

# Multi-omics Comparative Analysis Reveals Multiple Layers of Host Signaling Pathway Regulation by the Gut Microbiota

Nathan P. Manes,<sup>a</sup> Natalia Shulzhenko,<sup>b</sup> Arthur G. Nuccio,<sup>a</sup> Sara Azeem,<sup>a</sup> Andrey Morgun,<sup>c</sup> Aleksandra Nita-Lazar<sup>a</sup>

Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA<sup>a</sup>; College of Veterinary Medicine, Oregon State University, Corvallis, Oregon, USA<sup>b</sup>; College of Pharmacy, Oregon State University, Corvallis, Oregon, USA<sup>c</sup>

**ABSTRACT** The bodies of mammals are hosts to vast microbial communities composed of trillions of bacteria from thousands of species, whose effects on health and development have begun to be appreciated only recently. In this investigation, an integrated analysis combining proteomics and transcriptomics was used to quantitatively compare the terminal ilia from conventional and germfree mice. Female and male mice responded similarly to the microbiota, but C57BL/10A mice responded more strongly than BALB/c mice at both the transcriptome and proteome levels. The microbiota primarily caused upregulation of immunological pathways and downregulation of metabolic pathways in the conventional mice. Many of the affected pathways were altered only at either the transcriptome or proteome level. Of the pathways that were affected at both levels, most were affected concordantly. The discordant pathways were not principally involved in the immune system but instead were related to metabolism, oxidative phosphorylation, protein translation, transport, and turnover. To broaden the discovery of affected host pathways, a meta-analysis was performed using intestinal transcriptomics data from previously published studies of germ-free versus conventional mice with diverse microbiota populations. Similar transcript-level responses to the microbiota were found, and many additional affected host pathways were discovered.

**IMPORTANCE** Multiple host pathways were affected by its adaptation to the microbiota. We have found significant transcriptome-proteome discordance caused by the microbiota. This discovery leads to the definite conclusion that transcript-level analysis is not sufficient to predict protein levels and their influence on the function of many specific cellular pathways, so only analysis of combinations of the quantitative data determined at different levels will lead to a complete understanding of the complex relationships between the host and the microbiota. Therefore, our results demonstrate the importance of using an integrative approach to study host-microbiota interaction at the molecular level.

**KEYWORDS** cell signaling, host-microbiome interaction, microbiota, multi-omics, proteomics, systems biology, transcriptomics

The bodies of mammals are homes to vast populations of indigenous microbes. Each collection of microbes is called a microbiota, and each collective metagenome, sometimes referred to as a “second genome,” is called a microbiome. The human microbiome contains ~7,000 species and ~1,000,000 genes, and most of it is located in the gut (1). The microbiota is heterogeneous and dynamic: it changes with age, diet, and health status (2). Numerous aspects of host physiology are influenced by the gut

Received 28 August 2017 Accepted 29 September 2017 Published 24 October 2017

**Citation** Manes NP, Shulzhenko N, Nuccio AG, Azeem S, Morgun A, Nita-Lazar A. 2017. Multi-omics comparative analysis reveals multiple layers of host signaling pathway regulation by the gut microbiota. *mSystems* 2:e00107-17. <https://doi.org/10.1128/mSystems.00107-17>.

**Editor** Sean Michael Gibbons, MIT

**Copyright** © 2017 Manes et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](http://creativecommons.org/licenses/by/4.0/).

Address correspondence to Andrey Morgun, [anemorgun@hotmail.com](mailto:anemorgun@hotmail.com), or Aleksandra Nita-Lazar, [nitalazarau@niaid.nih.gov](mailto:nitalazarau@niaid.nih.gov).

N.P.M. and N.S. contributed equally to this article.

microbiota, including the development of the immune system, drug metabolism, nutritional status, and pathological states such as obesity, diabetes, cardiovascular disease, chronic gastrointestinal disorders such as inflammatory bowel disease, and asthma (3, 4). The significance of resident microbe-microbe and host-resident microbe interactions as a factor for human health (5) and biomedical research (6) is becoming apparent. Consequently, interest in the human microbiota is rapidly growing, and the NIH Human Microbiome Project (HMP) and the Metagenomics of the Human Intestinal Tract (MetaHIT) programs were recently launched to coordinate efforts to elucidate its biology. In parallel, model systems are being used to study the influence of the microbiota on the host, and germfree (GF) mice are an important tool for studying the effects of the microbiota on the host's physiology.

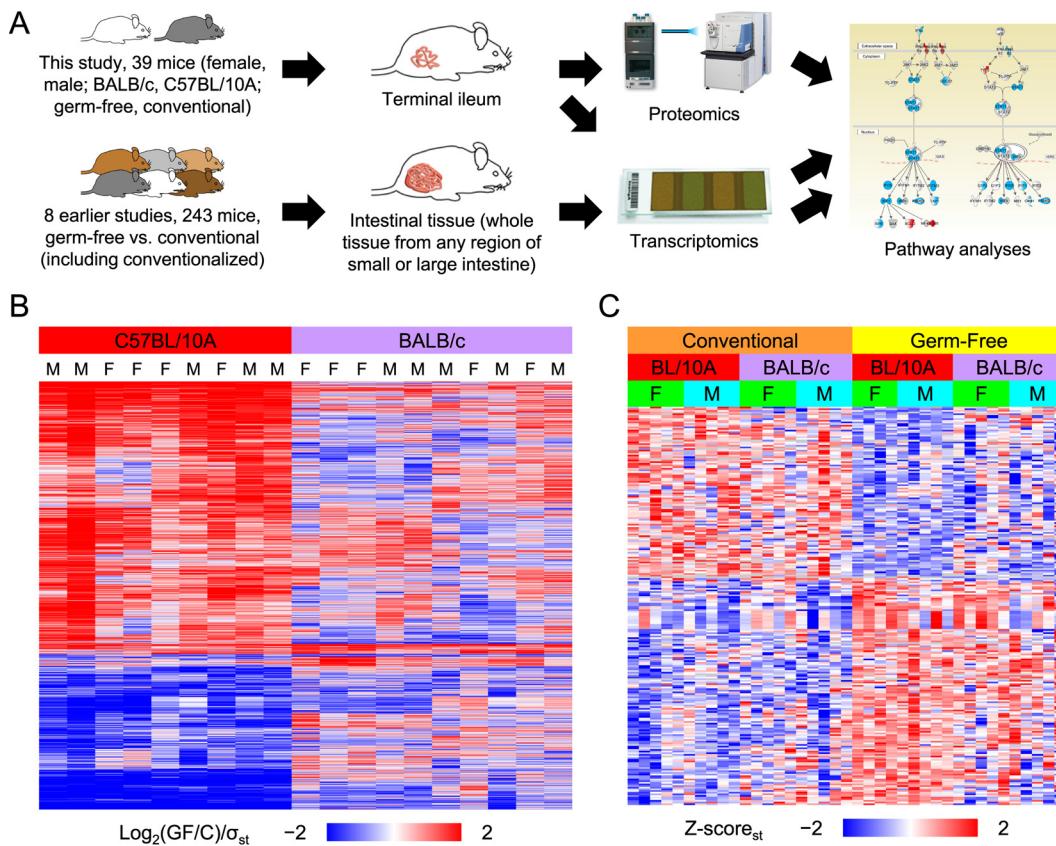
The characteristics of GF mice include dramatic physiological changes and metabolic and immune impairments (4). The cecum of GF mice is ~10-fold larger in volume than that of normal mice (7). In addition, GF mice have relatively small spleens and reduced Peyer's patches (8). Our investigation focused on the terminal ileum (the final section of the small intestine) because it is a site of intense host-microbiota interaction frequently affected in humans during intestinal infections and autoimmune and inflammatory bowel disease (9–11). The terminal ileum also exhibits a response to the microbiota that is stronger than that seen with other locations of the intestine, and the terminal ileum displays concordant changes in mRNA and cytokine levels for the key T cell subsets (12). At the cellular level, the terminal ileum is highly heterogeneous; it is composed of enterocytes, goblet cells, Paneth cells, enteroendocrine cells, microfold cells, stem cells, fibroblasts, blood and lymphatic vessels, lymphoid nodules, immune cells, smooth muscle, adipose tissue, and neurons; and contains very abundant Peyer's patches (13). We have previously shown that the ilea of GF mice have greatly reduced immune cell populations (14).

System-level studies comparing GF, conventionally raised (C), and conventionalized animals have focused on gene expression analysis using DNA microarrays and next-generation sequencing. Only a small number of studies have used proteomics to compare GF, conventional, and conventionalized animals. The metaproteome of the microbiota has been investigated (15–19). Proteomics was used to compare the intestines of conventional, monoassociated, and GF pigs (20, 21), to study the ceca of GF and conventionalized chickens (22), to study the liver and proximal colon of GF and conventionalized mice (23), to compare the liver and kidneys of conventional, GF, and antibiotics-treated mice (24), and to examine the effect of both gut and eye microbiota on immunity to ocular infections (25). Additionally, characteristics of histone modification within the proximal colon, liver, and white adipose tissue were compared between conventional, conventionalized, and GF mice (26). However, proteomics has never before been used to compare the gastrointestinal tracts of conventional and GF mice.

Here we quantitatively compared the ilea of conventional and GF mice at both the transcriptomic and proteomic levels and performed a comprehensive, comparative pathway analysis. Numerous studies have identified host sex-mediated (27–29) and strain-mediated (30–32) differences in host-microbiota interactions, so we included female and male mice from two strains (BALB/c and C57BL/10A) in our study. We discovered pathway differences between the GF and conventional mice at both the transcript and protein levels. The effects of the microbiota on female and male mice were similar, but the magnitude of the responses was larger for the C57BL/10A mice than for the BALB/c mice. We were able to conclude that transcript-level analysis is not sufficient to predict protein levels and their influence on many specific cellular signaling pathways. We also performed a meta-analysis (M) using previously published transcriptomics data sets from comparisons of conventional and GF mouse whole-tissue intestine samples to broaden our analysis to more-diverse microbiota effects on hosts.

## RESULTS AND DISCUSSION

**The ileal transcriptome and proteome adapt to the microbiota.** To examine the effect of the microbiota on the mouse ileum, we performed a comparative analysis at



**FIG 1** Study overview. (A) Terminal ilea from conventional and GF mice were quantitatively compared to produce the "ileum data set." Transcriptomics was used with DNA microarrays to produce the "transcriptome data subset," and the same ileum samples were also quantitatively compared using shotgun mass spectrometry to produce the "proteome data subset." In addition, the "meta-analysis data set" was produced by analyzing microarray data from previous publications (the meta-analysis did not use any of the ileum data set). Genes significantly affected by germ status were analyzed using Ingenuity pathway analysis. GAS, gamma interferon (IFN- $\gamma$ )-activating sequence; ISRE, IFN-stimulated response element; TC-PTP, T-cell-protein tyrosine phosphatase. (B) The transcriptome data subset was analyzed using an HCA (both rows and columns were clustered; gray = missing value). Each row depicts a microarray probe that was significantly affected by germ status ( $n = 3,817$ ), and each column depicts an array. Each array was used to analyze two ilea (one conventional and 1 GF; sex and strain matched). Each standard deviation ( $\sigma_{st}$ ) was calculated across the columns independently for each strain using the  $\log_2$ -transformed GF/C ratios. F, female; M, male. (C) The proteome data subset was analyzed using an HCA (only rows were clustered; gray = missing value). Each row depicts a protein group that was significantly affected by germ status ( $n = 242$ ), and each column depicts a mouse. Each z score was calculated across the columns independently for each strain using the  $\log_2$ -transformed abundance values.

the transcriptomic and proteomic levels (Fig. 1A). The abundances of many gene products were statistically significantly affected by germ status (3,063 and 219 genes in the transcriptome [ $\Pi$ ] and proteome [ $P$ ] data subsets, respectively). We treated the germfree status as the perturbed state and therefore compared GF mice to conventional mice throughout the analysis, defining “upregulated” (up) as corresponding to GF/C ratio values of  $>1$  and “downregulated” (down) as corresponding to GF/C ratio values of  $<1$ .

The ileum transcriptome data were analyzed to discover transcripts that were affected by germ status (see Tables S1 and S2 in the supplemental material). A hierarchical cluster analysis (HCA) partitioned the probe-level data into two clusters: upregulated probes (approximately the top two-thirds), and downregulated probes (approximately the bottom third) (Fig. 1B). Most of the probes were also affected by strain. Specifically, germ status had a stronger effect on the transcriptome of C57BL/10A mice than on that of BALB/c mice, and these effects were largely sex independent (Fig. 1B). The ileum transcriptome data were also analyzed using InfernoRDN to discover transcripts affected by sex and/or strain (see Fig. S1B in Text S1 in the

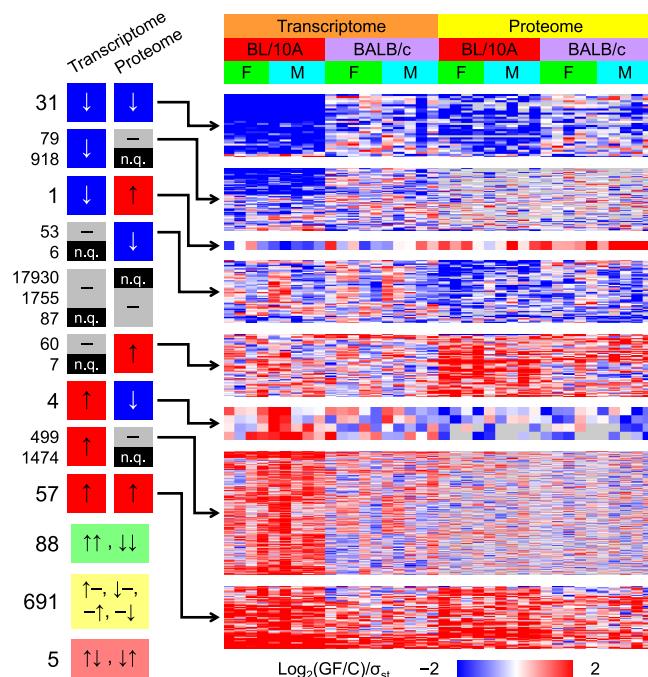
supplemental material) (Table S3). The effect of mouse strain on the transcriptome was significant and largely independent of germ status (see Fig. S1A and B in Text S1). These results are consistent with earlier discoveries of strain-mediated differences in transcriptomes and proteomes of animal tissues. The hepatic transcriptome and proteome displayed significant variance across 97 mouse strains (33), and a similar result was found in a multi-omic study of liver from two rat strains (34). Additionally, numerous studies have identified host strain-mediated differences in host-microbiota interactions in the composition of the microbiota across 113 mouse strains (30); in Paneth cell population, antimicrobial peptide composition, and microbiota composition (31); or in microbiota composition and the colonic mucosa transcriptome (32).

