bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Metagenome-wide measurement of protein synthesis in the human fecal microbiota using MetaRibo-Seq

Brayon J. Fremin[1] and Ami S. Bhatt[1,2*]

[1]Department of Genetics, Stanford University, CA, 94305, USA

[2]Department of Medicine (Hematology), Stanford University, CA, 94305, USA

*Correspondence: asbhatt@stanford.edu

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The healthy human fecal microbiota is too diverse to comprehensively study with the current throughput of proteomic methods. Shotgun sequencing technologies allow for much more comprehensive profiling. Here, we develop and apply MetaRibo-Seq, a method for simultaneous ribosome profiling of multiple taxa within a complex bacterial community. This approach captures taxonomic diversity in fecal samples. As expected, the detected ribosome-bound transcripts are relatively enriched within coding regions and significantly correlate to detectable protein abundances. In a low diversity fecal sample, we show that MetaRibo-Seq is more strongly correlated than metatranscriptomic data to protein abundance. This significant correlation of metatranscriptomics and MetaRibo-Seq with protein levels is maintained, though with decreased strength as taxonomic diversity increases. Finally, we identify genes that are consistently regulated at the translational level across bacterial taxa within fecal communities. In conclusion, MetaRibo-Seq enables comprehensive translational profiling in complex bacterial communities for the first time.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

There is great interest in determining the potential functions of the human fecal microbiota. To date, methods have excelled at describing the taxonomy of such communities; however, assigning and defining functions of the community of bacteria or individual organisms within these communities has been challenging[1]. An ideal method to study functions within a complex community would allow simultaneous enumeration of all of the proteins, lipids, and other macromolecules within the mixture. Unfortunately, this is not feasible with current technologies. While 10^2 to 10^4 proteins can be simultaneously quantified with metaproteomics[2], it is challenging to obtain accurate measurements of the full array of bacterial proteins that likely exist in human fecal samples, estimated to be 10^7 to 10^8 proteins[3]. Thus, current proteomic methods lack the dynamic range required to comprehensively study the human fecal microbiota[4].

Given the challenges in direct protein measurement and need for databases of protein sequences, some have focused on enumerating the gene content of a community to determine the potential function; indeed, progress has been made in predicting genes present within a metagenome[5]. However, the presence of a gene in a complex bacterial community does not imply that the gene is transcribed or translated. Acknowledging this limitation, recent work has demonstrated that differential transcription of bacterial genes can be used to derive biologically meaningful insights[6,7,8,9]. Yet, we still have a very limited understanding of regulation occurring post-transcriptionally in the human fecal microbiota.

In contrast to transcriptomic profiling, ribosome profiling (Ribo-Seq) is a method that quantifies protein synthesis[10]. In eukaryotes, Ribo-Seq generally correlates more strongly to protein abundance than transcriptomics[11,12,13]; this correlation has not been described in bacteria. Bacterial ribosome profiling studies have been performed in model organisms such as *Escherichia coli* and *Bacillus subtilis*, with minor modifications from the eukaryotic protocols, such as using chloramphenicol to inhibit translation and micrococcal nuclease (MNase) to enrich for ribosome footprints[13,14,15,16,17]. These methodological modifications enable a high-throughput snapshot of translation, but often compromise the ability to achieve codon-level resolution[11].

In bacteria, many genes are regulated at the translational level. For example, genes involved in translation itself are known to be regulated at a translational level via feedback mechanisms[18,19,20,21,22]. Translational regulation is critical for generating proteins at the correct stoichiometry for many protein complexes. For example, the multiprotein complex that forms

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bacterial ATP synthase has multiple genes whose translation is regulated; the stoichiometry of this complex is best predicted by Ribo-Seq[13,17]. This even extends to pathway-specific enzyme stoichiometry in which protein synthesis remains conserved as compensation for transcript abundance and architecture divergence across taxa[17]. Moreover, specific translational regulation has been extensively observed upon a variety of perturbations to bacteria[23,24,25,26,27,28,29]. For example, bacteria employ translational quality control and regulation of amino acid biosynthesis in response to amino acid stress[30]. Thus, translation is a conserved, critical, dynamic, and regulated process in bacteria. However, this level of regulation has thus far been overlooked in mixed bacterial communities. Previous studies of protein synthesis in bacteria have been restricted to pure large-scale cultures (media and RNA inputs up to liters and milligrams, respectively). To date, studying translational regulation in mixed communities or in culture-free contexts has been hindered by low extraction yield, low purity, and the lack of informatic frameworks to study organisms without reference genomes. Consequently, we have a very limited understanding of how widespread bacterial translational regulation may be outside of cultured model organisms.

In this work, we develop a method that allows for simultaneous ribosome profiling in a complex community of bacteria without the need for a large-scale, purified cultures. With three lines of evidence, we confirm that MetaRibo-Seq effectively enables translation to be studied in the fecal microbiota. First, the signal consists of footprints that capture the taxonomic diversity of metagenomics, while being locally enriched within coding regions. We identify most enrichment at gene start and stop codons, characteristic of chloramphenicol-treated ribosome profiling[31]. Second, ribosome footprint densities significantly correlate to detectable protein abundances and are significantly enriched in signal for these abundant proteins in these complex bacterial communities. In low diversity human fecal samples, we show that MetaRibo-Seq better correlates with protein abundance and *E. coli* ATP synthase stoichiometry than transcriptomics. Third, biological processes known to be translationally regulated, such as translation itself, are consistently detected as such across multiple samples and taxa. We catalog tens of thousands of genes with evidence of translational regulation in fecal samples across diverse taxa, providing a widespread view of consistent, bacterial translational regulation in these systems. Overall, we show that MetaRibo-Seq facilitates metagenome-wide measurement of bacterial protein synthesis across taxa directly in fecal samples.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

The MetaRibo-Seq workflow.

