for each protein or gene divided by the minimum value are plotted on the *x* axis. The numbers of proteins corresponding to each maximum/minimum (Max/Min) range are plotted on the *y* axis. (d) Variability comparison. The analyses were performed as described for panel c but according to the standard deviation of each gene or protein. (e) GO categories with the largest fluctuations. Proteins and genes were summed according to their GO categories, and the maximum values were compared to the minimum values. The highest metagenomic fluctuations for each category are recorded at the top, and the highest metaproteomic fluctuations are displayed at the bottom.

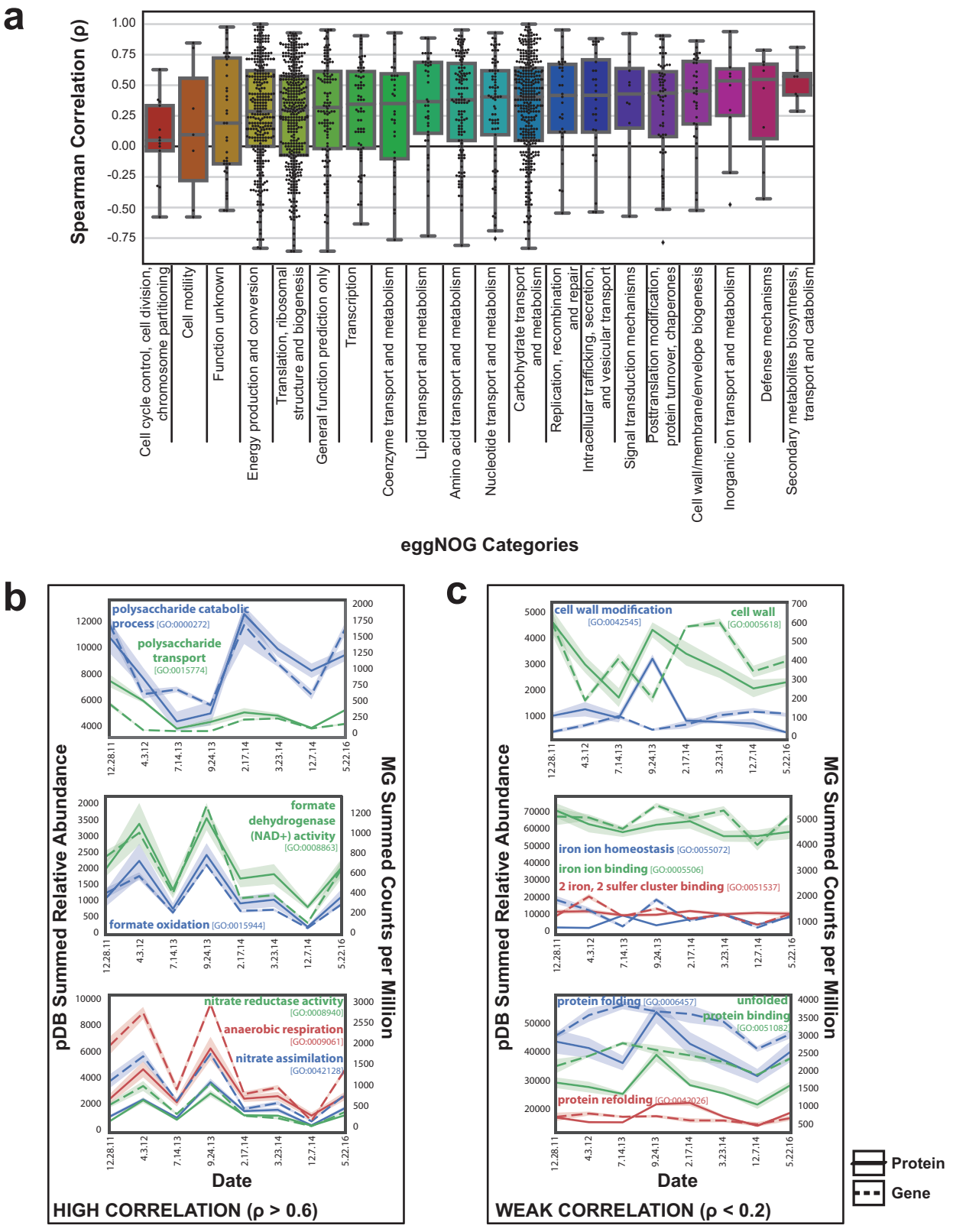


FIG 3 Functional categories with strong or weak genomic prediction of proteome fluctuation. (a) Box plot demonstrating the distribution of Spearman correlations for each gene with an associated eggNOG functional category. The Spearman correlation (ρ) between the summed metagenomic counts per million per time point and the average relative abundance of associated metaproteomic protein is displayed. Summary statistics for these data can be found in Table S1. (b) Summed GO categories with strong genomic and proteomic correlation. (c) Summed GO categories with weak genomic and proteomic correlation.

and membrane proteins and metal binding proteins and chaperones were among the categories with poor correlations (Fig. 3c). These results suggest that there are some categories of genes that better represent protein expression levels, which may be the result of constitutive versus inducible expression. However, the techniques used also influence particular categories, such as membrane proteins, whose hydrophobic nature presents a challenge to MS workflows (42). All of the described categories had greater than 200 proteins and genes contributing to these relationships, which indicates that the findings was not related to differences based on the presence of high- or low-abundance proteins.

Taxonomic correlations with inflammatory markers are largely shared at the protein and gene levels. We next sought to determine whether fluctuations related to inflammatory markers were conserved between genes and proteins. Taxonomic assignments for the pDB database were assigned based on the protein sequences to ensure consistent assignments for both data sets. Genus-level compositions were significantly different in the metagenome but not in the metaproteome (Friedman test $P = 8.9e{-}5$ and 0.69, respectively) (Fig. 4a and b). Dominant genera included *Escherichia*, *Bacteroides*, *Faecalibacterium*, and *Alistipes* (Fig. 4a). For easier interpretation of the abundances used for metagenome comparisons, the metaproteome composition was intentionally not adjusted for the lowest common ancestor of the peptides (43). Metaproteome taxonomic composition plots adjusted for lowest common ancestor also displayed stable compositions, though certain genera, such as *Blautia*, had a notably different composition after the adjustment (Fig. S2a-b).

To evaluate the relationship between species related to inflammation in CD and our biomarkers of interest, we evaluated each immune protein against a previously defined microbial dysbiosis index (19). This index was developed using hundreds of samples from both Crohn's disease patients and healthy controls to predict CD severity through analysis of log ratios of the species that were increased and decreased in abundance within CD (19). Nineteen of the species defined in the index were found in our data set. These included *Escherichia coli* and *Fusobacterium nucleatum*, which are increased in abundance in CD, and *Faecalibacterium prausnitzii*, *Eubacterium rectale*, and *Bacteroides vulgatus*, which are decreased in abundance in CD. After summing gene and protein abundances and determining the relationship between log ratios and each biomarker, fecal calprotectin was found to have the strongest association with the microbial dysbiosis index in both the metagenome and metaproteome. This result was not statistically significant, which was likely a result either of the small sample size or of the extrapolation of methods developed from hundreds of patients for use with a single subject (Fig. S2c).

Linear regression analyses were performed against inflammatory markers on each gene and protein. To evaluate our results, we compared the positively and negatively associated genes with large effect sizes (44) (correlation coefficient, $|r| > 0.7$). Interestingly, most of the individual genes and proteins associated with each of the inflammatory markers were unique, with only 0.5% (188/34,836) of associations shared between data types (Fig. 4c). Accounting for only the genes and proteins quantified in both data sets, 10% (188/1,814) of the strong associations were shared between data sets (Fig. S3).