The ileum proteome data were analyzed to discover protein groups that were affected by germ status, sex, and/or strain (Table S4). An HCA was used to analyze the protein groups significantly affected by germ status, and the rows partitioned into two clusters of roughly equal sizes (Fig. 1C): a downregulated (top half) protein group and an upregulated (bottom half) protein group. In addition to being affected by germ status, some of these protein groups were also affected by mouse strain. As with the transcriptome data (Fig. 1B), germ status had a stronger effect on the C57BL/10A proteome than on the BALB/c proteome. Interestingly, the effect was milder in the proteome than in the transcriptome.

HCAs were also performed using the set of protein groups that were significantly affected by the other experimental parameters (see Fig. S1C and D in Text S1). As with the transcriptome data (see Fig. S1A and B in Text S1), mouse strain had a strong effect on the ileum proteome. We discovered that 63 proteins were both significantly affected by germ status and significantly unaffected by both sex and strain. An HCA of these proteins was performed (see Fig. S1D in Text S1), and the rows partitioned into two clusters of roughly equal sizes: an upregulated protein group and a downregulated protein group. We observed that the highly expressed genes were more frequently identified and quantified at the protein level and were more frequently classified as significantly affected by germ status at both the transcript and protein levels (see Fig. S2 in Text S1). Our proteome data subset was compared to a quantitative shotgun proteomics data set from a previous report (10 male GF C57BL/6 mice; at 57 to 62 days of age, 5 were conventionalized and the other 5 were mock conventionalized; proximal colons were harvested 14 days later) (23). The genes that were significantly affected by germ status within both sets of data were used for a correlation analysis, and the correlations were statistically significant (see Fig. S3A in Text S1).

The significantly affected genes in the transcriptome and proteome data subsets were tallied and analyzed using HCAs (Fig. 2). Most of the genes that were quantified within both data subsets and significantly affected within at least one of them were unaffected within the other. However, among the genes that were significantly affected within both data subsets, the effects were much more likely to be concordant than discordant. Data from a correlation analysis showed a highly significant correlation (see Fig. S3B in Text S1). This observation is consistent with numerous reports of moderate correlation of transcriptome and proteome abundance ratios from analyses of perturbed versus control samples (35). Germ status had a relatively strong effect on the C57BL/10A mice compared to the BALB/c mice, similarly to the results depicted in Fig. 1B and C.

**Adaptation of the ileum to the microbiota primarily affects metabolism and the immune system.** The overlapping genes from the transcriptome and proteome data that were significantly affected by germ status were used to produce interset concordant and discordant gene sets (described in detail in Materials and Methods; a total of 30 gene sets were produced). Ingenuity pathway analysis (IPA) was used to analyze these sets of genes to discover enriched biological pathways and other gene annotations (a total of 428 biological pathways were significantly enriched [gene annotation enrichment false-discovery rate  $\{q_e\} = <0.1\}$  in at least 1 of the 30 analyses; Tables 1, S5, S6, and S7). The pathway annotations overlapped more than the functional annotations (see Fig. S4 in Text S1). The microbiota caused upregu-



**FIG 2** Concordance across the transcriptome and proteome data subsets. Genes were either upregulated due to germ status (red,  $GF/C > 1$ ,  $q_{\min} < 0.1$ ), downregulated (blue,  $GF/C < 1$ ,  $q_{\min} < 0.1$ ), unaffected (gray,  $q_{\min} \geq 0.1$ ), or unquantified (n.q.; black). Tallies were used to quantify the overall levels of concordance (light green) and discordance (light yellow and red). The overlapping, significantly affected genes were analyzed using HCAs (only rows were clustered; gray = missing value). Each row depicts a gene, and each column depicts a pair of mice (1 GF, one conventional, sex and strain matched, from one microarray; same pairing for the proteome and transcriptome columns). Each standard deviation ( $\sigma_{\text{st}}$ ) was calculated across the columns independently for each data subset and mouse strain using the  $\log_2$ -transformed GF/C ratios.

lation of many immune system pathways in the conventional mice, and most of this was transcriptome-proteome concordant. It is likely that the microbiota causes an upregulation of intestinal immune system pathway gene expression in nonimmune cells. Also, we have previously reported that the microbiota causes an increase in immune cell populations in the terminal ileum (14), and this was likely a major cause of the observed upregulation. The microbiota also caused downregulation of many metabolic pathways in the conventional mice. Some of this dysregulation effect was transcriptome-proteome concordant, and some was discordant. The microbiota is known to perform numerous metabolic functions, and its absence likely causes the host to upregulate some of the corresponding pathways.

**TABLE 1** Tally of the biological pathways that were significantly affected by germ status<sup>a</sup>

Gene set	Gene set description	No. of upregulated pathways ( $GF/C > 1$ )	No. of downregulated pathways ( $GF/C < 1$ )	No. of pathways showing intraset discordance
T	T (this study)	16	63	17
P	P (this study)	30	38	2
M	T (previous reports)	89	59	236
Concordant[T,P,M]	Concordant in T, P, and M	28	8	0
Concordant[T,P]	Concordant in T and P	39	32	1
Concordant[T,M]	Concordant in T and M	21	90	19
Discordant[T≠P]	In T and/or P; discordant between them	4	9	0
Discordant[T\P]	In T; discordant compared to P	4	0	0
Discordant[P\T]	In P; discordant compared to T	21	16	0
Invariant[P]	In P; unaffected by sex and strain	21	4	0

<sup>a</sup>T, transcriptome; P, proteome; M, meta-analysis.

Seventy-nine pathways were downregulated in the transcriptome and/or proteome data subsets, and many were immune system pathways. Twenty-two pathways were downregulated in both subsets, and almost all were immune system pathways such as the antigen presentation pathway (Fig. 3A). Forty-three pathways were upregulated in the transcriptome and/or proteome, and almost all were metabolism pathways. Only three pathways were upregulated in both: the glutathione-mediated detoxification pathway, the serotonin degradation pathway, and the xenobiotic metabolism signaling pathway. Thirty-two pathways were downregulated in the concordant[T,P] gene set, and most were immune system pathways. Thirty-nine pathways were upregulated in the concordant[T,P] gene set, and almost all were metabolism pathways.

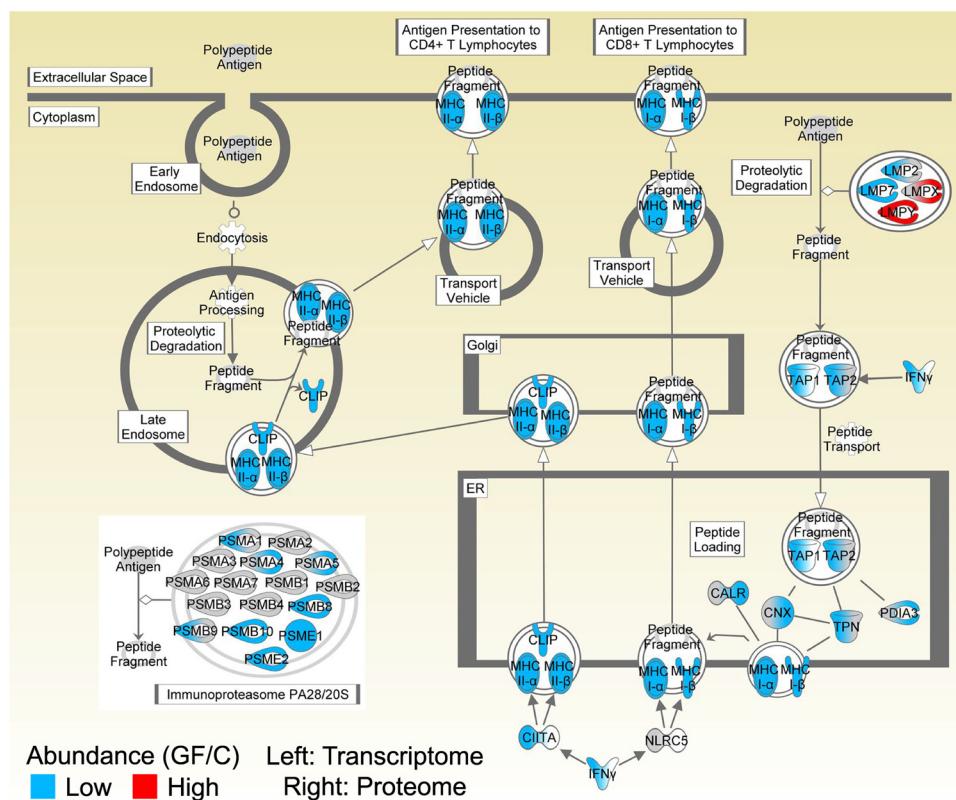
Network analyses of the significantly enriched biological pathways were performed. Each used 1 of the 30 sets of genes significantly affected by germ status. For each gene set, five networks were examined (the minimum number of overlapping significantly affected genes per edge examined ranged from one to five). Network analysis of the proteome data subset (see Fig. S5A in Text S1) and of the concordant[T,P] gene set (Fig. 3B) resulted in two clusters. One was principally composed of downregulated immune system pathways, and the other was principally composed of upregulated metabolic pathways. Network analysis of the invariant[P, up] gene set produced a cluster of upregulated metabolic pathways (see Fig. S5B in Text S1) similar to the proteome and concordant[T,P] network results. This network might constitute the core set of pathways that are normally resistant to perturbation but are nevertheless altered by the microbiota.

**Ideal pathway adaptation to the microbiota is partially transcriptome-proteome discordant.** Transcriptome-proteome concordance is consistent with regulation occurring mainly at the transcriptional level. In contrast, gene products being significantly affected at the transcriptome level and not at the proteome level (discordant[T\P]), or the reverse (discordant[P\T]), would mean that regulation occurs principally via posttranscriptional or cotranslational regulation, targeted protein degradation, or protein secretion. There were 5 instances of opposite regulation at the transcriptomic and proteomic levels, 113 instances of significantly affected proteins and unaffected transcripts, and 578 instances of the reverse (Fig. 2). These data were analyzed to discover differentially regulated biological pathways.

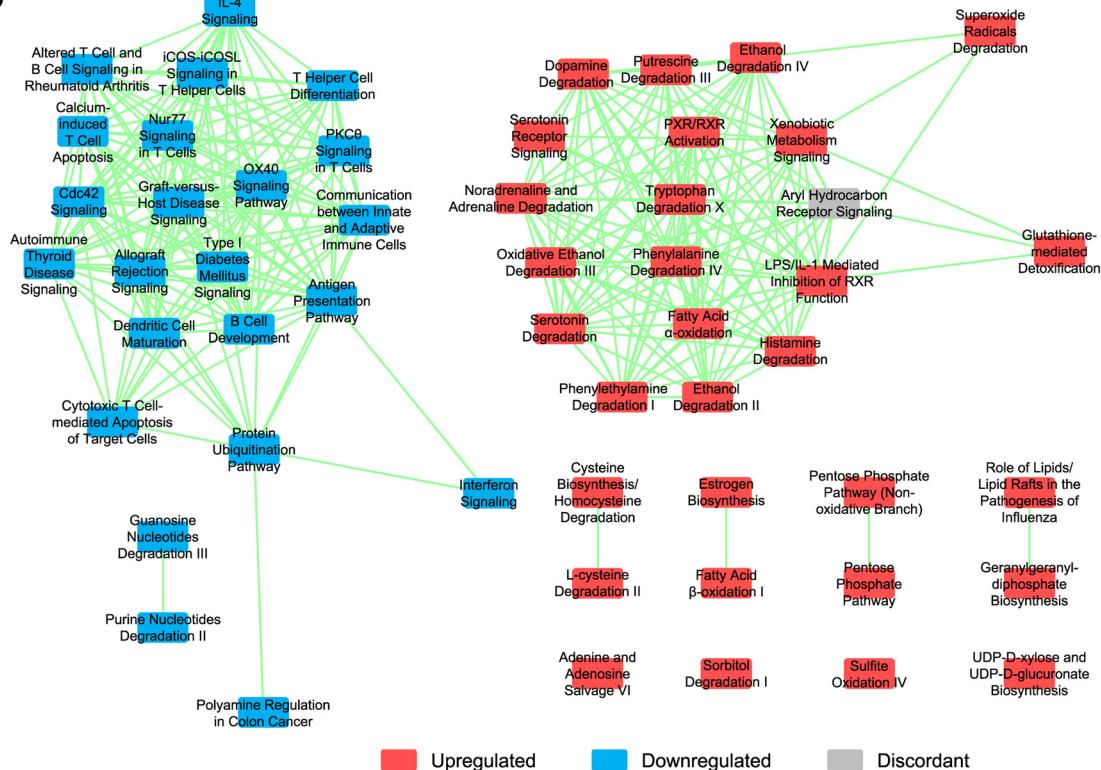
Four pathways were significantly enriched in the discordant[T\P] gene set, and all four were upregulated. Two of these pathways, the eukaryotic initiation factor 2 (eIF2) signaling and mechanistic target of sirolimus (mTOR) signaling (see Fig. S6 in Text S1) pathways, highly overlapped. These pathways included much of the protein translation machinery. The significantly affected genes annotated with the Ingenuity biofunctions "Processing of RNA" and "Translation" were analyzed, and a similar pattern was discovered (intratranscriptome concordancy and transcriptome-proteome discordancy; see Fig. S7 in Text S1). All of the significantly affected eIF and ribosomal protein (RP) genes were found to be upregulated in the transcriptome data subset, and almost all were transcriptome-proteome discordant (Fig. 4A). Inset concordancy of all of the eIF, 40S RP, and 60S RP transcript- and protein-level data was analyzed, and it was discovered that the eIF, 40S RP, and 60S RP transcripts were upregulated, the eIF proteins were unaffected, and the 40S and 60S RPs were downregulated (Fig. 4C). Generally, such differences may result from changes in posttranscriptional regulation, mRNA export and localization, cotranslational regulation, protein secretion, or targeted protein degradation (35).