MetaRibo-Seq enables sequencing of ribosome-protected footprints directly from human fecal samples (see Methods, Figure 1A). First, we show ribosome profiling can be performed on frozen fecal samples stored in RNAlater[32, 6] (Ambion), an RNA preserving solution. Unlike some existing protocols[33,16], our ribosome profiling protocol first introduces chloramphenicol during lysis. Bead beating lysis is performed to also lyse diverse Gram-positive bacteria[34]. An ethanol precipitation step post-lysis is introduced to both filter out fecal debris and concentrate RNA and any complexes bound. This has been demonstrated to effectively precipitate ribosomes[35]. MNase treatment is performed on an extremely crude purification of nucleic acids and complexes. Ribosome profiling performed here uses nearly an order of magnitude less RNA and MNase than isolate protocols typically use (see Methods)[16, 33]. After high-quality footprints are reliably generated using these methods, ribosome profiling converges to isolate protocols to purify monosomes and prepare libraries[33]. MetaRibo-Seq overcome challenges of sample storage, input requirement, bacterial purity, and uniform lysis to generate high quality RNA footprints from fecal samples.

Computationally, dealing with short reads and poor or incomplete reference genomes is challenging. To overcome these challenges, we use a *de novo* approach to build references, annotate genes, and map reads to those references (see Methods, Figure 1B). Mapping metrics to *de novo* references are provided (Table S1). We require perfect, unique matches of these ribosome footprints to references to ensure proper mapping. Multi-mapping varies sample to sample, ranging from 6.95 to 26.69 percent. We find that 4.1 - 10.4 percent of mapped reads from RNA technologies performed correspond to predicted coding regions. Given variable amounts of diversity and heterogeneity in any given sample, mapping statistics will vary sample to sample.

MetaRibo-Seq signal retains taxonomic diversity in human fecal samples.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MetaRibo-Seq captures taxonomic bacterial diversity in human fecal samples via ribosome-protected footprints. We perform MetaRibo-Seq on four diverse fecal samples. Sample A is from a healthy individual. Sample B is from a patient with a hematological disorder who is undergoing treatment. Sample C is from a patient with a solid malignancy who is undergoing treatment. Sample D is from a patient with Alzheimer's disease. We also perform MetaRibo-Seq on a low diversity fecal sample from a patient with a hematological disorder who is undergoing antibiotic treatment with metronidazole – Sample E. Metagenomic reads are subjected to *de novo* assembly and gene prediction and annotation for each sample (see Methods). These assemblies and gene predictions are provided (NCBI BioProject ####). Taxonomic differences at the genus level exist between technologies across samples, though most abundant taxa are largely consistent across technologies (Figure 2A). Shannon diversity is also concordant between technologies, including MetaRibo-Seq (Figure 2B). Thus we conclude that MetaRibo-Seq signal faithfully recapitulates the diversity of organisms present in the mixed bacterial communities.

MetaRibo-Seq signal is characteristic of bacterial ribosome profiling.

We find that MetaRibo-Seq signal is locally enriched within coding regions. We show average signal across all coding predictions and flanking regions for Samples A, B, C, and D (Figure 3A-D). We visualize strong signal corresponding to predicted ORFs with pronounced signal drop off outside of the start and stop codons for samples A through D (Figure 3A-D). Start and stop codons represent the strongest signal. Surprisingly, MetaRibo-Seq also shows some weak signs of overall codon resolution (Figure S1). In a more targeted analysis, MetaRibo-Seq can achieve stronger codon resolution of ribosomes in common genera. For Sample A and Sample B, assembled contigs of several shared genera are classified and binned appropriately (see Methods). Triplet periodicity is observed across footprint lengths in Bacteroides (Figure S2A), Faecalibacterium (Figure S2B), and Alistipes (Figure S2C). Based purely on raw signal, these findings collectively suggest that MetaRibo-Seq is capturing ribosome-bound footprints as expected.

MetaRibo-Seq outperforms metatranscriptomics as a proxy for protein abundance in a low diversity fecal sample.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In Sample E, we identify strong correlations of metatranscriptomics and MetaRibo-Seq to metaproteomics. First, Sample E is dominated by *E coli*. Interestingly, this *E. coli* is later isolated from the blood of the patient with no single nucleotide variants compared to fecal *E. coli*[36]. This blood isolate has been sequenced and represents our isolate reference for downstream analyses. In a separate study demonstrating strain identity, this patient is denoted as Patient 3, and Sample E in this study is specifically 27 days prior to bacteremia[36]. For the 1503 genes that were proteomically-detected in Sample E, we show genus-level representation for metatranscriptomics, MetaRibo-Seq, and metaproteomics (Figure 4A). Of note, 812 of these proteins belong to *E. coli*, while the remaining 691 belong to other taxa. There is greater representation of proteins predicated to arise from Klebsiella and Enterococcus than transcripts or MetaRibo-Seq signal (Figure 4A). We show a Pearson correlation of 0.46 between MetaRibo-Seq and metaproteomics across the 1503 detected proteins from the metagenomic Sample E (Figure 4B). The Pearson correlation is 0.64 when only considering the 928 detected proteins from the *E. coli* isolate reference (Figure 4C). We display correlations between metatranscriptomics, MetaRibo-Seq, and metaproteomics (Figure 5D). Same technology correlations between the 812 identical protein predictions from the metagenomic and isolate *E. coli* analyses all retain Pearson correlations of 0.99. Correlations are weaker in the metagenomic context specifically due to relatively poorer predictions upon addition of the 691 proteins that do not belong to *E. coli*. No previous study to our knowledge provides a correlation between ribosome profiling and proteomics in *E. coli* or any other bacteria; however, this correlation of 0.64 is stronger than any cited correlation between transcriptomics and proteomics in isolated, cultured, model *E. coli*[37]. ATP synthase is a well-characterized complex in terms of stoichiometry in *E. coli* (Figure 4E). We show that MetaRibo-Seq signal better correlates with known ATP synthase stoichiometry than transcriptomics, as expected (Figure 4F-G). Among sequencing technologies, MetaRibo-Seq serves as a better proxy for protein levels and ATP synthase stoichiometry in Sample E *E. coli*.