Despite the lack of overlap in the individual identities of the genes and proteins correlated with each clinical marker, we observed consistent trends in the taxonomic annotations among the correlated genes and proteins. With over 800 genes and proteins strongly correlated to each marker ($|r| > 0.7$), we contrasted the taxonomic compositions of the positive and negative correlations. Several genera had >30-fold differences between compositions (Fig. 4d). Genus-level differences were largely conserved between data types in both direction and magnitude of association (Fig. 4d). *Akkermansia* and *Anaerostipes* had the strongest proinflammatory relationship whereas *Faecalibacterium* and *Butyricicoccus* had the largest anti-inflammatory relationship as assessed through the number of proteins positively or negatively correlated to calprotectin (Fig. 4d). Several genus-level trends such as those corresponding to *Alistipes*,

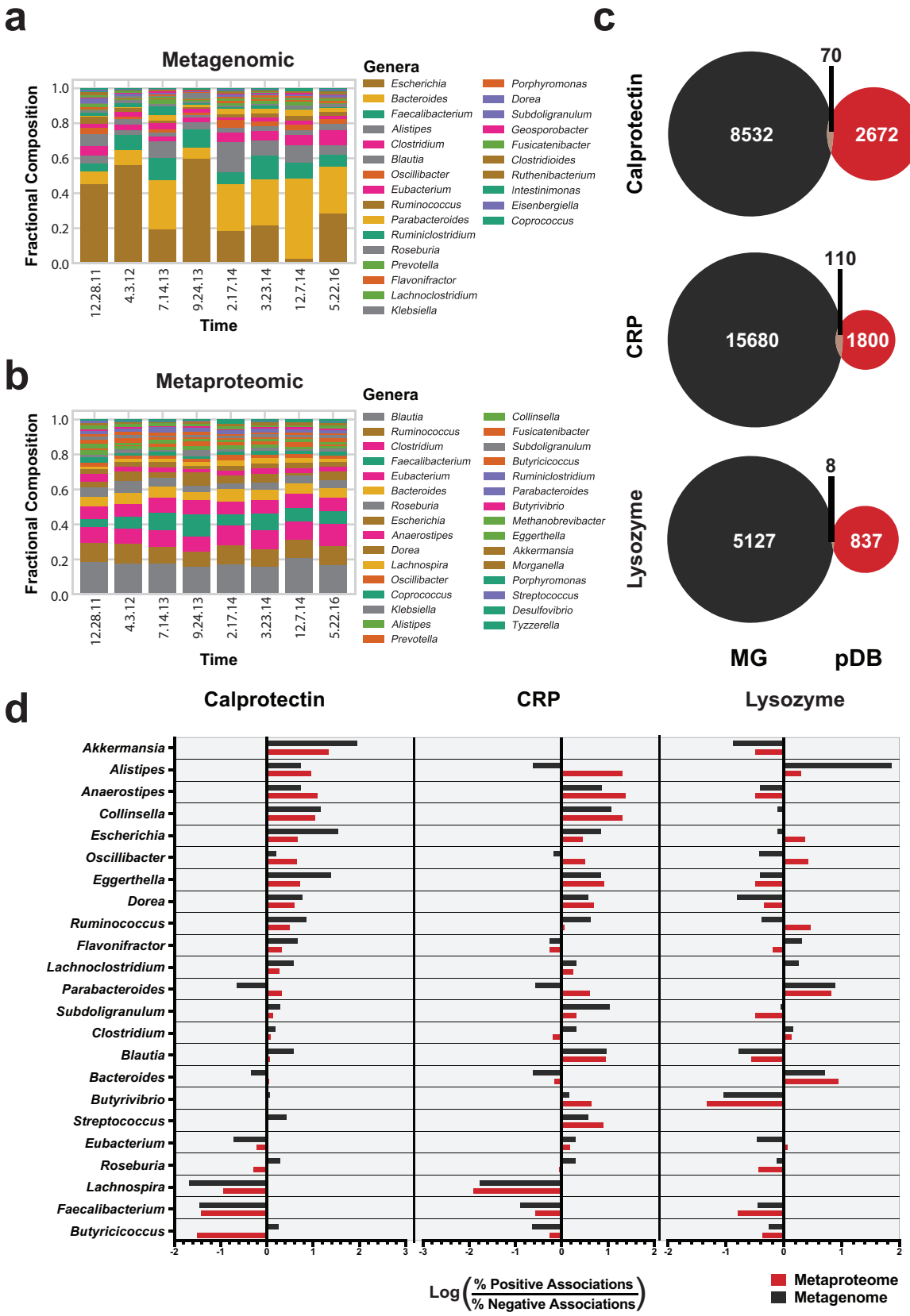


FIG 4 Genus-level associations with clinical markers. (a and b) Bar plot displaying the fractional composition of the most abundant genera (>0.03) in the metagenome (a) and the metaproteome (b) in each of the samples analyzed. (c) Comparison of genes and proteins significantly associated with each clinical marker. Venn diagrams show the number of genes and proteins with a large effect size ($|r| > 0.7$)

(Continued on next page)

Anaerostipes, *Faecalibacterium*, and *Lachnospira* were conserved between CRP and calprotectin, while lysozyme had largely different associated genera. Contextually, the number of proteins and genes used to generate these associations is important for the interpretation of these results as some associations were based on very few observations (Table S3).

Lysozyme is a component of the innate immune response that targets Gram-positive cell walls. Interestingly, proteins and genes correlated with lysozyme levels had large phylum-level changes (Fig. S4a). *Bacteroidetes* is a Gram-negative phylum, while *Firmicutes* is largely a Gram-positive phylum (45). The Gram-positive *Firmicutes* were enriched 1.4-fold among negative associations with lysozyme in both gene and proteins, while the Gram-negative *Bacteroidetes* were enriched 4.3-fold and 8.9-fold among positively correlated proteins and genes, respectively (Fig. S4a). Even though there were more than 800 genes and proteins from *Firmicutes* and *Bacteroidetes* that were correlated to lysozyme, very few from other phyla, such as the Gram-negative *Proteobacteria* and Gram-positive *Actinobacteria*, were observed. To validate these observations at the genus level, Gram staining information was cross-referenced (46). Although there were genera with both Gram-negative and Gram-positive species, the genus-level associations with lysozyme largely reflected the phylum-level observations (Fig. S4b).

Comparing functional interpretations of the genes and proteins associated with immunological biomarkers. Using the same identifications from linear regressions that provided the genus-level results, we next compared broad-scale functional groupings. The broad-scale functional associations were weaker than the genus associations. This observation may represent the effects of broad-scale categorization versus fine-scale categorization. Illustrating this point, the largest difference among the associations of genera was 90-fold, while the largest difference between functional groupings using assignments to the eggNOG database was 12-fold (Fig. 4d; see also Fig. 5a). Analyzing a broader taxonomic category, we observed that the maximum difference among comparisons of phyla was 8.9-fold (Fig. S4a), considerably closer to the 12-fold maximum for eggNOG categories. An additional consideration with respect to this result is the annotation rate for functional assignments. Only 15% of observed ORFs had an identifiable function, and this lower annotation rate may bias the results.

