Host-centric Proteomics of Stool: A Novel Strategy Focused on intestinal Responses to the Gut Microbiota*§

Joshua S. Lichtman‡, Angela Marcobal§, Justin L. Sonnenburg§, and Joshua E. Elias‡¶

The diverse community of microbes that inhabits the human bowel is vitally important to human health. Host-expressed proteins are essential for maintaining this mutualistic relationship and serve as reporters on the status of host-microbiota interaction. Therefore, unbiased and sensitive methods focused on host proteome characterization are needed. Herein we describe a novel method for applying shotgun proteomics to the analysis of feces, focusing on the secreted host proteome. We have conducted the most complete analysis of the extracellular mouse gut proteome to date by employing a gnotobiotic mouse model. Using mice colonized with defined microbial communities of increasing complexity or a complete human microbiota ("humanized"), we show that the complexity of the host stool proteome mirrors the complexity of microbiota composition. We further show that host responses exhibit signatures specific to the different colonization states. We demonstrate feasibility of this approach in human stool samples and provide evidence for a "core" stool proteome as well as personalized host response features. Our method provides a new avenue for noninvasive monitoring of host-microbiota interaction dynamics via host-produced proteins in stool. Molecular & Cellular Proteomics 12: 10.1074/mcp.M113.029967, 3310–3318, 2013.

Hundreds to thousands of microbial "species" and \( \sim 10^{13} \) individual organisms make up any one person’s gut microbiota (1), making the gastrointestinal (GI) tract one of the most complex biological ecosystems ever studied. The dynamic interaction between these communities and the host organism is linked to many aspects of health and disease in humans including inflammatory bowel diseases (2), obesity (3), allergies (4), and autoimmunity (5). Sequence-based approaches (e.g. metagenomics and 16S community profiling) have effectively elucidated the gene and species composition of several microbial communities that influence health and disease (3, 6, 7). However, sequencing alone is limited to defining microbial community constituents, providing little insight into the myriad ways hosts can respond to their resident microbes. Despite an individualized “fingerprint” (7) of microbiota composition, a major gap separates our understanding of how differently composed microbial communities specifically impact host responses in the gut. Enhanced methods that sensitively probe the microbial impact on host biology will be critical to expanding insight into the host-microbiota super-organism. Stool presents an easily sampled biological material that offers a window into complex host-microbe relationships.

Early studies of the host response to microbiota utilized laser-capture micro dissection (LCM) (8), followed by gene expression analysis of particular cell types in the GI epithelium. Although providing an unprecedented view into the ways microbiota can impact host biology, this approach is technically difficult, provides only a semiquantitative estimate of biologically pertinent protein expression, and requires the collection of intestinal tissue. Therefore, LCM and subsequent transcriptional profiling of host tissue prevents time-course experimentation in animal models and is not readily translated to patient studies.

The combination of liquid chromatography and tandem mass spectrometry (LC-MS/MS) provides a flexible, dynamic platform for the simultaneous identification and quantification of thousands of proteins in fecal samples. Implementing this technology to study gut biology has been inhibited by technical limitations stemming from the overwhelming complexity of the resident microbiota metagenome: it greatly overshadows the host’s genome, its composition varies between individuals, and it encodes only a sparsely defined proteome. Pioneering studies of this complex system focused on the metaproteome, attempting to identify as many host and bacterial proteins as possible using matched metagenomic sequencing and shotgun proteomics (9, 10). Although matched sequencing data can improve bacterial protein identifications, drawing biological conclusions from data that is composed predominantly of proteins with ill-defined functions and origins remains difficult (10).

Our approach acknowledges the contrast between the technical challenges posed by measuring bacterial proteins in
the context of complex microbial communities and the importance of elucidating the host response to microbial dynamics. By combining technical improvements in sample preparation before LC-MS/MS and subsequent data analysis, we have developed a workflow in which abundance changes of >3000 host proteins shed into the GI tract can be sensitively assayed. Applying these techniques to defined perturbations of the gnotobiotic mouse model establishes a pathway for discovering functional relationships between microbiota and host response. Furthermore, extending this approach to archived or freshly collected human stool samples makes possible the elucidation of specific host responses to microbiota for which extensive characterization is already complete or planned.