MetaRibo-Seq signal significantly correlates to protein abundance and is enriched in these proteins in mixed bacterial communities.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We find that MetaRibo-Seq significantly correlates to protein abundances as measured by shotgun metaproteomics. In Sample A, we detect 497 proteins. We measure a significant Pearson correlation of 0.32 between MetaRibo-Seq and metaproteomics (Figure 5A). MetaRibo-Seq is better correlated with protein abundance than metatranscriptomics in Sample A (Figure 5B). As expected, both metatranscriptomics and MetaRibo-Seq are significantly enriched in signal for these 497 proteomically-detected genes (Figure 5C). In Sample B, we detect 480 proteins and measure a Pearson correlation of 0.34 between MetaRibo-Seq and metaproteomics (Figure 5D). There is no significant difference in protein abundance prediction between metatranscriptomics and MetaRibo-Seq in Sample B (Figure 5E). Metaranscriptomics and MetaRibo-Seq are similarly enriched in signal for these proteins in Sample B (Figure 5F). These proteins detected by mass spectrometry represent highly abundant bacterial proteins in the fecal samples. These findings suggest that MetaRibo-Seq correlates well with highly abundant protein levels in mixed bacterial communities, and that MetaRibo-Seq signal may thus serve as a surrogate for protein abundance in the study of complex bacterial communities.

MetaRibo-Seq signal characteristics and predictive power of protein abundance suggests that it may also prove useful in predicting proteins in taxonomically diverse fecal samples. As a preliminary demonstration of this, we predict small proteins using Prodigal[38] with decreased length cutoff (see Methods). We show a histogram of the number of small predictions (20-29 amino acids) and the number of those predictions with MetaRibo-Seq RPKM greater than 0.5 across samples (Figure S4A). Due to metaproteomic limitations, we are unable to validate these proteins directly. However, we can use a comparative genomic approach to identify clusters of small proteins, all with evidence of translation, that also possess evolutionary signatures indicative of coding regions. We cluster proteins at 70 percent amino acid identity (see Methods). We discover 21 clusters (with at least 4 members) of small proteins across Samples A, B, C, and D that contain both translational evidence among all members in the cluster and significant coding signatures determined via RNAcode[39] (Figure S3B). We also show greater protein conservation among predictions with MetaRibo-Seq signal than by random chance (Figure S3D). Translational evidence of small proteins in diverse fecal samples decreases the number of predictions to a more conserved subset, suggesting it may be useful in gene prediction.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consistent translational regulation is observed across samples and taxa.

By contrasting metatranscriptomics with MetaRibo-Seq, we identify translationally regulated genes in fecal samples. This provides a widespread view of genes that are consistently translationally regulated within these systems. For Samples A, B, C, and D individually, we show the number of gene predictions and significantly translationally regulated genes (Figure 6A). We detect reasonable DESeq2 model fits in comparing between technologies, as shown by the dispersion plots for these analyses for each sample (Figure S4). These significantly translationally regulated genes are clustered at 70 percent amino acid identity for each sample (see Methods, Figure 6B). In a combined analysis (Figure 6C), we define any cluster containing five or more sequences as consistently translationally regulated. The representative sequences for all of these consistently translationally regulated clusters are assigned GO terms with Blast2GO[40] (see Methods). The top 10 most common biological process associated GO terms are displayed, with translation being the top hit (Figure 6D). These sequences and clusters are provided for reference (File S1). Across samples, we catalog 42,267 differentially translated genes and 607 consistently translationally regulated gene clusters in these fecal samples, many of which are involved in expected processes, like translation.

Other approaches, albeit more biased to known gene annotations, are to rely on Prokka[41] annotations. We count the number of times a gene symbol appears as differentially translated (Figure S5) and input those that appear at least five times into a GO Analysis (see Methods, Figure S6). Translation remains the top hit. For Samples A, B, C, and D, we provide GO analysis results of differential genes for each sample individually (Table S2). As a more pathway-oriented analysis, we also determine overrepresentation of pathways among globally differentially translated enzymes based on EC numbers assigned by Prokka[41] (see Methods, Table S3). Amino acid biosynthesis is among the most consistent top hits across samples (Table S3). Thus several approaches lead to expected conclusions of pathways and processes that are translationally regulated.

Discussion

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

One of the major limitations in advancing the functional knowledge of microbial communities is an inability to measure the macromolecular output of a given community in an unbiased manner. For example, until now, we have been unable to study fecal bacterial communities, or any *in vivo* system of bacteria, at the level of protein synthesis. Previous approaches have required *in vitro* growth of large, purified cultures; this limits both throughput and the diversity of the sample that can be studied. Here, we introduce a new method, MetaRibo-Seq, and provide evidence that this method enables the human fecal microbiota to be studied at a translational level. In five taxonomically varied samples from human subjects with variable health status, we show MetaRibo-Seq signal retains the taxonomic diversity of the samples. The signal itself is as expected for a chloramphenicol-treated ribosome profiling library – including local enrichment within coding regions and greater enrichment across the start and stop codons of genes. This suggests we have a way to measure bacterial protein synthesis *in vivo* for the first time.