Despite the weaker associations of functional categories, several functional relationships with the disease markers were of interest. In total, 19 eggNOG categories (12 from the metaproteome, 7 from the metagenome) had differences of 3-fold or greater (Fig. 5a). Comparing the categories with associations with different immune markers provided insight into how different data types might influence functional interpretation. For example, metagenomic data had several strong functional associations that were not confirmed by protein abundances. One such category, "Nucleotide transport and metabolism," had 147 genes positively correlated with CRP and 0 genes negatively correlated, indicating a positive association with CRP. The metaproteome data for this category had almost no association with CRP (Fig. 5a), with 6 proteins negatively correlated and 38 proteins positively correlated. We suspect that nucleotide metabolism undergoes protein expression in a manner independent of inflammatory conditions. The underlying reasons for this observation need to be further investigated.

Biologically relevant relationships were observed in the metaproteome that were not detectable in the metagenome. Free amino acids and urease enzymes have previously been associated with gut dysbiosis and Crohn's disease (47). Interestingly, the metaproteome data identified a functional association of amino acid metabolism proteins with calprotectin, while this observation was absent in the metagenomic data

FIG 4 Legend (Continued)
with respect to clinical markers based on linear regression. (d) Genera associated with clinical markers. The associated proteins with genus-level taxonomy analyzed as described for panel c were compared by determining the log ratios of the compositions of proteins with positive and negative associations. The log ratio is plotted on the *x* axis for each clinical marker, and bars represent the association with each genus. Metaproteome values are plotted in red, and metagenome values are plotted in black. The numbers of genes and proteins included in this analysis are listed in Table S2.

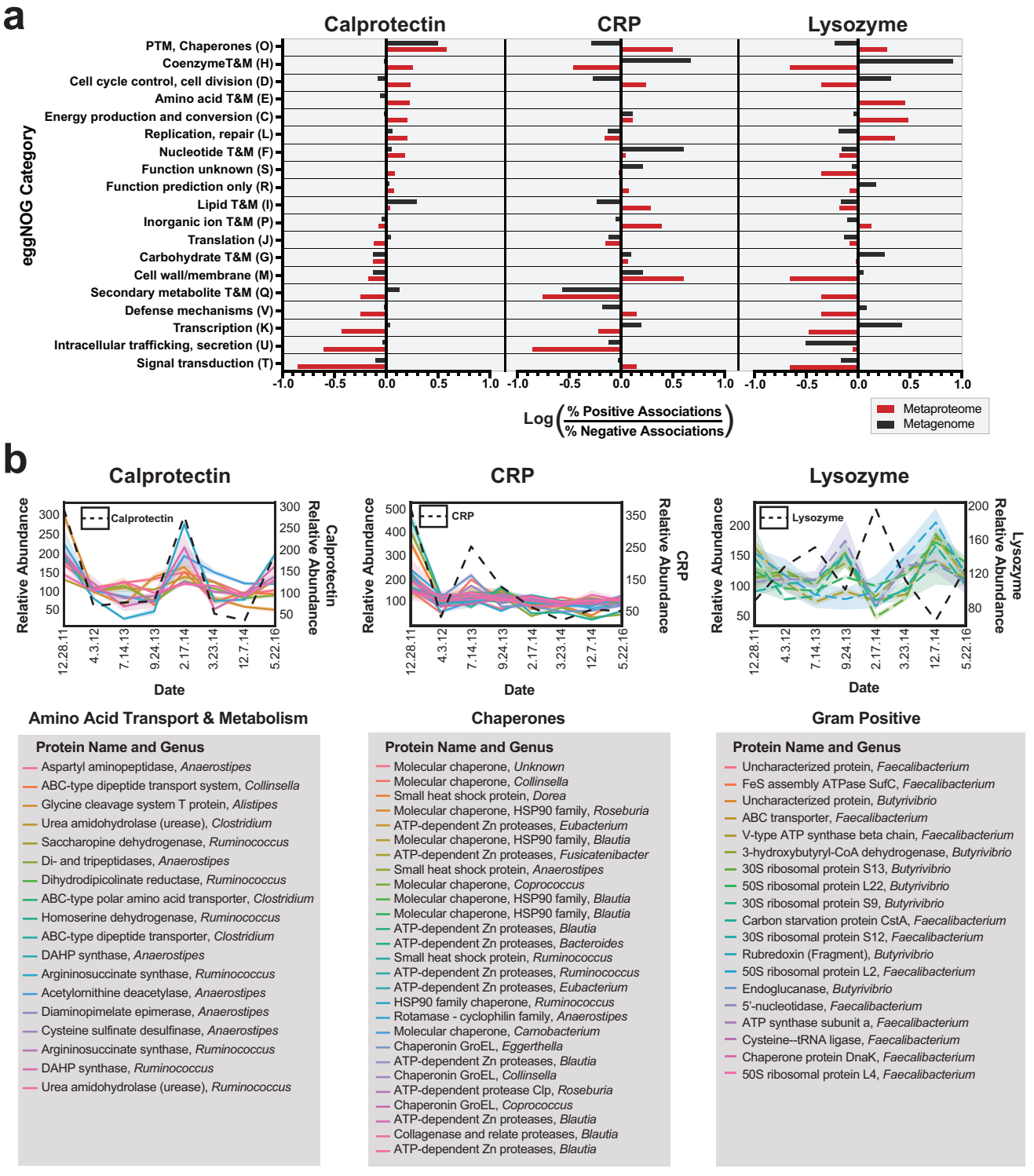


FIG 5 Functional associations with clinical markers. (a) Functions associated with clinical markers. Linear regressions to clinical markers were performed and the number of proteins or genes derived from each functional group with a large effect size ($|r| > 0.7$) were compared. The log ratio of the composition of positive and negative proteins is plotted on the *x* axis for each clinical marker. Metaproteome values are plotted in red and metagenome values are plotted in black. PTM, posttranslational modification; T&M, transport and metabolism. (b) Time series plots of selected proteins of interest. Protein abundances of one finding from each clinical marker are shown. A legend describing the protein names and associated genera is shown below each graph. DAHP synthase, 3-deoxy-D-arabinoheptulosonate 7-phosphate synthase.

(Fig. 5a). This observation included several urease proteins, as well as transporters for free amino acids, many of which were derived from the genera that had positive associations with inflammation (Fig. 5b). These ureases and transporters thus represent interesting targets for further investigation and represent further evidence of a previously established connection (47).

Another observation that was exclusively related to the metaproteome data was the relationship of chaperone proteins to several of the inflammatory metrics. There were 15 chaperone proteins with similar trends in expression with respect to CRP (Fig. 5b). This corresponded to posttranslational modification and chaperone proteins having 3.2-fold-higher representation in positively associated proteins and 1.9-fold-lower representation in genes (Fig. 5a). This unique observation from our patient's fecal metaproteome is a potential indication of microbial stress occurring in response to the acute phase response and may indicate a need for the microbiome to refold proteins.

Because lysozyme targets Gram-positive cell walls, we expected correlated genes and proteins to be influenced by taxonomy and to have functions related to cell walls or membranes. However, cell wall proteins were underrepresented in the metaproteomic data set relative to their occurrence in the metagenomic data set (Table S4). Of the cell wall proteins associated with lysozyme, two (COG1088 and COG0463) were related to cell wall biosynthesis, encoding a glycosyl transferase and a dTDP-glucose 4-6-dehydratase. In this case, the binding of lysozyme to peptidoglycan may have disrupted the binding of these cell wall/membrane/envelope biogenesis proteins, leading to the observed negative correlation. Even though we were not able to detect many membrane or cell wall proteins related to lysozyme, 19 negatively correlated proteins from the butyrate-producing (48), Gram-positive genera *Faecalibacterium* and *Butyrivibrio* were identified (Fig. 5b). These proteins included 6 ribosomal proteins, which may indicate decreased translation occurring in the presence of lysozyme.