EXPERIMENTAL PROCEDURES

Gnotobiotic Mouse Model—Gnotobiotic and conventional (RF, Taconic, Inc.) Swiss-Webster mice were maintained as previously described (11). Humanization was performed using human fecal samples from one healthy human donor (male, western diet) as previously described (12). Mice were fed with standard diet (Purina LabDiet 5K67).

Cultures of bacteria were prepared as following. Tryptone-yeast extract-glucose (TYG) medium was used for growing Bacteroides species (Bacteroides thetaiotaomicron/VPI-5482 and B. caccae ATCC 45185T) as described previously (11). Bifidobacterium species (Bif. longum NCC2705 and Bif. bifidum DSM 20456) were grown in Man-Rogosa-Sharpe (MRS) medium (Oxoid LTD., Basingstoke, Hampshire, England). Reinforced clostridium medium (RCM; Becton Dickinson, Franklin Lakes, NJ) was used for the growth of Clostridium scindens (version 2.3.01) using static carbamidomethylation of cysteines, differential oxidized methionines, 50 ppm precursor mass tolerance, 2 miscreavages and 0.8 Da fragment ion tolerance. FDRs were estimated with the target-decoy search strategy and peptide false-positives were filtered to 1% (15). Data from technical replicates was combined to improve identification and quantification. Protein FDR of the entire data set was filtered by eliminating proteins with a single spectral count with a homology factor (Mascot ion score/Mascot homology threshold)<1 and with the maximum protein ion score >20. Normalized spectral counts were calculated by dividing the spectral

Sample Preparation—Single fecal samples were collected directly from mice, immediately frozen in liquid nitrogen and stored at −80 °C before use. Fecal pellets were cut on dry ice, weighed, suspended in 300 μl 8 M Urea (Invtrogen, Carlsbad, CA; CAS# 57–13-6), 100 mM NaCl (Fisher Scientific, CAS# 7467–14-5), 25 mM Tris (Amresco, Solon, Ohio; CAS# 77–86–1), pH 8.2 with protease inhibitors (Roche Complete, EDTA free, Cat#05056849001). Pellets were physically disrupted with the tip of a pipette and vortexed thoroughly. Disrupted pellets were centrifuged (2500 × g, 5 min) to pellet insoluble material. The supernatant was then ultra-centrifuged (35000 × g, 30 min, 4 °C) to pellet bacteria. The final supernatant, containing soluble protein, was reduced and alkylated with iodoacetamide. Alkylated proteins were fractionated using a reverse-phase C-4 cartridge (Grace Vydac, Hesperia, CA; Cat# 214SP1E001), collecting fractions at 20%, 40%, 60% and 100% acetonitrile (ACN) (Fisher Scientific, CAS# 75–05–8) + 0.5% acetic acid (Fisher Scientific, CAS# 64–19–7). Fractions were vacuum centrifuged to dryness and suspended in 200 μl 50 mM ammonium bicarbonate (pH 8) + 5 μl/mi trypsin (Promega, VS111). Proteins were then digested overnight at 37 °C into peptides and de-salted using C-18 Sep-pak cartridges (Waters, Milford, MA; Cat# WAT054855), and analyzed via reverse-phase liquid chromatography tandem mass spectrometry (LC-MS/MS).