To conduct a fair comparison, we perform MetaRibo-Seq on diverse samples but also a lower diversity fecal sample (Sample E) so that more representative protein quantifications of select taxa are achievable. In taxonomically diverse stool samples, MetaRibo-Seq is comparably predictive to protein abundance as metatranscriptomics. However, we show it can be significantly more predictive in a lower diversity scenario. We show that the addition of lower abundant taxa weaken overall correlations in mixed communities. Biologically, metatranscriptomics and MetaRibo-Seq are snapshots of gene transcripts or proteins synthesized, respectively, not direct measurement of the proteins that currently exists. Moreover, significant post-translational differences between taxa likely exist. Technically, it becomes challenging to obtain accurate protein abundances for lowly abundant taxa and proteins in a sample, making such correlations to protein abundance themselves less representative. We conclude that while MetaRibo-Seq can outperform metatranscriptomics within a highly abundant organism, this effect is diminished, perhaps both for biological and technical reasons, when considering all taxa together in diverse communities.

There are several limitations to MetaRibo-Seq. First, MetaRibo-Seq does not include steps to degrade RNAs with secondary structure. This is a common issue for ribosome profiling protocols but exacerbated in this *de novo*, low input context[10, 13]. Though targeted approaches for specific bacteria have been successful for tRNA depletion,[26] an untargeted approach, which

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

would be necessary here, has yet to be implemented in literature. Utilization of sucrose density gradients instead of size-exclusion chromatography columns may also prove valuable in removing various structured RNAs; however, downstream ribosome profiling input requirements likely make this challenging to address. This limitation may, however, enable other types of investigation; for example, signal corresponding to structured RNAs will likely be useful to predict novel structured RNAs in non-coding regions. Experimental modifications will also likely improve these limitations.

We anticipate that MetaRibo-Seq will enable a clearer functional view of the fecal microbiota. With increasing use of metatranscriptomics, we envision that MetaRibo-Seq will be applied to various disease states to better probe the microbiota and its functions. We especially anticipate MetaRibo-Seq will be used longitudinally to study translational regulation of genomically-stable, clinically-relevant taxa. MetaRibo-Seq signal provides unique features like enrichment within coding regions, greater enrichment at the start and end of genes, and, as we show, some signs of codon-level resolution in some taxa. Future work will likely include using these features for gene prediction, which is particularly challenging when studying metagenomic samples and small proteins[42]. To validate such predictions, significant methods development and improvements in metaproteomics will be needed. With direct proteomic evidence often unattainable, coding potential, translational evidence, and conservation among predictions present themselves as the strongest lines of evidence proteins, especially small proteins, exists in the fecal microbiota (Figure S6 and File S2). We also expect MetaRibo-Seq to be applied to other culture-free conditions, perhaps requiring other modifications. Overall, we show that translation can be comprehensively studied in mixed bacterial communities in a culture-free manner. This method also sheds light on consistently translationally regulated genes *in vivo* in a comprehensive, metagenome-wide analysis.

Materials and Methods

Subject Recruitment

MetaRibo-Seq was performed on fecal samples from individuals from a variety of health states. Informed consent was obtained for all participants. None of the participants received bacterial translation inhibitors. All subjects were recruited at Stanford University as a part of one

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of three IRB-approved protocols for tissue biobanking and clinical metadata collection (PIs: Dr. Ami Bhatt, Dr. Victor Henderson, Dr. David Miklos).

Fecal Samples Storage

Stool was immediately stored in 2 mL cryovials and frozen at -80 °C. Stool was not thawed until lysis. For RNA extraction applications, 1.3 grams of fecal samples were preserved in 700 µL of RNALater (Ambion) at -80 °C.

Cell Lysis for Metatranscriptomics and MetaRibo-Seq

Stool (150 mg) was suspended in 600 µL Qiagen RLT lysis buffer supplemented with one percent beta-mercaptoethanol and 0.3 U/µL Superase-In (Invitrogen). For MetaRibo-Seq lysis, 1.55 mM of chloramphenicol was also added to this lysis solution, and the solution was incubated at room temperature for 5 minutes. The suspension was subjected to bead beating for 3 minutes using 1.0 mm Zirconia/Silica beads. This was performed with a MiniBeadBeater-16, Model 607. The lysed solution was centrifuged at room temperature for 3 minutes at 21,000 x g to pellet cellular debris, and the supernatant was extracted to 2 mL tubes.

Metagenomics

DNA was extracted from fecal samples with DNA Stool Mini Kit (Qiagen) using manufacture protocols. Samples were exposed to bead beating for 3 minutes. 1 ng of DNA was used to create Nextera XT libraries according to manufacturer's instructions (Illumina).

MetaRibo-Seq

The lysis supernatant was subjected to ethanol precipitation with 0.1 percent volume 3M sodium acetate and 2.5 volumes of 100 percent ethanol. To precipitate, samples were incubated at -80 °C for 30 minutes, then centrifuged at 21,000 x g for 30 minutes at 4 °C. This was a rough purification specifically implemented to enable suspension of concentrated RNA from reasonable input of fecal sample. The pellet of RNA and RNA-protein complexes was resuspended in MNase buffer. The buffer contained 25 mM Tris pH 8.0, 25 mM NH_4Cl, 10 mM MgOAc, and 1.55 mM chloramphenicol. To resuspend, we quickly broke the pellet apart with a pipette tip and vortexed for 15 seconds. 1 µL of solution was diluted 20 fold and quantified with