In addition to analyzing calprotectin, CRP, and lysozyme levels, we also evaluated S-IgA and lactoferrin levels. Secretory IgA is secreted in large quantities in the intestine to maintain favorable microbial compositions (49), and lactoferrin sequesters iron as an antimicrobial response (50). We observed similar expression patterns of lactoferrin, S-IgA, and calprotectin (Fig. S5a). The similar expression patterns resulted in minimal differences in both genus and functional relationships between calprotectin, lactoferrin, and S-IgA (Fig. S5b and c). Proteins positively associated with lactoferrin ($|r| = >0.7$) had a larger portion of GO terms related to iron (15.5% of 470 positive associations and 10% of 233 negative associations). Many of these proteins were pyruvate oxidoreductases, which are used in anaerobic bacteria for forming acetyl-coenzyme A (acetyl-CoA) from pyruvate (51) (Fig. S5d). These are crucial enzymes for certain anaerobic bacteria and have been suggested as potential drug targets (51). This result suggests that a connection exists between the iron-sequestering host proteins and the microbial proteins in our patient that are dependent on iron as a cofactor.

DISCUSSION

Our investigation of the fundamental relationship between changes in the metagenome and the metaproteome revealed important considerations for interpreting these data types. Currently, studies using shotgun metagenomics to dissect the functions of the microbiome are becoming more prevalent (52), and the current study showed that differences at the gene level may not reflect differences at the protein level. Though discordance between RNA and protein expression is widely acknowledged for individual species (4), the relationships between DNA and protein content in the complex ecology of the microbiome are less understood. As these systems have rarely been studied in parallel, it is possible that communities of microbes influence the fundamental relationships between genes and proteins that had been previously established in monoculture settings. Although the metaproteomics field is improving in depth of coverage (8) and scope (13), the technical hurdles that MS presents often make DNA-based studies a more practical, higher-throughput solution. That being the

case, functional insight from metagenomic studies requires a consideration of the relationship between protein abundances and metagenomic copy numbers.

Our results, although limited to a single patient, suggest that there is a degree of general agreement between changes in the metagenome and changes in the metaproteome. However, the relationship for individual genes/proteins is weak overall (our average Spearman $\rho = 0.3$). In the single-species context, bacterial systems have generally shown correlations between mRNA and proteins to range from $\rho = 0.5$ to 0.6 (38). Our experimental estimates indicate that DNA-to-protein correlations in complex microbial systems are notably lower. These associations do not appear to have obvious biases between large-scale functional groupings but do show certain trends in finer-resolution functional groupings such as individual GO terms. Representing an important notion in the field of IBD, formate- and nitrate-related categories had large fluctuations and consistent trends between the two data types. Formate oxidation has been implicated as a metabolic signature of inflammation-associated dysbiosis (53), indicating that metagenomic studies may predict protein abundances within this system. We do not believe that the consistency of the relationship between formate oxidation genes and proteins is a result of constitutive expression, as, at least in *E. coli*, related genes such as formate hydrogenase genes are regulated by the presence of formate (54). Nitrate-based anaerobic respiration is implicated in promoting the growth of facultative anaerobes such as the *Enterobacteriaceae*, which can lead to microbial dysbiosis and intestinal inflammation (45). Tables of the identified eggNOG and GO terms are provided and indicate how well the metagenomic copy number predicted the protein abundances within each identified category.

Identifying the genes and proteins with similar expression trends with respect to certain inflammatory and immune markers revealed that there were large differences in genus-level associations that were biologically relevant and generally consistent between data types. *Faecalibacterium* is a genus depleted in IBD (19, 55) and appears to have anti-inflammatory effects, possibly mediated by butyrate production (56). Both data types had a strong negative correlation in numerous *Faecalibacterium* proteins to our biomarker for local inflammation, calprotectin. While it was previously shown that there were consistent trends between these data types showing increased *Faecalibacterium* in healthy patients (23), our results show these relationships can occur within a patient through time in a manner that corresponds to the current level of inflammation. Other trends were also found for well-documented genera with inflammatory roles in IBD (19), including *E. coli*, which is of particular interest because of its adherent-invasive properties in CD (57, 58). Interestingly, these shared trends were found with almost entirely different genes. This may indicate that the underlying bacterial abundance influences both of these data types but that the individual proteins expressed at certain times are not directly associated with the amount of corresponding genetic material present. If this is the case, it is possible that functional associations made through some broad-scale categories, such as eggNOG, may have different results depending on the data type. This concept is supported by our results that indicate less-extensive and less-consistent associations with broad-scale groupings than with associations at the genus level.

Our analysis of clinical biomarkers was useful for understanding the biology associated with each immune component. As calprotectin had the strongest association with the microbial dysbiosis index (19), the results suggests that the level of calprotectin may be a better indication of microbial imbalances. Interestingly, CRP has been reported to represent a less useful diagnostic tool than fecal calprotectin for intestinal inflammation (29). CRP levels may be a better indication of systemic inflammation, and we observed here that the levels of many bacterial chaperone proteins may be increased in correspondence. We observed taxonomic trends with the abundances of lysozyme that were consistent with its biological function of acting upon cell walls. In general, predominately Gram-positive genera and phyla had a larger portion of anti-correlated genes and proteins, while Gram-negative bacteria had an opposite association.

Our observed discrepancies between gene and protein levels may have large implications for data interpretation, but it is important to replicate these results in a larger cohort of IBD patients. As certain GO categories present strong correlations between data types, it suggests that it may be possible to develop a metagenomic-metaproteomic reference guide for creating stronger functional hypotheses. This guide may be used to outline which groups of genes have a strong or weak association with protein abundances.

The relationship between genes and proteins may be influenced by several factors. Correlation between DNA and protein abundances might reflect the presence of DNA from dormant or dead cells (59), which may lead to a higher level of correlation (because the cells are not actively producing or secreting proteins). Other factors may include constitutive versus inducible genes or the stability of the proteins. For example, chaperone proteins were found in high abundance which may be a result of their high stability and of their stable concentrations within the cell (60). Ultimately, the associations between -omic data sets are influenced by the nature of the data collection techniques and normalization, and further benchmarking is necessary. Although, there are significant challenges in integrating multi-omic data types (61), further understanding these relationships is of paramount importance as the microbiome field progresses.

Our study presents several technical findings of interest. Leverage of the modern TMT-based LC-MS^3 quantification platform provided a highly accurate quantification method for comparison with gene counts. Our workflow designed for mediating comparisons between metagenomic and metaproteomic data expands our knowledge of data type differences and acts as a bioinformatic and technological update to previous studies (23). Additionally, the use of technical triplicates validates the reproducibility of these methods and helped increase our confidence in the quantification values at both the metagenomic and metaproteomic levels. However, outside validation from other technological pipelines may be necessary to further understand these biological systems. Our results are also derived from a small number of samples from one patient, and the time points were spread over large time spans. This design provided unique opportunities but limits our interpretation of the data to a single individual.

From a biological perspective, our results provide evidence that certain proteins and genera are correlated or anticorrelated with immunoprotein markers of inflammation. While the taxonomic insights that we observed were conserved between data types, our functional interpretations differed. This personalized perspective also demonstrates the extent of variability occurring within an individual, an important consideration to control for in studies with larger cohorts. Taking the results together, our study investigated the relationships between metagenomic and metaproteomic methods and highlighted important considerations for interpretation of meta-omic data.