LC-MS/MS—De-salted peptides were suspended in 20 μl sample buffer (5% ACN/5% Formic Acid (Acras, CAS# 64–18–6). A Spark-Holland 920 auto sampler loaded 2μl of sample onto a 18 cm, 100 μm, C18 fused silica micro-capillary column (inner diameter = 125 um) ending in an in-house pulled needle tip (internal diameter = 5 um). Ions were then eluted from this column at 300 μl/min over a 2.5-hour gradient (0–7% 0.5min, 7–25% 120min, 20–45% 25 min, 45–100% 3 min, 100% 7 min, 0% 7min) using an Agilent 1200 G1312B Binary pump (Agilent Technologies, San Jose, CA). The column was coupled to a home-built electrospray ionization source and LTQ-Velos Orbitrap mass spectrometer (Thermo Scientific, Santa Clara, CA). Full scans were collected in the orbitrap at 60,000 resolution in centroid mode followed by a standard top-10, data-dependent framework with CID fragmentation in which singly charged precursor ions were excluded from fragmentation. Masses within 2.5 Da of precursor masses were dynamically excluded for 60 s.

Data Analysis—All eukaryotic protein databases were downloaded from Uniprot (13) and bacterial databases were downloaded from the Integrated Microbial Genomes database (14). The combined host, food and microbial database was constructed by concatenating mouse proteins with those of wheat, corn, soybean, yeast, alfalfa, and the six inoculating microbes (Bacteroides thetaiotaomicron, Bacteroides caccae, Bifidobacterium longum, Clostridium scindens, Edwardsiella tarda, and Clostridium spiroforme) and common contaminants (90,152 protein, Uniprot, released Oct. 30, 2012). The BT and M2 databases (77,541 and 89,145 proteins respectfully, Uniprot, released Oct 30, 2012) were constructed similarly with their respective microbial constituents. The readRAW program (version 4.3.0) was used to generate peak lists from the original data. All protein sequences were reversed and concatenated to the original forward-oriented proteins (15). Spectra were assigned to peptides using a semi-specific enzyme specificity with the Mascot (16) search algorithm (version 2.3.01) using static carbamidomethylation of cysteines, differential oxidized methionines, 50 ppm precursor mass tolerance, 2 miscreavages and 0.8 Da fragment ion tolerance. FDRs were estimated with the target-decoy search strategy and peptide false-positives were filtered to 1% (15). Data from technical replicates was combined to improve identification and quantification. Protein FDR of the entire data set was filtered by eliminating proteins with a single spectral count with a homology factor (Mascot ion score/Mascot homology threshold)<1 and with the maximum protein ion score >20. Normalized spectral counts were calculated by dividing the spectral

Molecular & Cellular Proteomics 12.11 3311
counts by the total number of spectra that were confidently assigned in a run and multiplied by 100. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (17) with the data set identifier PXD000240.

For cluster analysis, log2-normalized fold-change of normalized spectral counts was calculated relative to the average normalized spectral counts for a particular protein across all samples. Hierarchical clusters were constructed using Euclidean distance and average linkage metrics with the clustergram function of Matlab (MathWorks, v2012a) with no further normalization. Principal component analysis was conducted on normalized spectral counts using the princomp function in Matlab. GO-term functional enrichments were conducted using DAVID (18, 19) online functional annotation platform with the specified background protein set. Benjamini-corrected p values are reported. Line graphs were plotted in GraphPad Prism (GraphPad Software, v5.04).

RESULTS

A Novel Workflow for Fecal Proteomics Greatly Improves Host Proteome Coverage—Previous analyses of host-derived gut proteomes were inhibited by three major factors: (1) minimal contribution of host proteins to total stool proteome, (2) a large and ill-defined search space associated with the predicted microbiota proteome, and (3) partial proteolysis that occurs during gut transit. First, because of their expected modest representation in the extraordinarily complex fecal matrix, the proteins most biologically relevant to host responses are unlikely to be identified. This leads to incompatibilities between a mass spectrometer’s ability to sensitively detect very low-abundance proteins and downstream analysis tools’ ability to distinguish and correctly designate small signals in a vast sea of protein diversity. To cope with these challenges, we employed a gnotobiotic mouse model to control microbial complexity and enrichment strategies to better study host responses. Mice colonized with microbial communities of increasing complexity (Table I) with defined proteomes facilitated the development and testing of our method.