All RP genes and many eIF genes are known to be posttranscriptionally regulated via their 5'-terminal oligopyrimidine (5'TOP) tract (36, 37). In the inactive state, LARP1, TIA1, or TIAL1 is bound to the 5'TOP tract, which represses translation (38–40). LARP1 binding increases transcript stability, and TIA1 and TIAL1 binding targets the transcript to stress granules for mRNA triage. Mechanistic target of sirolimus (mTOR) complex 1 activation antagonizes repression of translation by LARP1, TIA1, and TIAL1. Posttranscriptional regulation by LARP1, TIA1, and/or TIAL1 may have caused the transcriptome-proteome discordancy that we observed. In addition, it is also possible that the

A



B



**FIG 3** (A) The antigen presentation pathway was downregulated in the GF mice. Each pathway node was dually colored using the transcriptome (left) and proteome (right) data subsets. The immunoproteasome complex is from the protein ubiquitination pathway. The statistical significance data were  $q_e$  (transcriptome, downregulated) =  $1.0E-14$ ,  $q_e$  (proteome, downregulated) =  $6.6E-10$ ,  $q_e$  (meta-analysis, downregulated) =  $1.9E-3$ ,  $P_c$  (transcriptome) =  $3.4E-3$ , and  $P_c$  (meta-analysis) =  $3.9E-2$ . The proteome node-level intraset concordancy data were not statistically significant, but they were close ( $P_c$  =  $6.5E-2$ ; significance level = 0.05). CALR, Cysteine/Leucine/Alanine Rich Protein; CLIP, CLIP protein; ER, endoplasmic reticulum; IFNγ, interferon-gamma; MHC, major histocompatibility complex; PSMA1–PSME2, proteasome subunits; TAP1/TAP2, transporter associated with antigen processing.

(Continued on next page)

microbiota caused a decrease in transcript stability via LARP1, TIA1, and/or TIAL1, resulting in the observed concordant downregulation of these transcripts in the conventional mice. A second 5' untranslated region tract called the pyrimidine-rich translational element is also involved in mTOR-mediated translation upregulation of many RP transcripts (41), and it is possible that regulation at this tract was involved.

The other two discordant[T\P] pathways, the OxP and mitochondrial dysfunction pathways, also highly overlapped. The OxP pathway nodes were intratranscriptome concordant (all 31 were upregulated in the GF ilea), and they were all transcriptome-proteome discordant except for Atp5a1 (Fig. 4B). Mouse gene expression of OxP genes was found to be coregulated using a coexpression network analysis (data not shown), as previously reported (42). Inset concordancy of all of the OxP transcript- and protein-level data was analyzed, and it was discovered that the OxP transcripts were upregulated and that the OxP proteins were unaffected (Fig. 4C).

Gene expression of OxP genes across 61 mouse tissues was previously reported to be coregulated (42). As with the eIF and RP genes, it is possible that the microbiota affects the transcription rate and/or transcript stability of OxP genes, resulting in the observed concordant downregulation of these transcripts in the conventional mice. Interestingly, we previously reported that transcripts encoding ribosomal and OxP proteins were significantly downregulated in the mouse ilea in response to treatment with either antibiotics or antibiotic-resistant microbes and that these effects were mediated by virulence factors and quorum sensing by antibiotic-resistant bacteria (14).

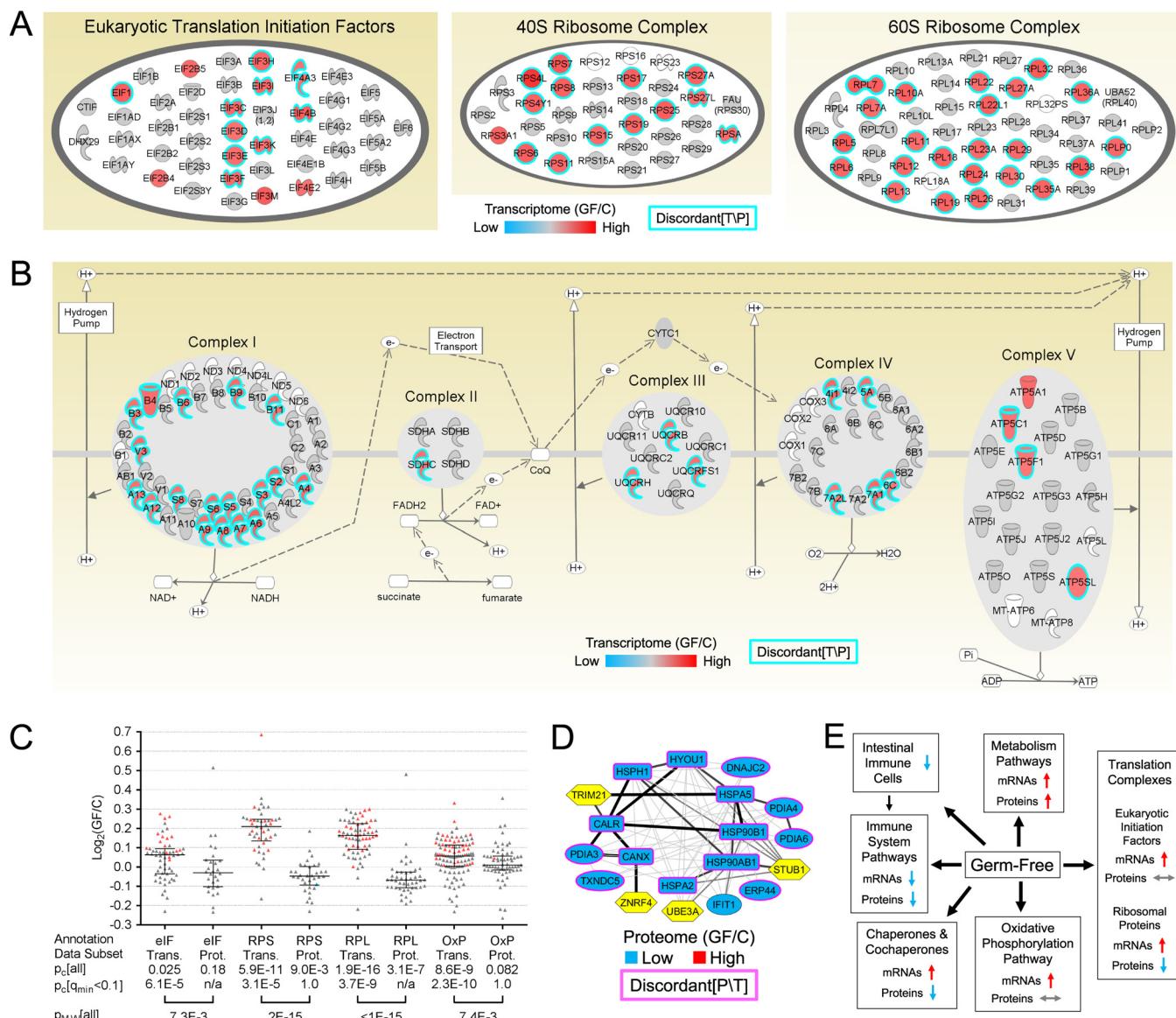
Thirty-seven pathways were significantly enriched in the discordant[P\T] gene set. Twenty-one were upregulated, 16 were downregulated, and none were inset discordant. A network analysis of the 21 upregulated pathways resulted in a cluster of upregulated metabolic pathways (see Fig. S8A in Text S1). Thus, the proteomic analysis discovered many upregulated metabolism pathway elements that were missed by the transcriptomic analysis. Interestingly, of the 21 upregulated discordant[P\T] pathways, 12 were also upregulated invariant[P] pathways.

Fifteen of the 16 downregulated discordant[P\T] pathways contained chaperone proteins that were significantly affected by germ status. Therefore, the interset concordancy and inset concordancy/discordancy of the chaperones and cochaperones in the transcriptome and proteome data subsets were analyzed. There were 310 (co)chaperone genes corresponding to 428 microarray probes in the transcriptome; 206 probes were upregulated, 222 probes were downregulated, and the results were found not to be significantly inset concordant ( $P_c = 0.47$ ). Forty-three (co)chaperone genes (44 microarray probes) were significantly affected by germ status in the transcriptome data subset. Of the 44 significantly affected probes, 31 were upregulated and 13 were downregulated, which represented significant inset concordance ( $P_c = 9.6E-3$ ). All 43 of the (co)chaperone genes affected at the transcript level were unaffected at the protein level, with one exception: Ifit1, an interferon-induced cochaperone, was downregulated at both levels.

In the proteome data subset, there were 107 (co)chaperone genes corresponding to 100 nonredundant quantified protein groups. Of the 100 nonredundant quantified protein groups, 37 were upregulated and 63 were downregulated, representing significant concordance ( $1.2E-2$ ). Of the 100 nonredundant quantified protein groups, 15 were significantly affected by germ status, and all 15 were downregulated, representing significant inset concordance ( $P_c = 6.1E-5$ ). Many of these 15 proteins have been reported to directly associate with each other, and all 15 genes were unaffected in the

### FIG 3 Legend (Continued)

calreticulin; CLIP, class II-associated invariant chain peptide; CNX, calnexin; ER, endoplasmic reticulum; MHC, major histocompatibility complex; TPN, tapasin. (B) The discordant[T,P] network of biological pathways formed two independent clusters related to the immune system (left) and metabolism (right). Each node depicts a significantly enriched biological pathway, and each edge depicts at least one instance of overlapping genes that were significantly affected by germ status. Pathways were classified as upregulated if the annotation enrichment was maximally significant by analyzing only the upregulated genes and likewise for the downregulated pathways. IL-4, interleukin-4.



**FIG 4** Inset concordancy and transcriptome-proteome discordancy of gene products related to protein translation, OxP, and chaperone proteins. (A) The eIF and RP complexes displayed significant intraset concordancy ( $P_c$  [transcriptome] = 8.9E-16) and transcriptome-proteome discordancy (cyan outlines). (B) The OxP pathway also displayed significant intraset concordancy [ $q_e$  (transcriptome, up) = 3.4E-6,  $P_c$  [transcriptome] = 9.3E-10] and transcriptome-proteome discordancy (cyan outlines). (C) All of the following categories of transcriptome and proteome data related to the four annotations were analyzed: eIFs, the 40S ribosomal complex (RPS), the 60S ribosomal complex (RPL), and the OxP pathway. Each triangle depicts a microarray probe (transcriptome [Trans.] data subset) or a protein group (quantified nonredundantly; proteome [Prot.] data subset). Red and blue triangles depict significantly increased and decreased gene products, respectively (gray triangles = not significantly affected). The lines indicate each median and interquartile range. Each binomial test either used all of the data or was restricted to just the significantly affected data, as indicated. n/a, not applicable. (D) A protein-protein interaction network was prepared using all of the chaperones (rectangles) and cochaperones (ovals) that were significantly affected by germ status within the proteome data subset, and the intraset concordance was significant ( $P_c$  = 6.1E-5). Also included are the E3 ubiquitin ligases (yellow hexagons) that have been reported to associate with chaperones and typically target misfolded/unfolded proteins. Edge weight data depict the protein-protein interaction confidence (increasing weight values depict low, medium, high, and very high confidence). (E) Summary of the effects on the GF mouse intestine. The decrease in immune cell populations was reported previously by Morgan et al. (14).

transcriptome data subset except for the Ifit1 gene (Fig. 4D). Thus, the (co)chaperone transcripts and proteins were affected oppositely (transcripts upregulated and proteins downregulated). Consistently, the full (co)chaperone transcriptome and proteome data subsets were found to be significantly discordant (Mann-Whitney  $P$  value [ $P_{MW}$ ] = 5.0E-3). To identify concordance within the (co)chaperone data more precisely, the chaperone and cochaperone data were analyzed separately. The chaperones were not concordantly affected at the transcript level, the cochaperones were concordantly

upregulated at the transcript level, and the (co)chaperones were concordantly downregulated at the protein level (see Fig. S8B in Text S1). This finding is consistent with numerous reports that the microbiota causes upregulation of host intestinal chaperone protein abundance, principally in the epithelium, and that the level of upregulation is typically highest in the cells that are closest to the microbiota (43, 44). The innate immune response is known to cause upregulation of heat shock protein gene expression, at least partially as a consequence of cross talk between the innate immune response and the unfolded protein response (43, 45).