MATERIALS AND METHODS

Ethics statement. The patient had stool samples collected by consent under two protocols: HRPP 141853 (American Gut Project) and HRPP 150275 (Evaluating the Human Microbiome). Both protocols were approved by the Human Research Protection Program (HRPP) of the University of California, San Diego. Written informed consent obtained from the patient concerning dissemination and scientific publication of the results is also included in the approved protocols.

Longitudinal sample collection. Naturally passed fecal samples were collected and immediately stored without buffer at −80°C. Eight samples were selected. A personal symptom log entry was generated at the time that each fecal sample was passed. Additionally, the weight and body mass index (BMI) of the patient were determined on the day associated with each sample.

Generation of metagenomic reads. Samples were extracted according to the Earth Microbiome Project (2) protocol using a Qiagen MagAttract PowerSoil DNA kit as previously described (62). Briefly, swabbed fecal material was plated into 96-well PowerBead DNA plates containing garnet beads. DNA extraction was performed once on each of the eight samples according to the manufacturer's instructions, with an additional incubation at 65°C for 10 min following the addition of lysis solution and immediately prior to shaking (Qiagen TissueLyser II; Qiagen catalogue 85300). Magnetic DNA purification was performed using a KingFisher Flex purification system. Then, whole-genome shotgun libraries were made using a Nextera DNA library preparation kit (Illumina, San Diego, CA, USA) and a 1:10 miniaturized-reaction volume. Unique barcodes were used per triplicate totaling 24 metagenomic samples. The

median insert sizes by sample ranged from 183 bp to 366 bp. Libraries were sequenced using Illumina MiSeq paired-end (2 by 250 bp) sequencing, filling a total of one lane.

Processing of metagenomic reads for a shared reference library (pDB). Because typical metagenomics and metaproteomics workflows require a reference database, it was necessary to use a minimal approach to create from scratch a single reference database that could be used for both metagenomics and metaproteomics from the individualized data. All reads from the technical triplicates of each sample were concatenated. Next, the MEGAHIT alignment program (63) was utilized for assembling short reads into larger contigs. Assembled contigs were searched for possible coding regions through the program Prodigal (64). Next, the program Diamond (65) was used for gene alignment to the uniref50 database (66). Finally, the most likely uniref50 entry, determined through bitScore, was used for the functional annotations. KEGG orthology annotations were cross-referenced using GhostKOALA (67). Taxonomic assignments were determined by Diamond alignment (65) to an in-house library of microbial genomes. Taxonomy was assigned from the translated amino acid sequence of each predicted ORF in the pDB. This database was used as a reference database for both mass spectrometry data and sequencing data. Scripts used for data processing are available online (https://github.com/knightlab-analyses/Crohns-MG-MP-Comparisons).

Generating copy numbers of metagenomic genes. The program Salmon (68) was applied to determine the reads present for each gene from the pDB. First, an index was created with Salmon, inputting the pDB fasta file. Next, reads were aligned to this index in quasimapping mode for each of the 24 metagenomic samples. The results were represented in counts per million sequences, with missing values padded as zeroes.

Protein abundances from the shared reference library (pDB). The generation of mass spectra data is described below. Spectral data were searched against the pDB with a concatenated human reference library (https://www.uniprot.org/; accessed 28 November 2016) using Proteome Discoverer 2.1 (Thermo Fisher Scientific). Further data processing is described below.

Protein digestion and TMT labeling. Fecal samples were measured out to ~0.5 g and suspended in 10 ml of ice-cold, sterilized Tris-buffered saline (TBS). Samples were suspended through vortex mixing and homogenized through the use of a blender apparatus. A Steriflip (Millipore) filter (20 μM vacuum) was used to remove particulate from the samples. Cells were pelleted through centrifugation at 4,000 rpm for 10 min. Next, cells were lysed in 2 ml of buffer containing 75 mM NaCl (Sigma), 3% sodium dodecyl sulfate (SDS; Fisher), 1 mM NaF (Sigma), 1 mM beta-glycerophosphate (Sigma), 1 mM sodium orthovanadate (Sigma), 10 mM sodium pyrophosphate (Sigma), 1 mM phenylmethylsulfonyl fluoride (PMSF; Sigma), 1× Complete Mini EDTA free protease inhibitors (Roche), and 50 mM HEPES (Sigma), pH 8.5 (69). An equal volume of 8 M urea–50 mM HEPES (pH 8.5) was added to each sample. Cell lysis was achieved through two 10-s intervals of probe sonication at 25% amplitude. Proteins were then reduced with dithiothreitol (DTT; Sigma), alkylated with iodoacetamide (Sigma), and quenched as previously described (70). Proteins were then precipitated via chloroform-methanol precipitation, and the protein pellets were dried (71). Protein pellets were resuspended in 1 M urea–50 mM HEPES (pH 8.5) and digested overnight at room temperature with LysC (Wako) (72). A second, 6-h digestion was performed using trypsin at 37°C, and the reaction was stopped through addition of 10% trifluoroacetic acid (TFA; Pierce). Samples were then desalted through the use of C_{18} Sep-Paks (Waters) and eluted with 40% and 80% acetonitrile solutions containing 0.5% acetic acid (73). Concentrations of desalted peptides were determined with a bicinchoninic acid (BCA) assay (Thermo Scientific). Aliquots (50 μg) of each sample were dried in a SpeedVac, additional bridge channels consisting of 25 μg from each sample were created, and 50-μg aliquots of this solution were used in duplicate per TMT 10-plex as previously described (16). These bridge channels were used to control for labeling efficiency, interrun variation, mixing errors, and the heterogeneity present in each sample (74). Each sample or bridge channel was resuspended in 30% dry acetonitrile–200 mM HEPES (pH 8.5) for TMT labeling with 7 μl of the appropriate TMT reagent (14). Reagents 126 and 131 (Thermo Scientific) were used to bridge between MS runs. The remaining reagents were used to label samples in random order. Labeling was carried out for 1 h at room temperature and quenched by adding 8 μl of 5% hydroxylamine (Sigma). Labeled samples were acidified by adding 50 μl of 1% TFA. After TMT labeling, the products of the 10-plex experiments were combined, desalted through the use of C_{18} Sep-Paks, and dried using a SpeedVac.

Basic pH reverse-phase liquid chromatography sample fractionation. Sample fractionation was performed by basic pH reverse-phase liquid chromatography with concatenated fractions as previously described (75). Briefly, samples were resuspended in 5% formic acid–5% acetonitrile and separated over a C_{18} column (Thermo Scientific) (4.6 mm by 250 mm) on an Ultimate 3000 high-performance liquid chromatography (HPLC) system fitted with a fraction collector, degasser, and variable-wavelength detector. The separation was performed over a 22% to 35%, 60-min linear gradient of acetonitrile–10 mM ammonium bicarbonate (Fisher) at 0.5 ml/min. The resulting 96 fractions were combined as previously described (75). Fractions were dried under vacuum and resuspended in 5% formic acid–5% acetonitrile and analyzed by liquid chromatography (LC)-MS^2/MS^3 for identification and quantitation.

LC-MS^2/MS^3 for protein identification and quantitation. All LC-MS^2/MS^3 experiments were carried out on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) with an in-line EASY-nLC 1000 instrument (Thermo Fisher Scientific) and a chilled autosampler. Separation and acquisition settings were as previously defined (76).