Having established a defined representation of the intestinal environment, we next extended a previously described differential centrifugation protocol (20) (Fig. 1A) to a fecal sample from a bi-colonized mouse (community M2, Table I). Recovering supernatant from our centrifugation procedure enriched proteins shed by the host into the gut environment. This further improved host protein identifications (1.78-fold) and overall spectrum assignment (4.5-fold) compared with the analysis of the cell pellet. The host proteins we identified from the supernatant were enriched for expected biologically relevant GO annotations including secreted (p = 5e-10), generation of precursor metabolites and energy (p = 2e-15) and protease (p = 1e-10) whereas the host proteins found in the pellet were enriched for ontologies less likely to be relevant to extracellular host-microbe interactions: methylation (p = 1e-8), acetylation (p = 3e-9), and filament (p = 3e-9). Notably, the two-step centrifugation procedure we used separated large particulate matter from intact free-living cells. It is therefore compatible with parallel studies of the microbiota, an important dimension for future studies, which we do not address here.

Second, the >3 million predicted ORFs in the bacterial metagenome greatly increase the hypothetical proteome from the ~70,000 annotated human proteins (13). This search space explosion can easily overwhelm current state-of-the-art, database-dependent search engines commonly used in proteomics investigations (21). Search space complexity is further compounded by interindividually variable variability in microbiota composition and function. By searching against the single, smaller, and well-characterized host genome, we efficiently identified a larger proportion of host-derived peptides while maintaining high search algorithm accuracy (supplemental Fig. S1). We estimated this improvement by subjecting the feces collected from mice stably colonized with five known, sequenced bacteria (community M5, Table I) to our workflow (Fig. 1A). The resulting spectra were searched with a mouse-only database (68,998 proteins) or a composite database consisting of the mouse proteome, the meta-proteome of the five-member microbiota, and the available proteomes of the plants that make up the standard chow (315,278 proteins). Of 32,595 spectra (5587 unique peptides) confidently assigned to host proteins in the host-only search, just 186 spectra (0.5%) and 52 unique peptides (0.9%) were re-assigned to microbial or plant proteomes when using the larger database. This is consistent with our FDR threshold of <1% for selecting high-confidence unique peptides. By avoiding the distraction that occurs with larger search spaces (21), the host-only search returned 16.6% more high confidence host IDs (32,595 vs. 27,953 IDs) (supplemental Fig. S1).

The third challenge posed in fecal proteomics is that proteins can be partially digested throughout the gastrointestinal tract. To improve overall identification rates, it is important to ensure that partially digested proteins are collected and analyzed. We accomplished this through C-4 solid phase extraction, which has previously been applied to effectively fractionate both peptides and intact proteins (22). C-4 fractionation was conducted using four increasing steps of acetonitrile. The consequent eluted fractions were digested with trypsin, ana-
lyzed by LC-MS/MS, and the resulting spectra were searched requiring that peptides had at least one tryptic terminus (semi-tryptic). The consideration of semispecific peptides approximately doubled the number of identified spectra in both germ-free and conventional mice feces (Fig. 1B). While peptides showing partial trypsin specificity are commonly excluded as incorrect identifications (23, 24), complete analysis of the mouse fecal proteome appears to require that these peptides be considered. As expected, tryptic peptides, likely derived from intact proteins, were more abundant than semitryptic peptides, which may be the remnants of partial digestion in the lower GI (Fig. 1C).