All but 1 of the 15 (co)chaperones downregulated at the proteome level were not affected at the transcriptome level (Fig. 4D). This may have resulted from the microbiota affecting (co)chaperone posttranscriptional regulation, protein secretion, and/or protein degradation. Cellular export of chaperones has been reported, though the mechanism is unclear (46, 47), and it is possible that the microbiota somehow downregulates (co)chaperone export from cells of the mouse ileum, resulting in the observed discordancy. Some chaperones are known to associate with the E3 ubiquitin ligases that typically target unfolded/misfolded client proteins, resulting in their degradation (48).

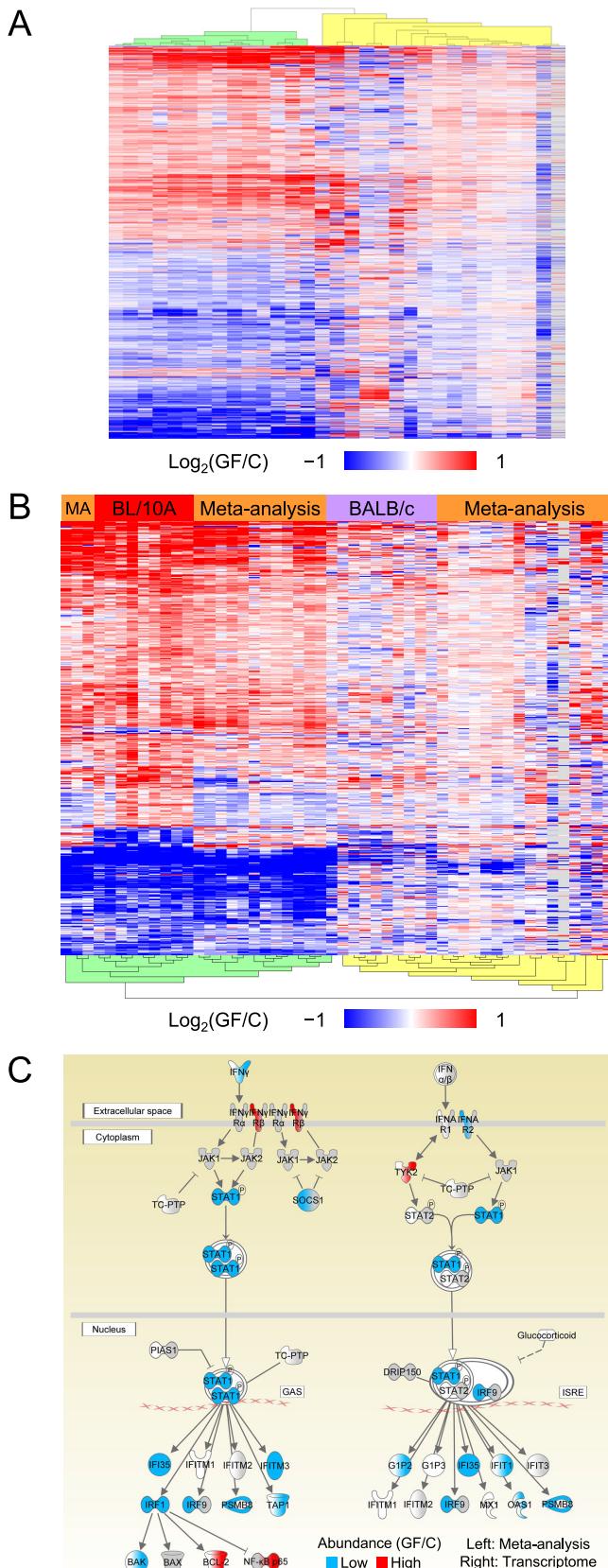
We noted that some (co)chaperones were upregulated in the GF ilea in the transcriptome data subset and were unaffected in the proteome data subset. It is possible that transcription, translation, and protein degradation of these (co)chaperones were upregulated in the GF mice, resulting in increased abundance in the transcriptome and no significant net change in the proteome. We also noted that Stub1 itself was upregulated in the transcriptome data subset but was unaffected in the proteome data subset. STUB1 is known to autopolyubiquitinate itself (49), which may result in its degradation, and this may explain the discordancy that we observed. Intriguingly, PARK2 has a key role in derepressing translation of O<sub>x</sub>P transcripts, and it is also a chaperone-associated E3 ubiquitin ligase that typically targets unfolded/misfolded client proteins, resulting in their degradation (48). It is possible that PARK2 has a dual role in the host response to the microbiota, linking derepression of translation with chaperone-associated ubiquitination.

A summary of the ileum data set results is provided as Fig. 4E.

**Adaptation of the intestinal transcriptome to the microbiota varied significantly across globally diverse murine and microbiota populations.** A meta-analysis was performed to discover host genes that were affected by the microbiota (Fig. 1A). The meta-analysis used the microarray data originating from a globally diverse set of samples that varied in microbiota population, host strain, host sex, tissue location within the intestine, and sample preparation protocol (Table S8). The abundances of 2,132 genes were statistically significantly affected by germ status (Table S9). This is the first report of such an analysis. Ideally, a corresponding meta-analysis using proteomics data would also have been performed, but our proteomics comparison of conventional and germfree mouse intestinal tissue is the first of its kind.

An HCA of the meta-analysis data partitioned the rows into two clusters of roughly equal sizes: one of upregulated genes and the other of downregulated genes (Fig. 5A). The HCA also partitioned the columns (depicting comparison group pairs) into two clusters. Germ status had a relatively strong effect on the leftmost cluster (14 columns; dendrogram highlighted green in Fig. 5A). This cluster contained mice from two studies: male C57BL/6J mice (jejunum, ileum, and colon; conventionalized for 4, 8, 16, and 30 days; GEO dataset accession no. [GSE32513](#)), and female C3H/HeN mice (ileum, conventional and conventionalized using a mouse microbiota for 8 and 60 days; GEO dataset accession no. [GSE18056](#)). From left to right, the columns of the green cluster Fig. 5A represent 11 GSE32513 columns (3 colon columns followed by 4 ileum columns and 4 jejunum columns) followed by 3 GSE18056 columns. The colon 30-day GSE32513 column data were aberrantly absent from the green cluster.

In comparison to the green cluster, germ status had a relatively weak effect on the other cluster (17 columns; dendrogram highlighted in yellow in Fig. 5A). It is likely that six columns of the yellow cluster were weakly affected because they originated from



**FIG 5** The transcriptome data subset and meta-analysis data set were partially concordant. (A) An HCA of the meta-analysis data set resulted in clustering of relatively strongly and weakly affected mice (indicated by the green and yellow dendrogram highlighting, respectively). Each row depicts a significant gene.

(Continued on next page)

comparisons of GF versus conventionalized mice at early time points (days 1 and 2; all from the GSE32513 study). There was no obvious reason for the relatively weak effect shown in the other 11 columns. However, two columns of the yellow cluster (columns 4 and 5, counting from the left) represented data from mice that had been conventionalized using a human microbiota (GSE18056 study; 8 and 60 days, respectively), and it is possible that human-mouse microbiota differences somehow caused the relatively weak effect. We noticed that one column of the yellow cluster (column 7, counting from the left) was the colon 30-day GSE32513 column mentioned above and that the bottom half of this column roughly matched the corresponding data in the green cluster. In summary, the meta-analysis revealed that the mice were affected by germ status at two levels: a relatively low level and a high level. It is possible that this dissimilarity was due to the different mouse strains and/or different microbiota populations examined. The pattern of relatively strong and weak effects in the meta-analysis data set was consistent with the strain-dependent patterns seen in both the transcriptome and proteome data subsets (Fig. 1B and C).

An HCA of the significantly affected genes in the transcriptome data subset and the meta-analysis data set was performed (Fig. 5B), and the row and column partitioning was similar to that described above (Fig. 1B and C, 2, and 5A). With a single exception (the aberrant colon 30-day GSE32513 column), the strong/weak partitioning of the meta-analysis columns was the same in Fig. 5A and B. Similarly, the strong/weak partitioning of the transcriptome columns was the same in Fig. 1B and 5B except for a single BALB/c column. A correlation analysis was performed, and the correlation was highly significant (see Fig. S9A in Text S1). Reassuringly, the concordant changes between the transcriptome data subset and meta-analysis data set validated each other. The overall concordance across the transcriptome, proteome, and meta-analysis sets of data were examined (see Fig. S9B in Text S1). A significantly affected gene in one set was most likely to be unaffected or unquantified within the other two sets. However, if a gene was significantly affected within multiple sets, the effects were more likely to be concordant than discordant.

**Adaptation of the intestine to the microbiota affects hundreds of biological pathways.** Three hundred eighty-four Ingenuity pathways were significantly enriched in the meta-analysis data set (Table 1). Two hundred sixty-eight (70%) of the pathways were not enriched in the transcriptome or proteome data subsets (28 were downregulated, 60 were upregulated, and 180 were intraset discordant). Thus, the meta-analysis greatly broadened the discovery of pathways affected by germ status. As with the ileum data set, some of these pathways were downregulated immune system pathways and upregulated metabolic pathways, but the pathways varied greatly overall.

Twenty-eight pathways were downregulated in both the transcriptome and meta-analysis sets of data, and almost all were immune system pathways such as the interferon signaling pathway (Fig. 5C). Only three pathways, the glutathione-mediated detoxification pathway, the serotonin degradation pathway, and the xenobiotic metabolism signaling pathway, were upregulated in both the transcriptome analysis and

#### FIG 5 Legend (Continued)

cantly affected gene ( $n = 2,132$ ), and each column depicts an intrastudy pair of groups of mice (a “comparison group pair” consisting of a conventional group and a GF group) (both rows and columns were clustered; gray = missing value). (B) An HCA of the intersection of the transcriptome data subset and the meta-analysis (MA) data set resulted in clustering of strongly and weakly affected mice (indicated by the green and yellow dendrogram highlighting, respectively). Each row depicts a significantly affected gene ( $n = 530$ ), each transcriptome column depicts a pair of mice (1 GF and one conventional from one microarray; sex and strain matched; red and lavender bars), and each meta-analysis column depicts an intrastudy “comparison group pair” (orange bars) ( $q_{min}$  [germ status] < 0.1 required for both sets of data [concordance of GF/C ratios was not required]; both rows and columns were clustered; gray = missing value). (C) The interferon signaling pathway was downregulated in the GF mice. Each pathway node was dually colored using GF/C ratios from the sets of meta-analysis (left) and transcriptome (right) data. The statistical significance data were as follows:  $q_e$  (transcriptome, downregulated) = 1.6E-6,  $q_e$  (proteome, downregulated) = 4.6E-2,  $q_e$  (meta-analysis, downregulated) = 1.7E-3,  $P_c$  [meta-analysis] = 3.9E-2. The transcriptome node-level intraset concordancy was not statistically significant, but it was close ( $P_c$  = 5.7E-2; significance level = 0.05).

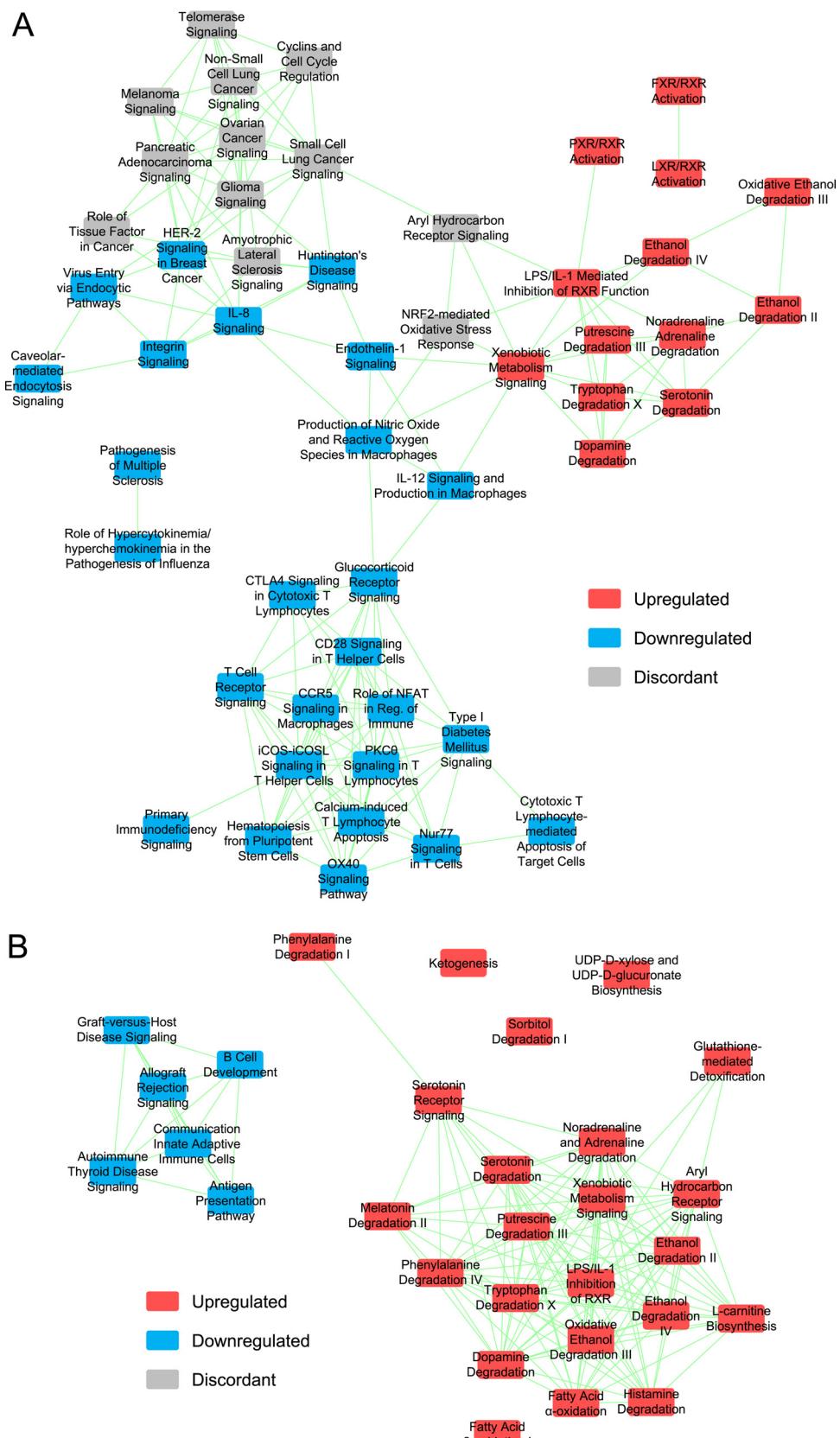
meta-analysis, and these three were also upregulated in the proteome. Ninety pathways were downregulated in the concordant[T,M] gene set, and most were immune system pathways. Twenty-one pathways were upregulated in the concordant[T,M] gene set, and almost all were metabolism pathways.