Proteomic data processing. Data were processed using Proteome Discoverer 2.1 (Thermo Fisher Scientific). MS^2 data were searched against the pDB and Uniprot human database (https://www.uniprot.org/; accessed 28 November 2016). The Sequest searching algorithm (77) was used to align spectra to database peptides. A precursor mass tolerance of 50 ppm (78, 79) and 0.6-Da tolerance were specified for

the MS^2 fragments. Static modification of TMT 10-plex tags on lysine and peptide N termini (+229.162932 Da), carbamidomethylation of cysteines (+57.02146 Da), and variable oxidation of methionine (+15.99492 Da) were included in the search parameters. Raw data were searched at a peptide and protein false-discovery rate (FDR) of 1% using a reverse-database-search strategy (80–82).

TMT reporter ion intensities were extracted from MS^3 spectra for quantitative analysis, and signal-to-noise values were used for quantitation. Additional stringent filtering was used, removing any moderate-confidence peptide spectral matches (PSMs) or ambiguous PSM assignments. Additionally, any peptides with a spectral interference level above 25% were removed, as well as any peptides with an average signal-to-noise ratio of less than 10. In accordance with false discovery rate benchmarking (83), proteins matching only one high-confidence PSM were not removed. As metaproteome data contain a high degree of similarity in levels of identity between proteins, several decisions were made to reduce false assignments. Standardized methods in Proteome Discoverer (Version 2.1) preferentially assign peptides to proteins that had previously had peptides reported. If this does not resolve the assignment, the peptide is assigned to the longest protein. Additionally, a duplicate peptide filter was applied according to the Proteome Discoverer report. Normalization occurred as previously described (76). Briefly, relative abundances are normalized first to the pooled standards for each protein and then to the median signal across the pooled standard. An average of these normalizations was used for the next step. To account for slight differences in the amounts of protein labeled, these values were then normalized to the median of the entire data set and reported as final normalized summed signal-to-noise ratios per protein per sample.

Use of an integrated gene catalog for reference library comparison. The integrated reference catalog was downloaded from http://meta.genomics.cn/meta/home (accessed 22 December 2016). A two-step database search method was utilized (12). Briefly, the full database was used as a first-pass screen. Second, both forward and reverse database identifications were used to create a study-specific database. This database was used to search mass spectrometry data, and identifications were filtered at a 1% FDR for peptides and proteins.

Data analysis. Data analysis was performed in python version 3.5 (https://www.python.org/), and records of the code are available in corresponding Jupyter Notebooks for this project (https://github.com/knightlab-analyses/Crohns-MG-MP-Comparisons). All displayed metaproteomic data were generated using the pDB metaproteomic data unless otherwise specified. Qiime was used for principal-coordinate analysis (37). Spearman correlations were performed through the use of the pandas python package (http://pandas.pydata.org/). Linear regressions were performed on metagenome sums and metaproteome averages against the metaproteome abundances of each of the biomarker abundances. Protein and gene associations were ranked by the associated coefficient of correlation, and taxonomic and functional annotations of the top associated genes and proteins ($|r| < 0.7$) were compared. Linear regressions were performed using the python package scipy (https://www.scipy.org). Friedman tests were also performed through scipy, comparing genus compositions within the metagenome and metaproteome between samples.

Data availability. Proteomic data and supplementary files are available online at https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp (study identifier [ID] MSV000082113). Metagenomic data are available through the European Bioinformatics Institute (EBI) (https://www.ebi.ac.uk/ena) under the study identifier PRJEB28712 (ERP110957).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00337-18.

FIG S1, PDF file, 0.7 MB.
FIG S2, PDF file, 1.1 MB.
FIG S3, PDF file, 0.3 MB.
FIG S4, PDF file, 0.4 MB.
FIG S5, PDF file, 1.6 MB.
TABLE S1, DOCX file, 0.03 MB.
TABLE S2, XLSX file, 2.1 MB.
TABLE S3, DOCX file, 0.03 MB.
TABLE S4, DOCX file, 0.03 MB.

ACKNOWLEDGMENTS

We thank William Sandborn, Tomasz Kosciolek, Jon Sanders, and John Lapek for helpful discussions and editing contributions. Additionally, we thank our anonymous reviewers for the feedback that we have incorporated into our manuscript.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. We express no conflict of interest. We acknowledge the SDSC for donating supercomputer time for data processing.