As with any technology, it is important to ensure technical reproducibility to better differentiate true biological signal

Fig. 1. A novel method for the analysis of the mammalian gut proteome. A, In this workflow, the resulting supernatant from differential centrifugation is fractionated with C-4 reverse phase solid-phase extraction to enrich for proteins and peptides. The resulting fractions are digested with trypsin, de-salted and applied to reverse-phase LC-MS/MS to study the host contribution to the gut proteome. The resulting data can then be applied to univariate and multivariate analyses. B, The distribution of peptide-spectral matches between fully-tryptic and semitryptic peptides in germ-free and conventional mice indicating the importance for utilizing semitryptic parameters when searching this data. C, The frequency of identification across biological replicates in germ-free and conventional mice. Comparing parts b and c, there are more spectra per unique peptide in the fully-tryptic category but fewer total unique peptides, indicating that the fully-tryptic ones are more abundant and less diverse.
Host-centric Proteomics of Stool

from noise. We have shown that a sample from a mouse colonized with two microbes (M2, Table I), run on separate days and with separate chromatographic columns, yields raw spectral counts that are highly correlative \( R^2 = 0.988 \). Based on this high level of reproducibility, we combined technical replicates in all subsequent analyses to improve quantification and proteome coverage.

**Host Protein Secretion is Altered by Microbiota Composition**—We investigated the effect the gut microbiota composition has on host protein production by analyzing stool from (1) germ-free mice, (2) gnotobiotic mice colonized with defined microbial communities ranging in complexity from one to five microbes, and (3) mice colonized with a complex microbiota derived from conventional mice or from a human fecal microbiota (“humanized”) mice (Table I). Microbiota were allowed to stabilize for 20 days following host intragastric inoculation, and single fecal samples (~60 mg) were collected from each mouse. Samples were processed using our validated protocol and analyzed with technical replicate LC-MS/MS runs. Across the six colonization states, we identified 2529 unique host proteins from 17,221 unique peptides, with estimated protein and peptide FDRs of 3.7% and 0.89%, respectively (supplemental Data Set S1, supplemental Table S1). We identified an average of 612 proteins and 3250 unique peptides per individual animal. Each protein was represented by a median of two unique peptides and four total spectra (respective ranges: 1–230 and 1–20,909). Proteome complexity and intragroup variation increased with microbial community complexity (Fig. 2A, 2B). Sixty-eight proteins were identified in all 22 mice (supplemental Table S2). These proteins define a core secreted mouse gut proteome expressed independent of microbial colonization. These proteins were significantly enriched for gene ontologies one might expect for secreted gut luminal proteins, including signal peptide \( p = 7e^{-9} \), hydrolase \( p = 5e^{-8} \), and protease activity \( p = 1e^{-8} \). We also defined a core fecal proteome associated with microbial colonization: 26 proteins were absent in all germ-free mice but were identified in at least half of all colonized mice. These proteins were significantly enriched for immunoglobulin domains \( p = 8e^{-3} \), consistent with a known host response to microbes (supplemental Fig. S2, supplemental Table S3 (25)).

We detected several functionally related proteins in all microbial colonization states with highly reproducible expression within each condition (Fig. 2C, supplemental Fig. S3), indicating a direct relationship between the microbes and these particular aspects of the host biology. Likewise, we distinguished aspects of the host responses that were specific to each colonization state by hierarchical clustering (supplemental Fig. S4). These results demonstrate that increasing variability and complexity in host responses are linked to community complexity and to reproducible, colonization state-specific proteomic signatures.

**Microbiota-Elicited Host Responses Exhibit Host-Specificity And Temporal Stability**—Gut microbiota composition varies less within individuals over time than between similar individuals in both mice (26) and humans (27). It is unclear whether the observable host proteome is comparatively steady, or demonstrates similar variability. We addressed this by sampling feces from pairs of germ-free and conventional mice (1 male, 1 female for each condition) over 3 days, sampling every 24 h (supplemental Table S4). The fecal proteomes of the germ-free mice were indistinguishable, with minimal variation over time or between individuals. In contrast, conventional mice showed variation in their responses to the microbiota over time relative to GF mice, but this variation was less than differences observed between individuals (Fig. 3). Notably, germ-free mice consistently showed highly reproducible secreted protein profiles despite differences in age, cages, and litters (supplemental Fig. S5). These data indicate that intra- and intermouse variability in the host fecal proteome is largely dependent on the gut microbiota composition, and represents a sensitive reporter of host responses to the microbiota. Furthermore, our data confirm that germ-free mice provide an extremely stable baseline state for monitoring host responses. This highlights the power of the gnotobiotic mouse model as a platform for using simplified communities to define the links between particular taxa and ecologies (collections of microbes) to specific host responses.