Fifteen pathways were downregulated in the sets of transcriptome, proteome, and meta-analysis data, and almost all were immune system pathways such as the antigen presentation pathway (Fig. 3A) and the interferon signaling pathway (Fig. 5C). As mentioned above, three pathways were upregulated in all three sets of data. Eight pathways were downregulated in the concordant[T,P,M] gene set, and almost all were immune system pathways. Twenty-eight pathways were upregulated in the concordant[T,P,M] gene set, and almost all were metabolism pathways. Notably, the aryl hydrocarbon receptor signaling pathway was significantly enriched and intraset discordant in all three sets of data.

Network analysis of the concordant[T,M] gene set resulted in three clusters (Fig. 6A). The top-right cluster in the figure panel primarily consists of upregulated metabolic pathways, the bottom-center cluster primarily consists of downregulated immune system pathways, and the third cluster (upper left) partially consists of downregulated immune system pathways. Interestingly, the upper half of the third cluster was principally composed of intraset discordant cancer pathways. In a related result, the colorectal cancer metastasis pathway was enriched and intraset discordant in the meta-analysis data set ( $q_e$  [up + down] = 1.6E-12). This pathway was not enriched in either the transcriptome data subset or the proteome data subset. Thus, the microbiota affected numerous cancer pathways at the transcript level. Network analysis of the concordant[T,P,M] gene set resulted in two clusters (Fig. 6B). One was principally composed of downregulated immune system pathways, and the other was principally composed of upregulated metabolic pathways. This network might constitute the core set of pathways that are most broadly and strongly affected by the microbiota.

**Conclusions.** The microbiota is an important factor for human health and biomedical research, and GF animals are a valuable tool for discovering microbiota effects on the host. A hypothetical model of some of the results is summarized in Fig. S10 in Text S1, and compendia of illustrations of the affected pathways have been prepared and are available upon request. So far, most system-level studies comparing conventional and GF animals have focused on transcriptomic analyses. Although it has been found that protein abundance under steady-state conditions is primarily determined by transcript abundance, the spatial and temporal variations of mRNAs, the local availability of resources for protein biosynthesis, posttranscriptional regulation, and targeted protein degradation strongly influence the relationship between protein levels and their coding transcripts. Transcript fluctuations can be buffered at the level of protein concentration; consequently, analysis of transcript levels alone is generally not sufficient to predict protein levels in many scenarios (50), including during host-microbiota interactions. To address this, we performed the first proteomic comparison of the gastrointestinal tract of conventionally raised and GF mice alongside a corresponding transcriptomic analysis. Multiple gene products and hundreds of biological pathways were found to be affected by germ status in the mouse terminal ileum, and protein-level variance often did not reflect a similar change in the corresponding transcript level.

While transcriptomics analysis of small samples (e.g., from laser capture microdissection) is now routine, proteomics technology needs to continue to be researched and developed to enable deep analyses of small-mass samples. When such analyses become possible, this work will provide the foundation for the studies of cell-specific changes in the proteome caused by the microbiota. Now, on the basis of our analysis and our previous work, we can speculate that some of the changes that we observed at the proteome and transcriptome levels are caused by changes in the abundance of specific cell populations in the ileal tissue. Our previous study (14) had shown enrichment of T and B lymphocytes caused by the microbiota. However, this type of analysis



**FIG 6** Genes affected concordantly within the ileum and meta-analysis data sets formed biofunctional clusters of overlapping pathways. Each node depicts a significantly enriched biological pathway, and each edge depicts overlapping genes that were significantly affected by germ status. Pathways were classified as upregulated if the

(Continued on next page)

does not discriminate between subpopulations, which would be more interesting to analyze. Therefore, future studies using single-cell transcriptome sequencing (RNA-seq) would enable addressing this issue.

Our report demonstrates that transcript abundance cannot always reliably predict protein abundance differences caused by the perturbations of the microbiota; provides a resource to study the effect of the microbiota on the host at the functional, protein, and pathway levels; and reinforces the idea of the importance of measuring protein abundance alterations to reveal molecular mechanisms behind the phenotypic changes caused by changes in the microbiota.

## MATERIALS AND METHODS

**Mouse ilea.** This study was approved by the NIH Animal Care and Use Committee. Rederived GF and conventional mice (female and male; BALB/c and C57BL/10A strains; 6 to 8 months old) were prepared by Taconic Biosciences, Inc. (Hudson, NY). The conventional mice were housed under normal conditions and maintained using standard procedures. Six GF mice were prepared for each experimental class (two sexes, two strains, 24 GF mice in total) with the goal of having 5 mice per experimental class (two germ statuses, two sexes, two strains, 40 mice in total). Samples (e.g., feces) from the GF mice were analyzed to ensure the complete absence of any microbial contamination, and mice were discarded if any sample tested positive. This resulted in only 4 GF female C57BL/10A ilea being available. The ilea were harvested, placed into and gently flushed with ice-cold phosphate-buffered saline (PBS) to wash away microbes, segmented, and immediately frozen on dry ice. One set of the segments was used for the transcriptomics experiments, and another set was used for the proteomics experiments. This resulted in the production of data referred to here as the "ileum data set," consisting of the transcriptome and proteome "data subsets."

**Transcriptomics.** Total RNA was isolated from the ilea using Total RNeasy minikits (Qiagen, Hilden, Germany). Fluorescently (green and red) labeled cDNA was prepared using 20 µg of total RNA and 2 µl of 1 mM Cy3-dUTP and Cy5-dUTP, respectively (GE Healthcare, Chicago, IL), 1 µg of oligo(dT)<sub>12-18</sub> primer, 0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.15 mM dTTP, 10 mM dithiothreitol, 1.5 µl of 200 U/µl SuperScript II reverse transcriptase, and 0.5 µl of 20 U/µl RNase inhibitor (all from Thermo Fisher Scientific Inc., Waltham, MA, unless noted otherwise). The resulting cDNA was purified using Vivaspin 500 columns (Sartorius AG, Göttingen, Germany) (10,000 molecular weight cutoff [MWCO]). Each sample pair (red and green; 38 samples in total) was cohybridized to a gene microarray (Agilent Technologies, Santa Clara, CA) (45,220 spots, 37,558 reporter spots, 35,958 reporter probes, 60-mer probes, "4 by 44 k" slide format, design number 017138) using a MicroArray User Interface (MAUI) hybridization system (BioMicro Systems Inc., Salt Lake City, UT). A reference-free experimental design was used. Each microarray was used to analyze 1 GF-mouse sample and one conventional-mouse sample (sex and strain matched). To mitigate label effects, half of the samples of each experimental class were randomly selected to be labeled green and the other half red. The dual-color microarrays were scanned using a Microarray Scanner System, and the resulting images were processed using Feature Extraction (both Agilent Technologies). These data are available at the Gene Expression Omnibus data repository at the National Center for Biotechnology Information (GEO Platform [GPL9354](#); GEO entry identifier [GSE95650](#); <http://www.ncbi.nlm.nih.gov/geo/>).

Gene symbols were retrieved from bioDBnet (12 March 2016; <https://biobdbnet-abcc.ncifcrf.gov/>) (51) using the gene and transcript identifiers within the GEO Platform data set. This resulted in 22,861 mouse gene symbols being associated with 29,819 reporter probes. Each reporter spot foreground and background intensity value represented the corresponding median pixel intensity value. Each spot background intensity value was set to be a maximum of four times the median background intensity value (calculated across each microarray and color channel; this affected 251 of the 1,427,204 background intensity values). For each ileum sample and spot, if the net intensity (foreground – background) value was <10, it was replaced with a random number between 10 and 20 (this range was roughly consistent with the range of the background values, as the median background intensity value was 35.5; this affected 31,139 of the 1,427,204 net intensity values). For each microarray/color channel/probe, the arithmetic mean of the net intensity values was calculated across replicate spots. The resulting values were log<sub>2</sub> transformed.

A statistical analysis was performed using BRB-ArrayTools (<http://brb.nci.nih.gov/BRB-ArrayTools/>) (52) to identify probe intensity values that were significantly affected by germ status. To this end, intra-array LOWESS normalizations (span = 0.4) were performed. The resulting data were used for analyses of variance (ANOVAs) that used a statistical model designed specifically for reference-free dually

## FIG 6 Legend (Continued)

annotation enrichment was found to be maximally significant by analyzing only the upregulated genes and as downregulated if the annotation enrichment was found to be maximally significant by analyzing only the downregulated genes. (A) The concordant[T,M] network ( $\geq 4$  overlapping genes per edge) formed three clusters related to cancer (top left), metabolism (top right), and the immune system (top left and bottom center). (B) The concordant[T,P,M] network (with at least one instance of overlapping genes per edge) formed two independent clusters, one composed of downregulated immune system pathways (left) and the other of upregulated metabolic pathways (right).

labeled microarray data (53). The BRB-ArrayTools “Random Variance Model” option was used. The ANOVA model estimated array effects as well as the effects due to the eight experimental classes (two germ statuses, two sexes, two strains). The resulting data were used for *post hoc* pairwise comparison testing across the experimental classes (35,958 probes  $\times$  28 pairs of experimental classes = 1,006,824 tests), which resulted in a *q* value (i.e., a false-discovery-rate value) for each test. For each probe, the minimum *q* value ( $q_{\min}$ ) was calculated across the four relevant pairs of experimental classes (GF mice versus conventional mice, sex and strain matched). The data were filtered by requiring  $q_{\min}$  values to be  $<0.1$ . This resulted in 3,817 probes (10.6%) passing this filter, and this corresponded to 3,063 genes (13.4%). Pearson and Spearman correlation *P* values were calculated using 2-tailed *t* tests. HCAs were performed using Genesis (complete linkage clustering using Euclidean distance) (54). For each probe, the arithmetic mean of the  $\log_2$ -transformed and LOWESS-normalized GF/C (germfree/conventional) ratios (described above) was calculated across the microarrays, and the anti- $\log_2$  values determined were used as the GF/C values for the pathway analyses.

An additional statistical analysis was performed using InfernoRDN (<https://omics.pnl.gov/software/infernordn>) (55) to identify probe intensity values that were significantly affected by sex and/or strain. To this end, an interarray LOESS (locally weighted scatterplot smoothing) normalization (reference = median value, span = 0.4) was performed. The resulting data were used to perform a 3-way ANOVA (germ status, sex, and strain; an array effect was not included) for each probe. Each 3-way ANOVA produced a *q* value for each of the three effects (germ status, sex, and strain) and for each of the four interactions (germ status plus sex, germ status plus strain, sex plus strain, and germ status plus sex plus strain). For each probe, a minimum *q* value ( $q_{\min}$ ) was calculated for each experimental parameter (germ status, sex, and strain). For example,  $q_{\min}(\text{sex})$  equaled the minimum of the *q* values that involved the sex parameter [ $q(\text{sex})$ ,  $q(\text{germ status plus sex})$ ,  $q(\text{sex plus strain})$ , and  $q(\text{germ status plus sex plus strain})$ ]. The data were filtered by requiring  $q_{\min}$  values to be  $<0.1$ . This resulted in 1,180 (3.28%), 574 (1.60%), and 22,758 (63.3%) probes passing the germ status, sex, and strain filters, respectively, and this corresponded to 917 (4.01%), 462 (2.02%), and 15,784 (69.0%) genes, respectively. HCAs were performed using Genesis as described above. Note that these data were not used for pathway analyses.

**Proteomics.** The ileum samples were blocked and randomized to mitigate any potential systematic biases present during sample preparation (56). Each sample was homogenized in 500  $\mu\text{l}$  of lysis buffer (freshly prepared 100 mM HEPES-NaOH [pH 8], 8 M urea, 10  $\mu\text{M}$  bestatin hydrochloride, 10  $\mu\text{M}$  pepstatin A, PhosStop phosphatase inhibitor cocktail [F. Hoffmann-La Roche Ltd., Basel, Switzerland] [1 tablet/10 ml]). Homogenization was performed using an Omni TH tissue homogenizer equipped with 7-mm-by-110-mm “hard tissue” disposable plastic tips (Omni International, Kennesaw, GA) (25% to 50% setting for 60 s). The homogenates were bath sonicated for 10 min at room temperature, and bicinchoninic acid protein concentration assays (Thermo Fisher Scientific Inc.) were performed.