REFERENCES

1. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower C, Jackson MA, Janssen S, Jeste DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, Metcalf JL, Mohimani H, Montassier E, Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray VG, et al. 2018. American Gut: an open platform for citizen science microbiome research. mSystems 3:e00031-18. https://doi.org/10.1128/mSystems.00031-18.
2. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vazquez-Baeza Y, Gonzalez A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R; Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463. https://doi.org/10.1038/nature24621.
3. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. Nature 449:804–810. https://doi.org/10.1038/nature06244.
4. Pandey A, Mann M. 2000. Proteomics to study genes and genomes. Nature 405:837–846. https://doi.org/10.1038/35015709.
5. Liu Y, Beyer A, Aebersold R. 2016. On the dependency of cellular protein levels on mRNA abundance. Cell 165:535–550. https://doi.org/10.1016/j.cell.2016.03.014.
6. Klaassens ES, de Vos WM, Vaughan EE. 2007. Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. Appl Environ Microbiol 73:1388–1392. https://doi.org/10.1128/AEM.01921-06.
7. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK. 2009. Shotgun metaproteomics of the human distal gut microbiota. ISME J 3:179–189. https://doi.org/10.1038/ismej.2008.108.
8. Zhang X, Chen W, Ning Z, Mayne J, Mack D, Stintzi A, Tian R, Figeys D. 2017. Deep metaproteomics approach for the study of human microbiomes. Anal Chem 89:9407–9415. https://doi.org/10.1021/acs.analchem.7b02224.
9. Kolmeder CA, de Vos WM. 2014. Metaproteomics of our microbiome - developing insight in function and activity in man and model systems. J Proteomics 97:3–16. https://doi.org/10.1016/j.jprot.2013.05.018.
10. Tanca A, Palomba A, Deligios M, Cubeddu T, Fraumene C, Biosa G, Pagnozzi D, Addis MF, Uzzau S. 2013. Evaluating the impact of different sequence databases on metaproteome analysis: insights from a lab-assembled microbial mixture. PLoS One 8:e82981. https://doi.org/10.1371/journal.pone.0082981.
11. Tanca A, Palomba A, Fraumene C, Pagnozzi D, Manghina V, Deligios M, Muth T, Rapp E, Martens L, Addis MF, Uzzau S. 2016. The impact of sequence database choice on metaproteomic results in gut microbiota studies. Microbiome 4:51. https://doi.org/10.1186/s40168-016-0196-8.
12. Zhang X, Ning ZB, Mayne J, Moore JI, Li J, Butcher J, Deeke SA, Chen R, Chiang CK, Wen M, Mack D, Stintzi A, Figeys D. 24 June 2016. MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota. Microbiome https://doi.org/10.1186/s40168-016-0176-z.
13. Zhang X, Deeke SA, Ning Z, Starr AE, Butcher J, Li J, Mayne J, Cheng K, Liao B, Li L, Singleton R, Mack D, Stintzi A, Figeys D. 2018. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. Nat Commun 9:2873. https://doi.org/10.1038/s41467-018-05357-4.
14. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C. 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75:1895–1904. https://doi.org/10.1021/ac0262560.
15. Ting L, Rad R, Gygi SP, Haas W. 2011. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods 8:937–940. https://doi.org/10.1038/nmeth.1714.
16. Lapek JD, Jr, Mills RH, Wozniak JM, Campeau A, Fang RH, Wei X, van de Groep K, Perez-Lopez A, van Sorge NM, Raffatellu M, Knight R, Zhang L, Gonzalez DJ. 2018. Defining host responses during systemic bacterial infection through construction of a murine organ proteome atlas. Cell Syst 6:579–592.e4. https://doi.org/10.1016/j.cels.2018.04.010.
17. Weekes MP, Tomasec P, Huttlin EL, Fielding CA, Nusinow D, Stanton RJ, Wang EC, Aicheler R, Murrell I, Wilkinson GW, Lehner PJ, Gygi SP. 2014. Quantitative temporal viromics: an approach to investigate host-pathogen interaction. Cell 157:1460–1472. https://doi.org/10.1016/j.cell.2014.04.028.
18. Lapek JD, Jr, Greninger P, Morris R, Amzallag A, Pruteanu-Malinici I, Benes CH, Haas W. 2017. Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. Nat Biotechnol 35:983–989. https://doi.org/10.1038/nbt.3955.
19. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, Gonzalez A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. 2014. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15:382–392. https://doi.org/10.1016/j.chom.2014.02.005.
20. Walters WA, Xu Z, Knight R. 2014. Meta-analyses of human gut microbes associated with obesity and IBD. FEBS Lett 588:4223–4233. https://doi.org/10.1016/j.febslet.2014.09.039.
21. Manichanh C, Borruel N, Casellas F, Guarner F. 2012. The gut microbiota in IBD. Nat Rev Gastroenterol Hepatol 9:599–608. https://doi.org/10.1038/nrgastro.2012.152.
22. Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, D'Amato M, Bonfiglio F, McDonald D, Gonzalez A, McClure EE, Dunklebarger MF, Knight R, Jansson JK. 2017. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol 2:17004. https://doi.org/10.1038/nmicrobiol.2017.4.
23. 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. 2012. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PLoS One 7:e49138. https://doi.org/10.1371/journal.pone.0049138.
24. Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, Andersen V, Andrews JM, Annese V, Brand S, Brant SR, Cho JH, Daly MJ, Dubinsky M, Duerr RH, Ferguson LR, Franke A, Gearry RB, Goyette P, Hakonarson H, Halfvarson J, Hov JR, Huang H, Kennedy NA, Kupcinskas L, Lawrance IC, Lee JC, Satsangi J, Schreiber S, Theatre E, van der Meulen-de Jong AE, Weersma RK, Wilson DC, International Inflammatory Bowel Disease Genetics Consortium, Parkes M, Vermeire S, Rioux JD, Mansfield J, Silverberg MS, Radford-Smith G, McGovern DP, Barrett JC, Lees CW. 2016. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet 387:156–167. https://doi.org/10.1016/S0140-6736(15)00465-1.
25. Juste C, Kreil DP, Beauvallet C, Guillot A, Vaca S, Carapito C, Mondot S, Sykacek P, Sokol H, Blon F, Lepercq P, Levenez F, Valot B, Carré W, Loux V, Pons N, David O, Schaeffer B, Lepage P, Martin P, Monnet V, Seksik P, Beaugerie L, Ehrlich SD, Gibrat J-F, Van Dorsselaer A, Doré J. 2014. Bacterial protein signals are associated with Crohn's disease. Gut 63:1566–1577. https://doi.org/10.1136/gutjnl-2012-303786.
26. Presley LL, Ye JX, Li XX, LeBlanc J, Zhang ZP, Ruegger PM, Allard J, McGovern D, Ippoliti A, Roth B, Cui XP, Jeske DR, Elashoff D, Goodglick L, Braun J, Borneman J. 2012. Host-microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal-luminal interface. Inflamm Bowel Dis 18:409–417. https://doi.org/10.1002/ibd.21793.
27. Kolmeder CA, de Been M, Nikkila J, Ritamo I, Matto J, Valmu L, Salojarvi J, Palva A, Salonen A, de Vos WM. 2012. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. PLoS One 7:e29913. https://doi.org/10.1371/journal.pone.0029913.
28. Kolmeder CA, Salojarvi J, Ritari J, de Been M, Raes J, Falony G, Vieira-Silva S, Kekkonen RA, Corthals GL, Palva A, Salonen A, de Vos WM. 2016. Faecal metaproteomic analysis reveals a personalized and stable functional microbiome and limited effects of a probiotic inter-