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Qubit dsDNA HS Assay Kit (Invitrogen). MNase reaction mix was prepared as described[33], except this was scaled down to an input of 80 µg of RNA and 1 µL of NEB MNase 500 U/µL in a total reaction volume of 200 µL. The MNase reaction was incubated at room temperature for 2 hours. All following steps were performed identically[33], except the tRNA removal steps were excluded. Briefly, 500 mL of polysome binding buffer was used to wash the Sephacryl S400 MicroSpin columns (GE Healthcare Life Sciences) three times - spinning the column for 3 minutes at 4 °C at 600 RPM. Polysome binding buffer consisted of 100 µL Igepal CA-630, 500 µL magnesium chloride at 1M, 500 µL EGTA at 0.5 M, 500 µL of NaCl at 5M, 500 µL Tris-HCl pH 8.0. at 1M, and 7.9 mL of RNase-free water. The MNase reaction was applied to the column and centrifuged for 5 minutes at 4 °C. The flow through was purified further with miRNAeasy Mini Kit (Qiagen) using manufacture protocols. Elution was performed at 15 µL volume. rRNA was depleted using RiboZero-rRNA Removal Kit for Bacteria (Illumina) using manufacture protocol, except all reaction volumes and amounts were reduced by 50 percent. This was purified with RNAeasy MinElute Cleanup Kit (Qiagen), eluting in 20 uL. The reaction, in 18 µL volume at a total of 100 ng, was subjected to T4 PNK Reaction (NEB M0201S) with addition of 1µL Superase-In (Invitrogen), 2.2 µL 10X T4 PNK Buffer, and 1 µL T4 PNK (10U/µL). This reaction was purified again with RNAeasy MinElute Cleanup (Qiagen). The concentration was determined with Qubit RNA HS Assay Kit (Illumina). With 100 ng as input, libraries were prepared using NEBNext Small RNA Library Prep Set for Illumina (NEB, E7330), using manufacture protocols. DNA was purified using Minelute PCR Purification Kit (Qiagen).

Small Metatranscriptomics of Fecal Samples

We performed metatranscriptomics as follows: 15 µL of proteinase K (Ambion, 20 mg/mL) was added to 600 µL of lysate. After incubation for 10 minutes at room temperature, samples were centrifuged at 21,000 x g for 3 minutes and the supernatant was collected. An equal volume of Phenol/Chloroform/Isoamyl Alcohol 25:24:1 (pH. 5.2) was applied and vortex for three minutes. The mixture was centrifuged at 21,000 x g for three minutes. The aqueous phase was extracted. This was repeated once more. The final aqueous phase was ethanol precipitated. The RNA was further purified using the RNAeasy Mini plus Kit (Qiagen) using manufacture protocols. Any remaining DNA was degraded via Baseline-ZERO-Dnase (Epicentre) using manufacture protocols. RNA was fragmented for 15 minutes at 70 °C using

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RNA Fragmentation Reagent (Ambion) using manufacture protocols. At this point, the MetaRibo-Seq and small metatranscriptomics protocol completely converge. The fragmented RNA was purified with miRNAeasy Mini Kit (Qiagen) using manufacture protocols. Elution was performed at 15 µL. rRNA was depleted using RiboZero-rRNA Removal Kit for Bacteria (Illumina) using half reactions of manufacture protocol. This was purified with RNAeasy MinElute Cleanup Kit (Qiagen), eluting in 20 uL. The fragments, in 18 µL volume, were subjected to T4 PNK Reaction (NEB M0201S) with addition of 1µL Superase-In (Invitrogen), 2.2 µL 10X T4 PNK Buffer, and 1 µL T4 PNK (10U/µL). This reaction was purified again with RNAeasy MinElute Cleanup (Qiagen). The concentration was determined with Qubit RNA HS Assay Kit (Invitrogen). With 100 ng as input, libraries were prepared using NEBNext Small RNA Library Prep Set for Illumina (NEB, E7330), using manufacture protocols. DNA was purified using MinElute PCR Purification Kit (Qiagen).

Differential Centrifugation and FASP for Metaproteomics

To remove human proteins, fecal samples were subjected to differential centrifugation. 100 mg of fecal sample was suspended in 1x PBS in 1.7 mL Eppendorf tubes. The tubes were centrifuged at 600 x g for 1 minute at room temperature. The supernatant was collected in a clean Eppendorf tube and centrifuged at 10,000 x g for 1 minute at room temperature. The supernatant was decanted and the pellet was resuspended in 1 mL of PBS. The process was repeated once more. The final pellet was resuspended in 2% SDS, 100 mM DTT, and 20 mM Tris HCl, pH 8.8 with protease inhibitor. These cells were subjected to bead beating for 3 minutes with a MiniBeadBeater-16, Model 607. 1mM zirconia/silica beads were used. Tubes were centrifuged for 3 minutes and clarified lysate in the supernatant was collected. Lysate was prepared using FASP [43] with the same minor modifications previously documented [44]. Every step involved a centrifugation step for 15 minutes at 14,000 x g. Samples were diluted tenfold in 8 M urea and loaded into Microcon Ultracel YM-30 filtration devices (Millipore). They were washed in 8 M urea, reduced for 30 minutes in 10 mM DTT, and alkylated in 50 mM iodoacetamide for 20 minutes. Samples were washed three times in 8M urea and two times in 50 mM ammonium bicarbonate. Trypsin (Pierce 90057) (1:100 enzyme-to-protein ratio) was added and incubated overnight at 37 °C. Into a new collection tube, samples were centrifuged and further eluded in 50 µL of 70 percent acetonitrile and 1 percent formic acid. The mixture was brought to dryness for

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

one hour using a Savant SPD121P SpeedVac concentration at 30°C, then resuspended in 0.2 percent formic acid[44].