**Translation to Human Samples**—We next tested whether our findings and method could be extended to the analysis of human fecal samples. We analyzed stool samples from three healthy human donors with distinct gut microbiota compositions (Purna Kashyap, Justin Sonnenburg, unpublished data). Using a single 100–150 mg stool sample from each donor, we identified a total of 234 human proteins with 1058 unique peptides from nearly 10,000 confidently assigned spectra with estimated protein and peptide FDRs of 2.0% and 0.7%, respectively (supplemental Table S5). Although the number of proteins identified from these samples was considerably lower than in mouse samples (2032 proteins identified in similar analysis of three CONV mice), these proteins were significantly enriched for the types of functions we would expect to find in the gut lumen (Fig. 4A, supplemental Table S6). Our data reveals a set of proteins biologically-distinct from a previous analysis of the fecal proteome, likely owing to our methodological focus on secreted host proteins (10). We further identified a core proteome that consisted of 57 proteins shared between these three individuals (Fig. 4C). Despite the reproducible presence of this core proteome, the relative abundance of most shared proteins varied between the three individuals consistent with host-specific proteomic signatures (Fig. 4B). The proteins that were unique to individuals had no functional enrichment over the entire set, whereas the core proteins were significantly enriched for secreted and extracellular annotations (Fig. 4C). Twenty-two of the proteins found in the human core proteome have orthologs that occur in the total mouse core proteome (supplemental Table S7), consist-
with a general conservation of the mammalian gut’s response to resident microbes (28).

A fecal sample from one human subject was used as the inoculum for the humanized mouse analysis (Fig. 2), allowing us to test whether individualized aspects of the host fecal proteome could be reconstituted in the recipient mice. As revealed by principal component analysis of 178 orthologous proteins present in both the mouse (Fig. 2) and human data sets (Fig. 4), the proteomic profile for the fecal donor was most similar to those of the recipient mice (supplemental Fig. S6). Recent reports illustrate faithful reconstitution of the human microbiota on transplantation into the mouse gut (29) and reconstitution of host phenotypes that are microbiota-dependent (30). Our study extends these findings, suggesting a
Fig. 3. Interindividual variation exceeds variation within an individual over time. Principal component analysis of the fecal proteomes of two germ free and two conventional mice sampled for 3 days.

Fig. 4. Application of secreted proteome enrichment method to human fecal samples. A, Functional annotation enrichment of the 244 human proteins over the human proteome compared with the enrichment found in a recent gut proteome analysis (10). B, GO term enrichment calculated using DAVID for the proteins found in the core versus all detected proteins. C, Hierarchical clustering of the normalized fold-change protein expression compared with the average of the three individuals.
general reconstitution of microbe-dependent host responses in humanized mice, including personalized aspects of these responses. Because our method is compatible with analysis of frozen human fecal samples, it should be easily applied to archived or newly acquired samples. Extending our method to analyze additional samples will expand the definition of the core secreted host proteome and establish proteomic biomarkers that reflect specific microbiota-host interactions in health and disease.

**DISCUSSION**

We report here a novel method for interrogating host-derived proteins from mouse and human fecal samples. Utilizing these methods in conjunction with a gnotobiotic mouse model, we have measured the effect of microbial community complexity on the secreted host proteome. As demonstrated by this set of proof-of-concept experiments, our methods circumvent several problems associated with gut microbiota proteomics including the complexity of feces, the size of the microbial metaproteome, and naturally occurring proteolysis in the GI tract. A host-centric approach, as described here, will complement ongoing metagenomic efforts, while facilitating the discovery of microbe-influenced host pathways with direct health relevance.