vention in adults. PLoS One 11:e0153294. https://doi.org/10.1371/journal.pone.0153294.
29. Chang S, Malter L, Hudesman D. 2015. Disease monitoring in inflammatory bowel disease. World J Gastroenterol 21:11246–11259. https://doi.org/10.3748/wjg.v21.i40.11246.
30. Iskandar HN, Ciorba MA. 2012. Biomarkers in inflammatory bowel disease: current practices and recent advances. Transl Res 159:313–325. https://doi.org/10.1016/j.trsl.2012.01.001.
31. Mosli MH, Zou G, Garg SK, Feagan SG, MacDonald JK, Chande N, Sandborn WJ, Feagan BG. 2015. C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. Am J Gastroenterol 110:802–820. https://doi.org/10.1038/ajg.2015.120.
32. Vermeire S, Van Assche G, Rutgeerts P. 2004. C-reactive protein as a marker for inflammatory bowel disease. Inflamm Bowel Dis 10:661–665. https://doi.org/10.1097/00054725-200409000-00026.
33. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Dore J, Ehrlich SD, MetaHIT Consortium, Bork P, Wang J. 2014. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 32:834–841. https://doi.org/10.1038/nbt.2942.
34. Tsilimigras MC, Fodor AA. 2016. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann Epidemiol 26:330–335. https://doi.org/10.1016/j.annepidem.2016.03.002.
35. Rosa Viner RB, Blank M, Rogers J. 2013. Increasing the multiplexing of protein quantitation from 6- to 10-plex with reporter ion isotopologues, p 1–7. Thermo Scientific Technical Note. Thermo Fisher Scientific, San Jose, CA.
36. Xiao J, Tanca A, Jia B, Yang R, Wang B, Zhang Y, Li J. 2018. Metagenomic taxonomy-guided database-searching strategy for improving metaproteomic analysis. J Proteome Res 17:1596–1605. https://doi.org/10.1021/acs.jproteome.7b00894.
37. 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. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.
38. Maier T, Guell M, Serrano L. 2009. Correlation of mRNA and protein in complex biological samples. FEBS Lett 583:3966–3973. https://doi.org/10.1016/j.febslet.2009.10.036.
39. Scanlan PD, Shanahan F, Marchesi JR. 2008. Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. BMC Microbiol 8:79. https://doi.org/10.1186/1471-2180-8-79.
40. Gaci N, Borrel G, Tottey W, O'Toole PW, Brugere JF. 2014. Archaea and the human gut: new beginning of an old story. World J Gastroenterol 20:16062–16078. https://doi.org/10.3748/wjg.v20.i43.16062.
41. Toyama BH, Hetzer MW. 2013. Protein homeostasis: live long, won't prosper. Nat Rev Mol Cell Biol 14:55–61. https://doi.org/10.1038/nrm3496.
42. Chandramouli K, Qian PY. 2009. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics 2009:239204. https://doi.org/10.4061/2009/239204.
43. Mesuere B, Debyser G, Aerts M, Devreese B, Vandamme P, Dawyndt P. 2015. The Unipept metaproteomics analysis pipeline. Proteomics 15:1437–1442. https://doi.org/10.1002/pmic.201400361.
44. Cohen J. 1988. Statistical power analysis for the behavioral sciences, 2nd ed. L. Erlbaum Associates, Hillsdale, NJ.
45. Winter SE, Lopez CA, Baumler AJ. 2013. The dynamics of gut-associated microbial communities during inflammation. EMBO Rep 14:319–327. https://doi.org/10.1038/embor.2013.27.
46. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. https://doi.org/10.1093/nar/gkr1044.
47. Ni J, Shen TD, Chen EZ, Bittinger K, Bailey A, Roggiani M, Sirota-Madi A, Friedman ES, Chau L, Lin A, Nissim I, Scott J, Lauder A, Hoffmann C, Rivas G, Albenberg L, Baldassano RN, Braun J, Xavier RJ, Clish CB, Yudkoff M, Li H, Goulian M, Bushman FD, Lewis JD, Wu GD. 15 November 2017. A role for bacterial urease in gut dysbiosis and Crohn's disease. Sci Transl Med https://doi.org/10.1126/scitranslmed.aah6888.
48. Sitkin S, Pokrotnieks J. 2018. Clinical potential of anti-inflammatory effects of Faecalibacterium prausnitzii and butyrate in inflammatory bowel disease. Inflamm Bowel Dis https://doi.org/10.1093/ibd/izy258.
49. Corthesy B. 2013. Multi-faceted functions of secretory IgA at mucosal surfaces. Front Immunol 4:185. https://doi.org/10.3389/fimmu.2013.00185.
50. Dai C, Jiang M, Sun MJ. 2018. Fecal markers in the management of inflammatory bowel disease. Postgrad Med 130:597–606. https://doi.org/10.1080/00325481.2018.1503919.
51. Chabriere E, Charon MH, Volbeda A, Pieulle L, Hatchikian EC, Fontecilla-Camps JC. 1999. Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. Nat Struct Biol 6:182–190. https://doi.org/10.1038/5870.
52. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35:833–844. https://doi.org/10.1038/nbt.3935.
53. Hughes ER, Winter MG, Duerkop BA, Spiga L, Furtado de Carvalho T, Zhu W, Gillis CC, Buttner L, Smoot MP, Behrendt CL, Cherry S, Santos RL, Hooper LV, Winter SE. 2017. Microbial respiration and formate oxidation as metabolic signatures of inflammation-associated dysbiosis. Cell Host Microbe 21:208–219. https://doi.org/10.1016/j.chom.2017.01.005.
54. Bohm R, Sauter M, Bock A. 1990. Nucleotide sequence and expression of an operon in Escherichia coli coding for formate hydrogenlyase components. Mol Microbiol 4:231–243. https://doi.org/10.1111/j.1365-2958.1990.tb00590.x.
55. Takahashi K, Nishida A, Fujimoto T, Fujii M, Shioya M, Imaeda H, Inatomi O, Bamba S, Sugimoto M, Andoh A. 2016. Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. Digestion 93:59–65. https://doi.org/10.1159/000441768.
56. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, Blugeon S, Bridonneau C, Furet JP, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas G, Blottiere HM, Dore J, Marteau P, Seksik P, Langella P. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A 105:16731–16736. https://doi.org/10.1073/pnas.0804812105.
57. Barnich N, Darfeuille-Michaud A. 2007. Adherent-invasive Escherichia coli and Crohn's disease. Curr Opin Gastroenterol 23:16–20. https://doi.org/10.1097/MOG.0b013e3280105a38.
58. Palmela C, Chevarin C, Xu Z, Torres J, Sevrin G, Hirten R, Barnich N, Ng SC, Colombel JF. 2018. Adherent-invasive Escherichia coli in inflammatory bowel disease. Gut 67:574–587. https://doi.org/10.1136/gutjnl-2017-314903.
59. Jansson JK, Baker ES. 2016. A multi-omic future for microbiome studies. Nat Microbiol 1:16049. https://doi.org/10.1038/nmicrobiol.2016.49.
60. Henderson B, Allan E, Coates AR. 2006. Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. Infect Immun 74:3693–3706. https://doi.org/10.1128/IAI.01882-05.
61. Palsson B, Zengler K. 2010. The challenges of integrating multi-omic data sets. Nat Chem Biol 6:787–789. https://doi.org/10.1038/nchembio.462.
62. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. 2017. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques 62:290–293. https://doi.org/10.2144/000114559.
63. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
64. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
65. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
66. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt C. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932. https://doi.org/10.1093/bioinformatics/btu739.
67. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA:

KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
68. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419. https://doi.org/10.1038/nmeth.4197.
69. Villen J, Gygi SP. 2008. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 3:1630–1638. https://doi.org/10.1038/nprot.2008.150.
70. Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villen J, Gygi SP. 2006. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. Mol Cell Proteomics 5:1326–1337. https://doi.org/10.1074/mcp.M500339-MCP200.
71. Wessel D, Flugge UI. 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 138:141–143. https://doi.org/10.1016/0003-2697(84)90782-6.
72. Van Rechem C, Black JC, Boukhali M, Aryee MJ, Graslund S, Haas W, Benes CH, Whetstine JR. 2015. Lysine demethylase KDM4A associates with translation machinery and regulates protein synthesis. Cancer Discov 5:255–263. https://doi.org/10.1158/2159-8290.CD-14-1326.
73. Tolonen AC, Haas W. 1 July 2014. Quantitative proteomics using reductive dimethylation for stable isotope labeling. J Vis Exp https://doi.org/10.3791/51416.
74. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. 2011. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol 7:461. https://doi.org/10.1038/msb.2010.116.
75. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, Moore RJ, Klemke RL, Camp DG, 2nd, Smith RD. 2011. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics 11:2019–2026. https://doi.org/10.1002/pmic.201000722.
76. Lapek JD, Jr, Lewinski MK, Wozniak JM, Guatelli J, Gonzalez DJ. 2017. Quantitative temporal viromics of an inducible HIV-1 model yields insight to global host targets and phospho-dynamics associated with Vpr. Mol Cell Proteomics 16:1447–1467. https://doi.org/10.1074/mcp.M116.066019.
77. Eng JK, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5:976–989. https://doi.org/10.1016/1044-0305(94)80016-2.
78. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. 2006. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol 24:1285–1292. https://doi.org/10.1038/nbt1240.
79. Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villen J, Haas W, Sowa ME, Gygi SP. 2010. A tissue-specific atlas of mouse protein phosphorylation and expression. Cell 143:1174–1189. https://doi.org/10.1016/j.cell.2010.12.001.
80. Elias JE, Gygi SP. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4:207–214. https://doi.org/10.1038/nmeth1019.
81. Elias JE, Haas W, Faherty BK, Gygi SP. 2005. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods 2:667–675. https://doi.org/10.1038/nmeth785.
82. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2:43–50. https://doi.org/10.1021/pr025556v.
83. Gupta N, Pevzner PA. 2009. False discovery rates of protein identifications: a strike against the two-peptide rule. J Proteome Res 8:4173–4181. https://doi.org/10.1021/pr9004794.
84. van der Sluys Veer A, Brouwer J, Biemond I, Bohbouth GE, Verspaget HW, Lamers CB. 1998. Fecal lysozyme in assessment of disease activity in inflammatory bowel disease. Dig Dis Sci 43:590–595.
85. Brophy MB, Nolan EM. 2015. Manganese and microbial pathogenesis: sequestration by the mammalian immune system and utilization by microorganisms. ACS Chem Biol 10:641–651. https://doi.org/10.1021/cb500792b.