Dithiothreitol was added to 500  $\mu\text{g}$  (protein mass) of each ileum homogenate, and the samples were incubated at 60°C for 45 min to reduce protein cysteine residues (final dithiothreitol concentration = 10 mM). Iodoacetamide was added to each sample (final concentration = 50 mM), and the samples were incubated at room temperature for 20 min in darkness to alkylate cysteine residues. Each sample was diluted using 100 mM HEPES-NaOH (pH 8) such that the final urea concentration was 1 M. Protein digestion was performed by adding sequencing-grade modified trypsin (Promega Corp., Madison, WI) (1:100 [wt:wt] protein trypsin/sample) and incubating the samples at 37°C for 19 h. Each sample was acidified by adding formic acid (final concentration = 1% [vol/vol]), and solid-phase extraction was performed using Sep-Pak C<sub>18</sub> columns (Waters Corp., Milford, MA).

A reference “pool” sample for relative quantifications was prepared by pooling 5  $\mu\text{l}$  (50  $\mu\text{g}$ ) of all 39 ileum samples. Six sets of iTRAQ 8-plex reagent (AB Sciex Pte. Ltd., Framingham, MA) were used to label the samples following the manufacturer’s instructions (100  $\mu\text{g}$  of peptide mass per labeling reaction;  $\geq 1$  reference sample per iTRAQ 8-plex set). The resulting six samples underwent TiO<sub>2</sub> phosphopeptide enrichment as described previously (57). The resulting flowthrough samples (i.e., the nonphosphorylated peptides used in this study) were fractionated using strong cation exchange high-pressure liquid chromatography (HPLC) and a 2.1-mm-by-200-mm PolySULFOETHYL A column (Poly LC Inc., Columbia, MD) (5- $\mu\text{m}$  diameter; 20-nm pore size; mobile phase A = 10 mM KH<sub>2</sub>PO<sub>4</sub>-HCl [pH 2.8] with 25% [vol/vol] acetonitrile; mobile phase B = mobile phase A with 1 M KCl; 30 [run 1] or 33 [runs 2 to 6] fractions per run). The fractions were subjected to solid-phase extraction using ZipTip C<sub>18</sub> tips (Millipore Corp., Billerica, MA).

The samples were analyzed using a Nano-LC Ultra two-dimensional (2D) HPLC system (Eksigent Technologies [acquired by AB Sciex Pte. Ltd.]) coupled via nano-electrospray (nano-ES) to a linear trap quadrupole (LTQ) Orbitrap Velos mass spectrometer (Thermo Fisher Scientific Inc.). The LC system was equipped with a trap column (8-cm length, 100  $\mu\text{m}$  inner diameter [i.d.]) and a packed-tip column (15-cm length, 50  $\mu\text{m}$  i.d.), and both were packed with Halo ES C<sub>18</sub> resin (Michrom Bioreources, Inc. [acquired by Bruker Corp., Billerica, MA]) (2.7- $\mu\text{m}$  diameter, 16-nm pore size). Each LC-tandem mass spectrometry (LC-MS/MS) run included a 120-min linear gradient (3% to 35% mobile phase B), a 10-min column regeneration step (90% mobile phase B), and a 30-min column reequilibration step (3% mobile phase B) (300 nl/min flow rate; mobile phase A = 0.1% [vol/vol] formic acid-H<sub>2</sub>O; mobile phase B = 0.1% [vol/vol] formic acid-acetonitrile). The top 6 most intense ions per precursor ion scan were selected for MS/MS (both collision-induced dissociation [CID] and higher-energy collisional dissociation [HCD]). This resulted in 236 LC-MS/MS runs and 4,434,402 MS/MS spectra (not including runs of blanks and quality control [QC] standards).

The resulting spectra were analyzed using Proteome Discoverer v2.0.0.802 (Thermo Fisher Scientific Inc.). The spectra were searched against the UniProt protein sequence database (taxonomy = *Mus*

*musculus*; 84,602 protein sequences; UniProt release 2015\_01; 7 January 2015; <http://www.uniprot.org/>) (58) supplemented with the sequences of protein standards and common contaminants (416 protein sequences). The spectra were searched using both Sequest HT (a component of Proteome Discoverer) (59) and Mascot Server v2.5.1.0 (Matrix Science, Inc., Boston, MA) (60). The search settings included the following: fully tryptic,  $\leq 2$  missed cleavages, and 20 ppm precursor mass tolerance. The fragment mass tolerance settings were 0.8 Da (CID; Mascot), 1 Da (CID; Sequest HT), 0.02 Da (HCD; Mascot), and 0.06 Da (HCD; Sequest HT). The search settings included one static modification: carbamidomethylation (Cys). The search settings included seven dynamic modifications: deamidation (Asn, Gln), oxidation (Met), iTRAQ8plex (Lys, Tyr, peptide N termini), acetylation (Lys, peptide N termini), amidation (protein C termini), Met loss (protein N-terminal Met), and Met loss plus acetylation (protein N-terminal Met).

Percolator (a component of Proteome Discoverer) was used to analyze all of the resulting peptide-spectrum matches (PSMs), and the PSMs were required to have Percolator  $q$  values of  $<0.01$  (61, 62). This resulted in 577,567 PSMs with a “refined” false-discovery rate (63) of 0.0046. This corresponded to 29,954 peptide groups (ignoring peptide modifications, this corresponded to 22,573 peptides). The data were further filtered by requiring each protein group identification to correspond to either  $\geq 2$  peptide groups and  $\geq 4$  PSMs or  $\geq 10$  PSMs. This resulted in 3,088 protein group identifications (not including protein standards and common contaminants; false-discovery rate = 0.018), which corresponded to 2,968 genes (gene symbols were retrieved from bioDBnet as described above). Of the 3,088 protein group identifications, 2,896 (93.8%) corresponded to  $\geq 2$  peptide groups, and 2,823 (91.4%) corresponded to  $\geq 2$  peptides (ignoring peptide modifications).

Proteome Discoverer was used to perform iTRAQ 8plex quantification relative to the quantification reference “pool” sample described above. Each protein group abundance value was equal to the median of the corresponding PSM sample/pool ratios. Proteome Discoverer was used to perform an abundance bias correction (a central-tendency normalization) using the median protein group abundance values. The resulting 39 sample normalization values (1 per ileum) ranged from 0.52 to 1.87. The geometric mean of the 39 “log-absolute values” was 1.216 (if a normalization value was  $>1$  or  $<1$ , then the corresponding log-absolute value was defined as the normalization value or the reciprocal normalization value, respectively), which indicated that the mean bias was 21.6%. Six subsets of the set of 39 normalization values were each defined by association with one of the six experimental parameters (conventional, GF, female, male, C57BL/10A, and BALB/c), and the geometric mean of each subset of normalization values was nearly unity (0.98, 0.96, 1.02, 0.92, 0.93, and 1.01, respectively), which indicated that the biases were largely independent of all six experimental parameters. The raw mass spectra and the Proteome Discoverer data were deposited in the ProteomeXchange Consortium database (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (64) with data set identifier [PXD003177](#).

The protein group abundance values were  $\log_2$  transformed, and InfernoRDN (<https://omics.pnl.gov/software/infernordn>) (55) was used to perform 3-way ANOVAs. As with the transcriptomics InfernoRDN 3-way ANOVAs (described above), each ANOVA produced a  $q$  value for each effect (germ status, sex, and strain) and interaction, and the data were filtered by requiring the minimum  $q$  value ( $q_{\min}$ , calculated for each experimental parameter) to be  $<0.1$ . This resulted in 242 (7.84%), 13 (0.42%), and 539 (17.5%) protein groups passing the germ status, sex, and strain filters, respectively, and this corresponded to 219 (7.38%), 12 (0.40%), and 495 (16.7%) genes, respectively. HCAs were performed using Genesis as described above. For each protein group, the geometric mean of the protein group abundance values (normalized; not  $\log_2$  transformed) was calculated across the GF samples, the same was calculated for the conventional samples, and the ratios of the two geometric means were used as the GF/C values for the pathway analyses.

**Meta-analysis.** The meta-analysis used eight previously generated, publicly available gene microarray data sets described in 10 publications (12, 65–73). Throughout this article, these data are explicitly referred to as the “meta-analysis data set” (the “transcriptome data subset” exclusively refers to the mouse ileum experiments described above; the meta-analysis did not use any of the transcriptome data subset). All of the meta-analysis data were from experiments that used GF and conventional (including conventionalized) mice, and all of the analyses were of small-intestine or large-intestine whole-tissue samples (243 samples in total).

The data were downloaded from the Gene Expression Omnibus data repository at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/geo/>) (identifiers [GSE1392](#), [GSE5198](#), [GSE8006](#), [GSE18056](#), [GSE23573](#), [GSE32084](#), [GSE32513](#), and [GSE46952](#)). The meta-analysis was performed using OMics Compendia Commons (OMiCC; <https://omicc.niaid.nih.gov/>) (74), a community-based online toolset for analyzing gene expression data. The results of the meta-analysis are available online (<https://omicc.niaid.nih.gov/myProject/showResult/1118?ac=OSW9E3XLH>). In OMiCC, the samples were arranged into 43 intrastudy “sample groups” (groups of biological replicates; minimum of three replicates per group), and intrastudy pairs of the sample groups (each pair contained one GF group and one conventional group) were arranged into 31 “comparison group pairs.” The meta-analysis tested for gene expression that was significantly affected by germ status (using only intrastudy comparisons), and it produced two  $q$  values for each gene [OMiCC meta-analyses always produce a  $q$ (up) value and a  $q$ (down) value for each gene]. For each gene, the minimum of the two  $q$  values was calculated ( $q_{\min}$ ), and the data were filtered by requiring  $q_{\min}$  values of  $<0.1$  (2,132 genes passed this filter). HCAs were performed using Genesis as described above. For each gene, the geometric mean of the GF/C (germfree/conventional) ratios was calculated across the “comparison group pairs,” and these values were used as the GF/C values for the pathway analyses.

**Pathway analyses.** Ingenuity pathway analysis (Qiagen) was used to perform “Core Analyses” to discover biological pathways that were significantly affected by germ status. Each Ingenuity analysis was performed using sets of genes that were significantly affected by germ status and categorized as follows: “meta-analysis,” “transcriptome,” “proteome,” “concordant[T,P,M],” “concordant[T,P],” “concordant[T,M],” “discordant[T≠P],” “discordant[T\|P],” “Discordant[P\T],” and “invariant[P].” For each of these 10 gene sets, two additional gene subsets were analyzed; one subset consisted of the upregulated genes, and the other consisted of the downregulated genes (e.g., transcriptome[up], discordant[P\T, down], proteome = proteome[up + down]). All 30 IPAs used the same IPA version (June 2016 release; Application Build 389077m; Content version 27821452).

The meta-analysis, transcriptome, and proteome gene sets (2,132, 3,063, and 219 genes, respectively) each consisted of the genes that were significantly affected by germ status ( $q_{min} < 0.1$ , as described above). The concordant[T,P,M] gene set (33 genes) consisted of the intersection of the aforementioned transcriptome, proteome, and meta-analysis gene sets, and the 3 GF/C ratios (transcriptome, proteome, and meta-analysis) for each gene were required to be concordant (i.e., all 3 GF/C ratios =  $<1$ , or all three =  $>1$ ). The concordant[T,P] gene set (88 genes) was the same except that only the transcriptome and proteome data subsets were considered. Likewise, the concordant[T,M] gene set (441 genes) was the same except that only the transcriptome and meta-analysis gene sets were considered.

Each gene in the discordant[T\|P] gene set (583 genes) was significantly affected by germ status in the transcriptome data subset ( $q_{min} < 0.1$ ), was confidently identified and quantified in the proteome data subset, and was not significantly affected by germ status in the proteome ( $q_{min} \geq 0.1$ ) or else the transcriptome and proteome GF/C ratios were in opposite directions (i.e., one represented upregulation and the other downregulation). The discordant[P\T] gene set (118 genes) was the same but with the transcriptome and proteome data subsets inverted. The discordant[T≠P] gene set (696 genes) consisted of the union of the discordant[T\|P] and discordant[P\T] sets. Lastly, the invariant[P] gene set (60 genes) consisted of the genes significantly affected at the proteome level by germ status ( $q_{min}[germ\ status] < 0.1$ ) and not significantly affected at the proteome level by sex and strain ( $q_{min}[sex] \geq 0.1$  and  $q_{min}[strain] \geq 0.1$ ).

Each Ingenuity analysis used one of the 30 sets of genes described above (10 “up + down” gene sets, 10 “up” gene subsets, and 10 “down” gene subsets). Each Ingenuity analysis also required a reference set of genes, and each was the corresponding unfiltered (by  $q_{min}$ ) experimental gene set (the whole-mouse genome was used for the meta-analysis analyses). For each Ingenuity analysis, a gene annotation enrichment  $q$  value ( $q_e$ ) was calculated for each gene annotation (e.g., representing a biological pathway or function) by using a right-tailed Fisher exact test and subsequently applying Benjamini-Hochberg multiple-hypothesis analysis. Note that this algorithm (the Fisher exact test plus Benjamini-Hochberg test) was itself indifferent to concordancy in the direction of the GF/C values. The definition of  $q_e$  is that it represents the estimated fraction of false-positive results within the discovered set of enriched annotations (the set defined by assuming that the  $q_e$  value represents the significance level).

The “up + down” analyses were used to discover gene annotations that were statistically significantly enriched within the set of gene products that were significantly affected by germ status. Likewise, the “up” analyses were used to discover annotations that were significantly enriched within the set of upregulated gene products and the “down” analyses were used to discover annotations that were significantly enriched within the set of downregulated gene products. For some of the annotations, the “up” or “down” analysis resulted in a  $q_e$  value that was lower (i.e., represented greater statistically significance) than the corresponding  $q_e$  value from the “up + down” analysis. Mathematically, this is because  $q_e$  is dependent on both the number of affected genes that are assigned the annotation and the number of affected genes that are not assigned the annotation. Consequently, compared to an “up + down” analysis, an “up” or “down” analysis can reduce either of these numbers, and this can result in an increase or decrease in the  $q_e$  value.