Metaproteomics

LC-MS/MS analysis was performed by the Stanford University Mass Spectrometry Facility using the Thermo Orbitrap Fusion Tribrid. A Thermo Scientific Orbitrap Fusion coupled to a nanoAcquity UPLC system (Waters, M Class) was used to collect mass spectra (MS). Samples were loaded on a 25 cm sub 100 micron C18 reverse phase column packed in-house with a 80 minute gradient at a flow rate of 0.45 µL/min. The mobile phase consisted of: A (water containing 0.2% formic acid) and B (acetonitrile containing 0.2% formic acid). A linear gradient elution program was used: 0–45 min, 6–20 % (B); 45-60 min, 35 % (B); 60-70 min, 45 % (B); 70-71 min, 70 % (B); 71-77 min, 95 % (B); 77-80 min, 2 % (B). Ions were generated using electrospray ionization in positive mode at 1.6 kV. MS/MS spectra were obtained using Collision Induced Fragmentation (CID) at a setting of 35 of arbitrary energy. Ions were selected for MS/MS in a data dependent, top 15 format with a 30 second exclusion time. Scan range was set to 400 – 1500 m/z. Typical orbitrap mass accuracy was below 2 ppm; for analysis. A 12 ppm window was allowed for precursor ions and 0.4 Da for the fragment ions for CID fragmentation detected in the ion trap. Prokka-predicted[41] proteins were used as a reference database for protein detection using the Byonic proteomics search pipeline v 2.10.5[45]. Byonic parameters include: spectrum-level FDR auto, digest cutter C-terminal cutter, peptide termini semi-specific, maximum number of missed cleavages 2, fragmentation type CID low energy, precursor tolerance 12.0 ppm, fragment tolerance 0.4 ppm, protein FDR cutoff 1 percent. These methods were performed by Stanford Mass Spectrometry Facility (SUMS). Using spectral count output, Normalized Spectral Abundance Factor (NSAF) was calculated by in house scripts.

De Novo Assembly

Quality trimmed metagenomic reads were assembled using metaSPAdes 3.7.0[47]. For all samples, a maximum of 60 million metagenomic reads was used to generate assemblies. Samples sequenced to higher depth were randomly subsetted to 60 million for assembly purposes to both ensure relatively similar numbers of gene predictions and limit computational requirements in assembly and downstream predictions.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

462

463 **Read Mapping, Gene Prediction and Annotation**

464 Reads were trimmed with trim galore version 0.4.0 using cutadapt 1.8.1[46] with flags –q
465 30 and –illumina. Reads were mapped to the annotated assembly using bowtie version 1.1.1[48].
466 To avoid all possible conservation conflicts in downstream differential analysis, only perfect,
467 unique short read alignments were considered. IGV[49] was used to visualize coverage. Prokka
468 v1.12[41] was used to predict genes from the metagenomics assemblies using the –meta option.
469 Annotations were facilitated by many dependencies[38,50,51,52]. For small protein predictions,
470 prodigal[38] was performed after lowering the size threshold from 90 bases to 60 bases.

471

472 **Read density as a function of position**

473 MetaRibo-seq reads were mapped to their metagenomic assemblies. The assembly and
474 aligned reads were analyzed with RiboSeqR[53]. CDSs (coding sequences) were predicted using
475 the findCDS function. Ribosome profiling counts for predicted CDSs were determined with the
476 sliceCounts function. CDSs were filtered to contain at least 10 reads.

477

478 **Taxonomic Classification of Technologies**

479 Reads mapping specifically to Prokka-predicted[41] coding regions were counted. That
480 entire genomic element was input into One Codex[54] for classification equal to the number of
481 reads mapping to it. This enabled fair comparisons between technologies, as the small
482 metatranscriptomics and MetaRibo-Seq reads can be too small to classify individually with *k*-
483 mer-based approaches. Though metagenomic reads were long enough to be classified directly,
484 they were also subject to the same analysis – entire genes are classified in equal number to the
485 reads overlapping them. Thus, all taxonomy plots represent entire gene classifications and are
486 dependent on the assembly.

487

488 **Differential Analysis**

489 The number of reads mapping to a given region were calculated with bedtools multicov[59].
490 Strandedness was enforced for metatranscriptomics and MetaRibo-Seq. All differential analyses
491 were performed using these counts with all conditions performed in duplicate via DESeq2[60]. A
492 gene was considered differential if it had log2fold change above 1 or below -1, while also

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

493 reaching an FDR < 0.05. Results were displayed as volcano plots or tables. Heatmaps were
494 created using gplots[61]. Reads per Kilobase Million (RPKM) calculations were performed using
495 in house scripts.
496
497 **Statistical Analysis**
498 All Pearson correlations were calculated in R using the Hmisc package[55]. Scatterplots
499 were created with ggplot2[56]. Significance between Pearson correlations was assigned via cocor[57].
500 Significant differences between RPKM values were assigned using the Kruskal-Wallis test.
501 Significance was assigned as * p value < 0.05, *** p value < 0.001. Zou's[58] 95 percent
502 confidence intervals were assigned *** if there is no overlap with 0 in the interval.
503
504 **Protein Clustering Analysis**
505 For analyses independent of gene annotation, significantly translationally regulated
506 proteins were clustered using Cd-hit[62] with 70 percent amino acid identity. Representative
507 sequences were input into Blast2GO[40] using the nr database. Small protein predictions with
508 translational evidence were also clustered using this same approach. Coding potential was
509 assessed using RNAcode[39] using the p value assigned to the predicted reading frame.
510
511 **Triplet Periodicity Analysis**
512 Using the same default parameters as read density as a function of position, triplet
513 periodicity was called using RiboSeqR[53]. To analyze triplet periodicity of specific genera,
514 assembled contigs were classified using One Codex[54]. Contigs that classified into a specific
515 genus were binned together. Only reads mapping specifically to these bins were considered.
516
517 **GO Analysis**
518 Based on differential genes from DESeq2[60] analyses, UniProt[63] genes annotated by
519 Prokka[41] were input into David Functional Annotation[64,65]. All species detected were used as
520 background for these metagenomic analyses.
521
522 **Pathway Analysis**

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prokka[41] predicted genes with associated EC (enzyme) numbers were considered. For a given sample, all the reads mapping to any gene with a specific EC number were summed for metatranscriptomics and MetaRibo-Seq. DESeq2[60] called differential enzymes using MicrobiomeAnalyst[66]. Network mapping is performed to identify pathways corresponding to differential enzymes.