This method represents the first targeted assessment of host responses from the GI tract in an unbiased and noninvasive manner. It will allow banked fecal samples previously used for microbial studies to be assayed for host protein expression. As such, it can be used to establish critical links between the gene and species composition of microbial communities, and the effects they have on host biology. Numerous studies have characterized microbiota in healthy versus disease states. While many of these have concluded that microbiota dysbiosis is linked to disease, few demonstrated specific host pathways that give rise to pathological states. Host stool proteomics will help establish mechanistic links between microbiota and host response and provide an avenue for discovering novel biomarker proteins. By directly sampling the matrix in which these interactions occur, it is likely that these proteins will be more tightly linked with gut-microbiota-influenced physiological status compared with markers found in blood or urine.

This method opens several opportunities for potential expansion and optimization of stool-based proteomics. First, by performing two centrifugation steps, we partitioned microbe-bound proteins from the secreted soluble host proteins on which our analysis focused. These microbial associated proteins likely include antimicrobial peptides secreted by the host (31), which may prove to be highly relevant to health and disease. Separating informative protein signatures from abundant microbial and microbe-adherent host proteins like immunoglobulins will likely include antimicrobial peptides, all of which result in spectra that are not readily expected from a standard tissue-based proteome study. We attribute this deficiency to proteins that undergo a large degree of post-translational processing (e.g. glycosylation), to nonpeptidyl food products, and to metabolites that co-elute with desired peptides, all of which result in spectra that are not readily identifiable. We expect further computational and methodological optimization will mitigate these limitations.

The complex interactions between the gut microbiota and their host require precise, unbiased measurements of host and microbe function. By coupling the described proteomics method with bacterial metagenomics and metabolomics, and by dissecting these interactions in the gnotobiotic mouse model, the field can continue to improve its understanding of gastro-intestinal biology. Likewise, this new layer of information, which directly assays host responses to any number of experimental and natural perturbations, will improve our ability to differentiate and diagnose complex GI diseases.

Acknowledgments—We thank Olga Vitek for statistics advice, Steve Higginbottom for work with the gnotobiotic mice and many members of the Elias and Sonnenburg laboratories for insightful discussions and contributions to the defined colonization experiments.
Host-centric Proteomics of Stool

This work was supported in part by the NIH Training Grant (JSL), Stanford Digestive Disease Center (JEE) and a grant from the National Institutes of Health (R01-DK080525)(JLS).

This article contains supplemental Figs. S1 to S6, Tables S1 to S7, and Data Set S1.

To whom correspondence should be addressed: Clark Center W300C, 318 Campus Drive, Stanford, CA 94305. Tel.: 650-724-3422; Fax: 650-724-5791; E-mail: josh.elias@stanford.edu.

Author Contributions: JSL performed the proteomics, JSL, JEE and JS took part in the data analysis, JSL and AM ran the mouse experiments, JSL, AM, JS and JEE wrote the manuscript.

REFERENCES

1. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005) Host-bacterial mutualism in the human intestine. Science 307, 191–20

2. Round, J. L., and Mazmanian, S. K. (2006) The gut microbiota shapes intestinal immune responses during health and disease. Nature Rev. 9, 313–23

3. Sanderson, S., Boardman, W., Cloft, C., and Gibson, R. (2006) Human gut microbes associated with obesity. Nature 444, 1022–1023

4. Maslow, K. M., and Mackay, C. R. (2011) Diet, gut microbiota and immune responses. Nat. Immunol. 12, 8–9

5. Reid, G., Younes, J. A., Van der Mei, H. C., Gloor, G. B., Knight, R., and Busscher, H. J. (2011) Microbiota restoration: natural and supplemented recovery of human microbial communities. Nat. Rev. Microbiol. 9, 27–38