The sets of biological pathways were filtered by requiring  $q_e$  values of  $<0.1$  (428 pathways passed this filter in at least 1 of the 30 analyses). Pathways were classified as “upregulated” if the annotation enrichment was maximally significant by analyzing the “up” gene subset (that is,  $q_e[up] < q_e[up + down]$  and  $q_e[up] < q_e[down]$  and  $q_e[up] < 0.1$ ), and likewise for the pathways classified as “downregulated.” This definition should not be confused with the biological activation or inactivation of the pathways, as a pathway may be inactivated due to the upregulation of specific genes/proteins, or, alternatively, a pathway can be activated due to the downregulation of specific genes/proteins. Pathways were classified as “intraset discordant” (not to be confused with interset discordance such as transcriptome-proteome discordance) if the annotation enrichment was maximally significant by analyzing the “up + down” gene set (that is,  $q_e[up + down] \leq q_e[up]$  and  $q_e[up + down] \leq q_e[down]$  and  $q_e[up + down] < 0.1$ ).

The pathway nodes colored red and blue were significantly ( $q_{min} < 0.1$ ) increased and decreased (respectively) in abundance due to germ status. The pathway nodes colored gray were not significantly affected ( $q_{min} \geq 0.1$ ), and the pathway nodes colored white represent those that were not experimentally observed. Note that OMICC performs an internal data filtration step; consequently, some meta-analysis nodes are colored white even though the corresponding genes were present in the original microarray data sets. Dual shading of a pathway node either depicted data from multiple gene products from a single data set or depicted data from two data sets (e.g., transcriptome and proteome), as indicated. For each of the 30 analyses, Ingenuity analysis was used to produce five networks of overlapping biological pathways. Each node depicted a biological pathway, and each edge depicted  $\geq 1$  overlapping gene products that were significantly affected by germ status. The minimum number of significantly affected overlapping gene products ranged between one and five. The nodes were arranged

using the Ingenuity “organic layout” algorithm. Inset concordancy  $P$  values ( $P_c$ ) were calculated using two-tailed binomial tests (significance level = 0.05). Prism v7.02 (GraphPad Software, Inc., La Jolla, CA) was used to perform two-tailed Mann-Whitney tests (significance level = 0.05) to produce interset discordancy  $P$  values ( $P_{MW}$ ) and also to produce scatter dot plots.

Chaperone, cochaperone, and E3 ubiquitin ligase protein-protein interactions were retrieved from the STRING database (v10.0) (<http://string-db.org/>) (75). Only “database” interactions (from manually curated pathways) and “experimental” interactions were used to analyze protein-protein interactions. The STRING database assigns a confidence score to every interaction and classifies interactions as low, medium, high, and very high confidence based on score thresholds of 0.15, 0.4, 0.7, and 0.9, respectively. The set of (co)chaperones (76) and the set of chaperone-associated E3 ubiquitin ligases that target misfolded proteins (48) were previously reported. This set of interactions was manually supplemented with additional data (77–82). Cytoscape was used for network visualization (<http://www.cytoscape.org/>) (83). The STRING database was also used to produce transcript coexpression networks.

**Data availability.** We have deposited the required mass spectrometry proteomics data in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (ProteomeXchange submission title, “Germ-free mouse ileum iTRAQ shotgun LC-MSMS”) under ProteomeXchange accession number **PXD003177**. The microarray data have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO Series accession number **GSE95650**.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00107-17>.

**TEXT S1**, DOCX file, 3.9 MB.

**TABLE S1**, XLSX file, 19.4 MB.

**TABLE S2**, XLSX file, 13.6 MB.

**TABLE S3**, XLSX file, 18.9 MB.

**TABLE S4**, XLSX file, 2.2 MB.

**TABLE S5**, XLSX file, 2.1 MB.

**TABLE S6**, XLSX file, 0.5 MB.

**TABLE S7**, XLSX file, 0.03 MB.

**TABLE S8**, XLSX file, 0.02 MB.

**TABLE S9**, XLSX file, 1.9 MB.

## ACKNOWLEDGMENTS

We thank Glenn A. Nardone for help with initial sample processing; Jialu Hu, Bhaskar Dutta, and Manikandan Narayanan for help with the statistical approaches; and John Tsang and Rachel Sparks for OMiCC training. Microarray hybridization and scanning was performed by the Genomic Technologies Section of the NIAID Research Technologies Branch.

This research was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health. This research was also supported by NIH grants AI109695 and DK103761.

The study was conceived by A.M. and A.N.-L. The methodology was devised by N.P.M., N.S., A.M., and A.N.-L. Formal analysis of the data was performed by N.P.M. N.P.M., N.S., A.G.N., and S.A. were responsible for the investigation performed in the study. A.M. and A.N.-L. were responsible for acquiring the resources used in the study. The original draft was written by N.P.M. Review and editing of the writing were performed by N.P.M., N.S., A.G.N., S.A., A.M., and A.N.-L. The study was supervised by A.M. and A.N.-L.

## REFERENCES

1. Relman DA. 2012. Microbiology: learning about who we are. *Nature* 486:194–195. <https://doi.org/10.1038/486194a>.
2. Ottman N, Smidt H, de Vos WM, Belzer C. 2012. The function of our microbiota: who is out there and what do they do? *Front Cell Infect Microbiol* 2:104. <https://doi.org/10.3389/fcimb.2012.00104>.
3. Kinross JM, Darzi AW, Nicholson JK. 2011. Gut microbiome-host interactions in health and disease. *Genome Med* 3:14. <https://doi.org/10.1186/gm228>.
4. Sekirov I, Russell SL, Antunes LC, Finlay BB. 2010. Gut microbiota in health and disease. *Physiol Rev* 90:859–904. <https://doi.org/10.1152/physrev.00045.2009>.
5. Clemente JC, Ursell LK, Parfrey LW, Knight R. 2012. The impact of the gut microbiota on human health: an integrative view. *Cell* 148:1258–1270. <https://doi.org/10.1016/j.cell.2012.01.035>.
6. Stappenbeck TS, Virgin HW. 2016. Accounting for reciprocal host-microbiome interactions in experimental science. *Nature* 534:191–199. <https://doi.org/10.1038/nature18285>.
7. Overstreet AC, Ramer-Tait AE, Jergens AE, Wannemuehler MJ. 2012. The role of the microbiota in gastrointestinal health and disease, chapter 3. In Szabo I (ed), *Inflammatory bowel disease*. InTechOpen, Rijeka, Croatia. <https://doi.org/10.5772/53920>.
8. Reikvam DH, Erofeev A, Sandvik A, Grcic V, Jahnson FL, Gaustad P, McCoy

- KD, Macpherson AJ, Meza-Zepeda LA, Johansen FE. 2011. Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS One* 6:e17996. <https://doi.org/10.1371/journal.pone.0017996>.
9. Peyrin-Biroulet L, Loftus EV, Jr., Colombel JF, Sandborn WJ. 2010. The natural history of adult Crohn's disease in population-based cohorts. *Am J Gastroenterol* 105:289–297. <https://doi.org/10.1038/ajg.2009.579>.
  10. Pizarro TT, Pastorelli L, Bamias G, Garg RR, Reuter BK, Mercado JR, Chieppa M, Arseneau KO, Ley K, Cominelli F. 2011. SAMP1/YitFc mouse strain: a spontaneous model of Crohn's disease-like ileitis. *Inflamm Bowel Dis* 17:2566–2584. <https://doi.org/10.1002/ibd.21638>.
  11. Dilauro S, Crum-Cianflone NF. 2010. Ileitis: when it is not Crohn's disease. *Curr Gastroenterol Rep* 12:249–258. <https://doi.org/10.1007/s11894-010-0112-5>.
  12. Gaboriau-Routhiau V, Rakotobe S, Lécuyer E, Mulder I, Lan A, Bridonneau C, Rochet V, Pisi A, De Paepen M, Brandt G, Eberl G, Snel J, Kelly D, Cerf-Bensussan N. 2009. The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* 31:677–689. <https://doi.org/10.1016/j.jimmuni.2009.08.020>.
  13. Mescher AL. 2016. Digestive tract, chapter 15, p 295–328. In Weitz M, Kearns B (ed), Junqueira's basic histology, 14th ed. McGraw-Hill Education, New York, NY.
  14. Morgan A, Dzutsev A, Dong X, Greer RL, Sexton DJ, Ravel J, Schuster M, Hsiao W, Matzinger P, Shulzhenko N. 2015. Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut* 64:1732–1743. <https://doi.org/10.1136/gutjnl-2014-308820>.
  15. 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>.
  16. 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>.
  17. Lichtman JS, Marcabal A, Sonnenburg JL, Elias JE. 2013. Host-centric proteomics of stool: a novel strategy focused on intestinal responses to the gut microbiota. *Mol Cell Proteomics* 12:3310–3318. <https://doi.org/10.1074/mcp.M113.029967>.
  18. Muth T, Kolmeder CA, Salojärvi J, Keskkitalo S, Varjosalo M, Verdam FJ, Rensen SS, Reichl U, de Vos WM, Rapp E, Martens L. 2015. Navigating through metaproteomics data: a logbook of database searching. *Proteomics* 15:3439–3453. <https://doi.org/10.1002/pmic.201400560>.
  19. Lichtman JS, Alsentzer E, Jaffe M, Sprockett D, Masutani E, Ikwa E, Fragiadakis GK, Clifford D, Huang BE, Sonnenburg JL, Huang KC, Elias JE. 2016. The effect of microbial colonization on the host proteome varies by gastrointestinal location. *ISME J* 10:1170–1181. <https://doi.org/10.1038/ismej.2015.187>.
  20. Danielsen M, Hornshøj H, Siggers RH, Jensen BB, van Kessel AG, Bendixen E. 2007. Effects of bacterial colonization on the porcine intestinal proteome. *J Proteome Res* 6:2596–2604. <https://doi.org/10.1021/pr070038b>.
  21. Jiang P, Sangild PT, Siggers RH, Sit WH, Lee CL, Wan JM. 2011. Bacterial colonization affects the intestinal proteome of preterm pigs susceptible to necrotizing enterocolitis. *Neonatology* 99:280–288. <https://doi.org/10.1159/000317807>.
  22. Volf J, Polansky O, Varmuzova K, Gerzova L, Sekelova Z, Faldynova M, Babak V, Medvecky M, Smith AL, Kaspers B, Velge P, Rychlik I. 2016. Transient and prolonged response of chicken cecum mucosa to colonization with different gut microbiota. *PLoS One* 11:e0163932. <https://doi.org/10.1371/journal.pone.0163932>.
  23. Simon GM, Cheng J, Gordon JI. 2012. Quantitative assessment of the impact of the gut microbiota on lysine epsilon-acetylation of host proteins using gnotobiotic mice. *Proc Natl Acad Sci U S A* 109: 11133–11138. <https://doi.org/10.1073/pnas.1208669109>.
  24. Kuno T, Hirayama-Kurogi M, Ito S, Ohtsuki S. 2016. Effect of intestinal flora on protein expression of drug-metabolizing enzymes and transporters in the liver and kidney of germ-free and antibiotics-treated mice. *Mol Pharm* 13:2691–2701. <https://doi.org/10.1021/acs.molpharmaceut.6b00259>.
  25. Kugadas A, Christiansen SH, Sankaranarayanan S, Surana NK, Gauguet S, Kunz R, Fichorova R, Vorup-Jensen T, Gadjeva M. 2016. Impact of microbiota on resistance to ocular *Pseudomonas aeruginosa*-induced keratitis. *PLoS Pathog* 12:e1005855. <https://doi.org/10.1371/journal.ppat.1005855>.
  26. Krautkramer KA, Kreznar JH, Romano KA, Vivas El, Barrett-Wilt GA, Rabagli ME, Keller MP, Attie AD, Rey FE, Denu JM. 2016. Diet-microbiota interactions mediate global epigenetic programming in multiple host tissues. *Mol Cell* 64:982–992. <https://doi.org/10.1016/j.molcel.2016.10.025>.
  27. Gomez A, Luckey D, Taneja V. 2015. The gut microbiome in autoimmunity: sex matters. *Clin Immunol* 159:154–162. <https://doi.org/10.1016/j.clim.2015.04.016>.
  28. Jašarević E, Morrison KE, Bale TL. 2016. Sex differences in the gut microbiome-brain axis across the lifespan. *Philos Trans R Soc Lond B Biol Sci* 371:20150122. <https://doi.org/10.1098/rstb.2015.0122>.
  29. Org E, Mehrabian M, Parks BW, Shipkova P, Liu X, Drake TA, Lusis AJ. 2016. Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes* 7:313–322. <https://doi.org/10.1080/19490976.2016.1203502>.
  30. Org E, Parks BW, Joo JW, Emert B, Schwartzman W, Kang EY, Mehrabian M, Pan C, Knight R, Gunsalus R, Drake TA, Eskin E, Lusis AJ. 2015. Genetic and environmental control of host-gut microbiota interactions. *Genome Res* 25:1558–1569. <https://doi.org/10.1101/gr.194118.115>.
  31. Gulati AS, Shanahan MT, Arthur JC, Grossniklaus E, von Furstenberg RJ, Kreuk L, Henning SJ, Jobin C, Sartor RB. 2012. Mouse background strain profoundly influences Paneth cell function and intestinal microbial composition. *PLoS One* 7:e32403. <https://doi.org/10.1371/journal.pone.0032403>.
  32. Brodziak F, Meharg C, Blaut M, Loh G. 2013. Differences in mucosal gene expression in the colon of two inbred mouse strains after colonization with commensal gut bacteria. *PLoS One* 8:e72317. <https://doi.org/10.1371/journal.pone.0072317>.
  33. Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, Park CC, Wen PZ, Brewer H, Weitz K, Camp DG, II, Pan C, Yordanova R, Neuhaus I, Tilford C, Siemers N, Gargalovic P, Eskin E, Kirchgessner T, Smith DJ, Smith RD, Lusis AJ. 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* 7:e1001393. <https://doi.org/10.1371/journal.pgen.1001393>.
  34. Low TY, van Heesch S, van den Toorn H, Giansanti P, Cristobal A, Toonen P, Schafer S, Hübner N, van Breukelen B, Mohammed S, Cuppen E, Heck AJ, Guryev V. 2013. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Rep* 5:1469–1478. <https://doi.org/10.1016/j.celrep.2013.10.041>.
  35. Wang X, Liu Q, Zhang B. 2014. Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* 14:2676–2687. <https://doi.org/10.1002/pmic.201400184>.
  36. Fonseca BD, Smith EM, Yelle N, Alain T, Bushell M, Pause A. 2014. The ever-evolving role of mTOR in translation. *Semin Cell Dev Biol* 36: 102–112. <https://doi.org/10.1016/j.semcd.2014.09.014>.
  37. Fonseca BD, Graber TE, Hoang H, González A, Soukas AA, Hernández G, Alain T, Swift SL, Weisman R, Meyer C, Robaglia C, Avruch J, Hall MN. 2016. Evolution of TOR and translation control, chapter 15, p 327–411. In Hernández G, Jagus R (ed), Evolution of the protein synthesis machinery and its regulation. Springer International, Cham, Switzerland. <https://doi.org/10.1007/978-3-319-39468-8>.
  38. Ivanov P, Kedersha N, Anderson P. 2011. Stress puts TIA on TOP. *Genes Dev* 25:2119–2124. <https://doi.org/10.1101/gad.17838411>.
  39. Damgaard CK, Lykke-Andersen J. 2011. Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR. *Genes Dev* 25:2057–2068. <https://doi.org/10.1101/gad.17355911>.
  40. Fonseca BD, Zakaria C, Jia JJ, Graber TE, Svitkin Y, Tahmasebi S, Healy D, Hoang HD, Jensen JM, Diao IT, Lussier A, Dajadian C, Padmanabhan N, Wang W, Matta-Camacho E, Hearnden J, Smith EM, Tsukumo Y, Yanagiya A, Morita M, Petroulakis E, González JL, Hernández G, Alain T, Damgaard CK. 2015. La-related Protein 1 (LARP1) represses terminal oligopyrimidine (TOP) mRNA translation downstream of mTOR Complex 1 (mTORC1). *J Biol Chem* 290:15996–16020. <https://doi.org/10.1074/jbc.M114.621730>.
  41. Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ, Wang S, Ren P, Martin M, Jessen K, Feldman ME, Weissman JS, Shokat KM, Rommel C, Ruggero D. 2012. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 485:55–61. <https://doi.org/10.1038/nature10912>.
  42. van Waveren C, Moraes CT. 2008. Transcriptional co-expression and co-regulation of genes coding for components of the oxidative phosphorylation system. *BMC Genomics* 9:18. <https://doi.org/10.1186/1471-2164-9-18>.
  43. Arnal ME, Lallès JP. 2016. Gut epithelial inducible heat-shock proteins