Acknowledgements

The authors would like to thank Tessa M Andermann, Ekaterina Tkachenko, and Joyce B. Kang for major contributions to the fecal biobank of bone marrow transplant patients at Stanford hospital. We would like to thank the patients and nurses involved in collection. We thank Christina Wyss-Coray for collecting Alzheimer's samples used in this study. We appreciate sample collection feedback from Victor Henderson and Tony Wyss-Coray. We thank Anshul Kundaje and Georgi Marinov for helpful computational analysis guidance. We would also like to thank Stanford University Mass Spectrometry for performing and analyzing mass spectrometry on the FASP fecal samples. Sequencing costs were supported via NIH S10 Shared Instrumentation Grant (1S10OD02014101) and Damon Runyon Clinical Investigator Award to ASB, Stanford ADRC grant # P50AG047366. B.J.F is supported by National Science Foundation Graduate Research Fellowship DGE-114747.

Figure Legends

Figure 1

Workflow of ribosome profiling.

(A) Experimental workflow of MetaRibo-Seq. Chloramphenicol halts translation, the bacterial community is lysed, MNase is used to create footprints, and footprints are converted to sequencing libraries.

(B) Computational workflow of the multi-omics approach. *De novo* assemblies are created and annotated, predicted genes are quantified at multi-omic levels, and taxonomy, correlations, and differential abundance are determined from these results.

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

554 MetaRibo-Seq signal captures diversity in a metagenomic context.
555 (A) Genus-level classifications of all sequencing technologies performed on Samples A, B, C, D,
556 and E. Replicates for metatranscriptomics and MetaRibo-Seq are shown for reproducibility. Taxa
557 represented below three percent are grouped into "Other" for visual purposes.
558 (B) Shannon diversities across technologies for these samples are displayed.
559
560 **Figure 3**
561 MetaRibo-Seq signal is characteristic of chloramphenicol-treated ribosome profiling in bacteria
562 across diverse samples.
563 (A-D) Average MetaRibo-Seq signal across genes and flanking regions for Sample A, B, C, and
564 D, respectively. Every predicted open reading frame containing at least 10 reads are included in
565 the analysis**.**
566
567 **Figure 4**
568 In a low diversity fecal sample, MetaRibo-Seq is a significantly better predictor than
569 metatranscriptomics of protein abundance and ATP synthase stoichiometry in *E. coli*.
570 (A) In Sample E, there are 1503 proteomically-detectable genes. Only focusing on these genes,
571 we taxonomically classify the entire gene in equal number to the reads (for metatranscriptomics
572 and MetaRibo-Seq) or spectral counts (for metaproteomics) assigned to it (see Methods).
573 (B) Scatterplot of MetaRibo-Seq RPKM and metaproteomics NSAF log-scaled for these 1503
574 genes.
575 (C) Scatterplot of MetaRibo-Seq RPKM and metaproteomics NSAF log-scaled only for the 928
576 proteomically-detected genes predicted from isolate *E. coli*.
577 (D) Pearson correlations for pairwise comparisons across technologies. Blue bars indicate
578 Pearson correlations pertaining to the entire metagenomic Sample E. The correlations between
579 metatranscriptomics vs. MetaRibo-Seq, metatranscriptomics vs. metaproteomics, and MetaRibo-
580 Seq vs. metaproteomics are 0.85, 0.39, and 0.46, respectively. All are significant (p value $< 2^{-16}$).
581 MetaRibo-Seq is a significantly better predictor of protein levels with Zou's[58] 95 % confidence
582 interval between -0.0917 to -0.0487. Red bars indicate Pearson correlations pertaining to the
583 isolated *E. coli* in the sample. The correlations between metatranscriptomics vs. MetaRibo-Seq,
584 metatranscriptomics vs. metaproteomics, and MetaRibo-Seq vs. metaproteomics are 0.90, 0.53,

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and 0.64, respectively. All are significant (p value < 2^{-16}). MetaRibo-Seq is a significantly better predictor of protein levels with Zou's[58] 95 % confidence interval between -0.1352 to -0.0868.
(E) F0F1 ATP synthase in *E. coli* forms with a specific stoichiometry as visualized.
(F) Correlation between log-scaled metatranscriptomics RPKM and expected ATP synthase stoichiometry in complex. The Pearson correlation is 0.56 (p value = 0.1475)
(G) Correlation between log-scaled MetaRibo-Seq RPKM and expected ATP synthase stoichiometry in complex. The Pearson correlation is 0.92 (p value = 0.0012).