6. Gill, S. R., Pop, M., Deboy, R. T., Eckberg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., and Nelson, K. E. (2006) Metagenomic analysis of the human distal gut microbiome. Science 312, 1355–9

7. Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. and Relman, D. A. (2005) Diversity of the human intestinal microbial flora. Science 308, 1635–8

8. Stappenbeck, T. S., Hooper, L. V., Manchester, J. K., Wong, M. H., and Gordon, J. I. (2002) Laser capture microdissection of mouse intestine: characterizing mRNA and protein expression, and profiling intermediary metabolism in specified cell populations. Methods Enzymol. 356, 167–96

9. Verkerkboes, N. C., Russell, A. L., Shah, N., Leggitt, C. M., and Janssens, J. K. Shotgun metaproteomics of the human distal gut microbiota. ISME J. 3, 179–89

10. Erickson, A. R., Cantarel, B. L., Lamendella, R., Darzi, Y., Mongodin, E. F., Elias, J., and Gygi, S. (2007) Target-decoy search strategy for increased sensitivity and specificity of protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567

11. Vizcaino, J. A., Coté, R. G., Csordas, A., Di´ez, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O’Kelly, G., Schoenegger, A., Ovelleiro, D., P´erez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 41, D1063–9

12. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research 37, 1–13

13. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protocols 4, 44–57

14. Apajalahti, J. H., Särkilahti, L. K., Mäki, B. R., Heikkinen, J. P., Nurminen, P. H., and Holben, W. E. (1998) Effective recovery of bacterial DNA and percent-guanine-plus-cytosine-based analysis of community structure in the gastrointestinal tract of broiler chickens. Appl. Environ. Microbiol. 64, 1915–20

15. Elias, J., Haas, W., Faherty, B., and Gygi, S. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat. Methods 2, 667–75

16. Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol. Cell. Proteomics 3, 808–14

17. Janssens, J. K. Shotgun metaproteomics of the human distal gut microbiota. Nucleic Acids Res. 37, 167–96

18. Kruse, M. A., Feierman, L. J., Snodgrass, L. A., Winograd, S., and Holmsen, H. F. (2006) The human intestinal microbial flora. Science 312, 1915–20

19. Williams, K. A., Womack, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henriques, B., Crook, L. W., Russell, A., Verkerkboes, N. C., Hettrich, R. L., and Gordon, J. I. (2010) Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. Science 328, 1705–9

20. Thompson, C. L., Hofler, M. J., Campbell, I. L., and Holmes, A. J. (2010) Community dynamics in the mouse gut microbiota: a possible role for IRf8-regulated genes in community homeostasis. PloS one 5, e10335

21. Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. (2009) Bacterial community variation in human body habitats across space and time. Science 326, 1694–7

22. Rawls, J. F., Samuel, B. S., and Gordon, J. I. (2004) Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. Proc. Natl. Acad. Sci. U. S. A. 101, 4596–601

23. Turnbaugh, P., and Ridaura, V. (2009) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Translational Med. 1, 1–19

24. Turnbaugh, P. J., Ridaura, V. K., Faith, J. I., Rey, F. E., Knight, R., and Gordon, J. I. (2009) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Translational Med. 1, 6ra14

25. Agerberth, B., Charo, J., Werr, J., Olsson, B., Idali, F., Lindblom, B., Kiessling, R., Jörnvall, H., Wigzell, H., and Gudmundsson, G. H. (2008) The human antimicrobial and chemotactic peptides LL-37 and alpha-defensins are expressed by specific lymphocyte and monocyte populations. Blood 96, 3086–93

26. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A., and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1985–904

27. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154–69

28. Wollam, J. N., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henriques, B., Crook, L. W., Russell, A., Verkerkboes, N. C., Hettrich, R. L., and Gordon, J. I. (2010) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc. Natl. Acad. Sci. U. S. A. 106, 5859–64