- and their modulation by diet and the microbiota. *Nutr Rev* 74:181–197. <https://doi.org/10.1093/nutrit/nuv104>.
44. Liu H, Dicksved J, Lundh T, Lindberg JE. 2014. Heat shock proteins: intestinal gatekeepers that are influenced by dietary components and the gut microbiota. *Pathogens* 3:187–210. <https://doi.org/10.3390/pathogens3010187>.
  45. Grootjans J, Kaser A, Kaufman RJ, Blumberg RS. 2016. The unfolded protein response in immunity and inflammation. *Nat Rev Immunol* 16:469–484. <https://doi.org/10.1038/nri.2016.62>.
  46. Calderwood SK, Gong J, Murshid A. 2016. Extracellular HSPs: the complicated roles of extracellular HSPs in immunity. *Front Immunol* 7:159. <https://doi.org/10.3389/fimmu.2016.00159>.
  47. Binder RJ, Zhou YJ, Messmer MN, Pawaria S. 2012. CD91-dependent modulation of immune responses by heat shock proteins: a role in autoimmunity. *Autoimmune Dis* 2012:863041. <https://doi.org/10.1155/2012/863041>.
  48. Kriegerburg F, Ellgaard L, Hartmann-Petersen R. 2012. Molecular chaperones in targeting misfolded proteins for ubiquitin-dependent degradation. *FEBS J* 279:532–542. <https://doi.org/10.1111/j.1742-4658.2011.08456.x>.
  49. Soss SE, Rose KL, Hill S, Jouan S, Chazin WJ. 2015. Biochemical and proteomic analysis of ubiquitination of Hsc70 and Hsp70 by the E3 ligase CHIP. *PLoS One* 10:e0128240. <https://doi.org/10.1371/journal.pone.0128240>.
  50. 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>.
  51. Mudunuri U, Che A, Yi M, Stephens RM. 2009. bioDBnet: the biological database network. *Bioinformatics* 25:555–556. <https://doi.org/10.1093/bioinformatics/btn654>.
  52. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. 2007. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform* 3:11–17.
  53. Dobbin K, Simon R. 2002. Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* 18:1438–1445. <https://doi.org/10.1093/bioinformatics/18.11.1438>.
  54. Sturm A, Quackenbush J, Trajanoski Z. 2002. Genesis: cluster analysis of microarray data. *Bioinformatics* 18:207–208. <https://doi.org/10.1093/bioinformatics/18.1.207>.
  55. Polipitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG, II, Anderson GA, Smith RD. 2008. Dante: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24:1556–1558. <https://doi.org/10.1093/bioinformatics/btn217>.
  56. Oberg AL, Vitek O. 2009. Statistical design of quantitative mass spectrometry-based proteomic experiments. *J Proteome Res* 8:2144–2156. <https://doi.org/10.1021/pr8010099>.
  57. Manes NP, Dong L, Zhou W, Du X, Reghu N, Kool AC, Choi D, Bailey CL, Petricoin EF III, Liotta LA, Popov SG. 2011. Discovery of mouse spleen signaling responses to anthrax using label-free quantitative phosphoproteomics via mass spectrometry. *Mol Cell Proteomics* 10:M110.000927. <https://doi.org/10.1074/mcp.M110.000927>.
  58. UniProt Consortium. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42:D191–D198. <https://doi.org/10.1093/nar/gkt1140>.
  59. Eng JK, Fischer B, Grossmann J, MacCoss MJ. 2008. A fast SEQUEST cross correlation algorithm. *J Proteome Res* 7:4598–4602. <https://doi.org/10.1021/pr0800420s>.
  60. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567. [https://doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2).
  61. Brosch M, Yu L, Hubbard T, Choudhary J. 2009. Accurate and sensitive peptide identification with Mascot Percolator. *J Proteome Res* 8:3176–3181. <https://doi.org/10.1021/pr800982s>.
  62. Spivak M, Weston J, Bottou L, Käll L, Noble WS. 2009. Improvements to the percolator algorithm for peptide identification from shotgun proteomics data sets. *J Proteome Res* 8:3737–3745. <https://doi.org/10.1021/pr801109k>.
  63. Navarro P, Vázquez J. 2009. A refined method to calculate false discovery rates for peptide identification using decoy databases. *J Proteome Res* 8:1792–1796. <https://doi.org/10.1021/pr800362h>.
  64. Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O'Kelly G, Schoenegger A, Ovelleiro D, Pérez-Riverol Y, Reisinger F, Ríos D, Wang R, Hermjakob H. 2013. The proteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* 41:D1063–D1069. <https://doi.org/10.1093/nar/gks1262>.
  65. Mutch DM, Simmering R, Donnicola D, Fotopoulos G, Holzwarth JA, Williamson G, Corthéz-Theulaz I. 2004. Impact of commensal microbiota on murine gastrointestinal tract gene ontologies. *Physiol Genomics* 19:22–31. <https://doi.org/10.1152/physiolgenomics.00105.2004>.
  66. Rawls JF, Mahowald MA, Ley RE, Gordon JL. 2006. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* 127:423–433. <https://doi.org/10.1016/j.cell.2006.08.043>.
  67. Munakata K, Yamamoto M, Anjiki N, Nishiyama M, Imamura S, Iizuka S, Takashima K, Ishige A, Hioki K, Ohnishi Y, Watanabe K. 2008. Importance of the interferon-alpha system in murine large intestine indicated by microarray analysis of commensal bacteria-induced immunological changes. *BMC Genomics* 9:192. <https://doi.org/10.1186/1471-2164-9-192>.
  68. Shulzhenko N, Morgan A, Hsiao W, Battle M, Yao M, Gavrilova O, Orandle M, Mayer L, Macpherson AJ, McCoy KD, Fraser-Liggett C, Matzinger P. 2011. Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nat Med* 17:1585–1593. <https://doi.org/10.1038/nm.2505>.
  69. Munakata K, Takashima K, Nishiyama M, Asano N, Mase A, Hioki K, Ohnishi Y, Yamamoto M, Watanabe K. 2012. Microarray analysis on germfree mice elucidates the primary target of a traditional Japanese medicine juzentaihoto: acceleration of IFN-alpha response via affecting the ISGF3-IRF7 signaling cascade. *BMC Genomics* 13:30. <https://doi.org/10.1186/1471-2164-13-30>.
  70. El Aidy S, van Baarlen P, Derrien M, Lindenbergh-Kortleve DJ, Hooiveld G, Levenez F, Doré J, Dekker J, Samsom JN, Nieuwenhuis EE, Kleerebezem M. 2012. Temporal and spatial interplay of microbiota and intestinal mucosa drive establishment of immune homeostasis in conventionalized mice. *Mucosal Immunol* 5:567–579. <https://doi.org/10.1038/mi.2012.32>.
  71. El Aidy S, Merrifield CA, Derrien M, van Baarlen P, Hooiveld G, Levenez F, Doré J, Dekker J, Holmes E, Claus SP, Reijngoud DJ, Kleerebezem M. 2013. The gut microbiota elicits a profound metabolic reorientation in the mouse jejunum mucosa during conventionalisation. *Gut* 62:1306–1314. <https://doi.org/10.1136/gutjnl-2011-301955>.
  72. El Aidy S, Derrien M, Merrifield CA, Levenez F, Doré J, Boekschen MV, Dekker J, Holmes E, Zoetendal EG, van Baarlen P, Claus SP, Kleerebezem M. 2013. Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon. *ISME J* 7:743–755. <https://doi.org/10.1038/ismej.2012.142>.
  73. Joyce SA, MacSharry J, Casey PG, Kinsella M, Murphy EF, Shanahan F, Hill C, Gahan CG. 2014. Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. *Proc Natl Acad Sci U S A* 111:7421–7426. <https://doi.org/10.1073/pnas.1323599111>.
  74. Shah N, Guo Y, Wendelsdorf KV, Lu Y, Sparks R, Tsang JS. 2016. A crowdsourcing approach for reusing and meta-analyzing gene expression data. *Nat Biotechnol* 34:803–806. <https://doi.org/10.1038/nbt.3603>.
  75. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447–D452. <https://doi.org/10.1093/nar/gku1003>.
  76. Brehme M, Voisine C, Rolland T, Wachi S, Soper JH, Zhu Y, Orton K, Villella A, Garza D, Vidal M, Ge H, Morimoto RI. 2014. A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep* 9:1135–1150. <https://doi.org/10.1016/j.celrep.2014.09.042>.
  77. Kozlov G, Maattanen P, Schrag JD, Pollock S, Cygler M, Nagar B, Thomas DY, Gehring K. 2006. Crystal structure of the bb' domains of the protein disulfide isomerase ERp57. *Structure* 14:1331–1339. <https://doi.org/10.1016/j.str.2006.06.019>.
  78. Timchenko LT, Salisbury E, Wang GL, Nguyen H, Albrecht JH, Hershey JW, Timchenko NA. 2006. Age-specific CUGBP1-eIF2 complex increases translation of CCAAT/enhancer-binding protein beta in old liver. *J Biol Chem* 281:32806–32819. <https://doi.org/10.1074/jbc.M605701200>.
  79. Meunier L, Usherwood YK, Chung KT, Hendershot LM. 2002. A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins. *Mol Biol Cell* 13:4456–4469. <https://doi.org/10.1093/mbc.E02-05-0311>.
  80. Neutzner A, Neutzner M, Benischke AS, Ryu SW, Frank S, Youle RJ, Karbowski M. 2011. A systematic search for endoplasmic reticulum (ER) membrane-associated RING finger proteins identifies Nixin/ZNRF4 as a

- regulator of calnexin stability and ER homeostasis. *J Biol Chem* 286: 8633–8643. <https://doi.org/10.1074/jbc.M110.197459>.
81. Cheng ST, Nguyen TQ, Yang YS, Capra JD, Sontheimer RD. 1996. Calreticulin binds hYRNA and the 52-kDa polypeptide component of the Ro/SS-A ribonucleoprotein autoantigen. *J Immunol* 156:4484–4491.
82. Purcell AW, Todd A, Kinoshita G, Lynch TA, Keech CL, Gething MJ, Gordon TP. 2003. Association of stress proteins with autoantigens: a possible mechanism for triggering autoimmunity? *Clin Exp Immunol* 132:193–200. <https://doi.org/10.1046/j.1365-2249.2003.02153.x>.
83. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27:431–432. <https://doi.org/10.1093/bioinformatics/btq675>.