Figure 5

MetaRibo-Seq signal significantly correlates to metaproteomics and is enriched for these products
(A) For Sample A, scatterplot of MetaRibo-Seq RPKM (Reads per Kilobase Million) and metaproteomics NSAF (Normalized Spectral Abundance Factor) both log10-scaled. 497 genes are displayed.
(B) Pairwise Pearson correlations between log-scaled metatranscriptomics RPKM, MetaRibo-Seq RPKM, and metaproteomics NSAF for these 497 genes. Pearson correlations are 0.88, 0.26, and 0.32 for metatranscriptomics vs. MetaRibo-Seq, metatranscriptomics vs. metaproteomics, and MetaRibo-Seq vs. metaproteomics, respectively. All are significant (p value < 2^{-16}). MetaRibo-Seq is a significantly better predictor of protein levels than metatranscriptomics for these proteins in Sample A with a Zou's[58] 95 % confidence interval between -0.1322 and -0.0480.
(C) Metatranscriptomics and MetaRibo-Seq RPKM for all predicted genes compared to those detected by metaproteomics. Both metatranscriptomic and MetaRibo-Seq signal for proteomically-detected proteins are significantly enriched (p value < 2^{-16}).
(D) For Sample B, scatterplot of MetaRibo-Seq RPKM and metaproteomics NSAF both log10-scaled. 497 genes are displayed.
(E) Pairwise Pearson correlations between log-scaled metatranscriptomics RPKM, MetaRibo-Seq RPKM, and metaproteomics NSAF for these 480 genes. Pearson correlations are 0.89, 0.36, and 0.34 for metatranscriptomics vs. MetaRibo-Seq, metatranscriptomics vs. metaproteomics, and MetaRibo-Seq vs. metaproteomics, respectively. All are significant (p value < 2^{-16}).

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(F) Metatranscriptomics and MetaRibo-Seq RPKM for all predicted genes compared to those detected by metaproteomics. Both metatranscriptomic and MetaRibo-Seq signal for proteomically-detected proteins are significantly enriched (p value < 2^{-16}).

Figure 6

Genes that are consistently translationally regulated emerged across samples and taxa.

(A) For Samples, A, B, C, and D, we show the number of total gene predictions via Prokka[41]. We performed DESeq2[60] on these individually, comparing metatranscriptomics to MetaRibo-Seq. We show the number of those genes identified as translationally (absolute value(log2 fold change) > 1 and FDR < 0.05). We predict 223630, 196683, 272895, and 173624 genes from Sample A, B, C, and D, respectively. Among these, 11872, 6580, 15188, and 8647, respectively, are called significant translationally regulated genes.

(B) Significant translationally regulated genes are converted to proteins and clustered at 70 percent amino acid identity (see Methods). The number of clusters with specific numbers of sequences are displayed, jittered and color-coded for each sample.

(C) The number of clusters with specific numbers of sequences combined across Samples A, B, C, and D

(D) If the combined cluster contained at least 5 sequences, this gene is considered consistently translationally regulated. 607 clusters met this requirement. For these clusters, the representative sequence (see Methods) is selected to represent the entire cluster. These representatives were input into Blast2GO to assign GO terms based on protein sequence.

Figure S1

MetaRibo-Seq demonstrates some weak signs of overall codon-resolution.

(A-D) Triplet periodicity across footprint lengths for Sample A, B, C, and D, respectively. Colors indicate which frame a read falls within.

Figure S2

MetaRibo-Seq demonstrates stronger codon resolution in taxa-specific analyses.

(A) All contigs assigned to the genera Bacteroides are considered from Sample A and B, respectively. Only these contigs are considered in triplet periodicity analyses.

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(B and C) The sampe triplet periodicity analysis for Faecalibacterium and Alistipes, respectively.

Figure S3

Small protein predictions with translational evidence demonstrate strong conservation.

(A) Breakdown of small protein predictions across Samples A, B, C, and D, and those with translational evidence (MetaRibo-Seq signal above 0.5 RPKM).

(B) Genus-level classification of small protein predictions across Samples A, B, C, and D. Relative proportions of small proteins including all predictions and only those with translational evidence are provided.

(C) The 6,774 small proteins with translational evidence are clustered at 70 percent amino acid identity. 47 clusters with at least 4 members are identified. Among these, 21 clusters possess evolutionary signatures indicative of coding regions (p values < 0.001) via RNAcode[39].

(D) In dark red, clustering with 70 percent protein identity of small proteins with translational evidence – 6,774 proteins across all samples. In orange, clustering with 70 percent amino acid identity of 6,744 small proteins randomly chosen from prodigal predictions; an equal number of proteins as those with translational evidence from each sample were randomly chosen.

Figure S4

DESeq2 models adequately fit dispersion in comparing metatranscriptomics to MetaRibo-Seq.

(A-D) Dispersion plots of DESeq2 models fit to Samples A, B, C, and D, respectively.

Figure S5

Numerous genes are consistently translationally regulated in the fecal microbiota

(A) For Samples A, B, C and D, we show genera level classifications of translationally regulated genes assigned UniProt[31] protein IDs.

(B-E) Volcano plots comparing metatranscriptomics and MetaRibo-Seq in Sample A, B, C, and D, respectively. Significant genes are colored in red. The four most consistently translationally regulated genes are also denoted.

Figure S6

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In depth analysis of consistently translationally regulated genes with assigned names in the fecal microbiota across Samples A, B, C, and D.
(A) Top 40 most common translationally regulated genes across samples. Heatmap intensity represents the number of times the gene appeared differential in a sample.
(B) GO analysis including genes that are differentially translated at least 5 times across samples.

Table S1

Mapping statistics to *de novo* references. This table shows the number of reads, percentage of reads mapping to the assembly, and percentage of reads overlapping regions annotated as coding.

Table S2

GO analysis of significant translationally regulated genes for Samples A, B, C, and D. Only genes assigned UniProt[63] protein IDs are considered. Significant IDs are called with DESeq2[60]. These are input into David[64].

Table S3

Pairwise comparison of metatranscriptomics to MetaRibo-Seq across Sample A, B, C, and D. Only predicted genes with EC numbers are considered. Significant differences in EC numbers are called with DESeq2[60]. Negative log2FC means lower in the technology first listed in the tab under consideration. EC numbers and their associated adjusted p value are input into MicrobiomeAnalyst[66] to determine overrepresented pathways based on differential EC numbers.

Table S4

We display all correlations between replicates and technologies for Samples A, B, C, and D. For each sample, we provide correlations between metatranscriptomics replicates, MetaRibo-Seq replicates, metagenomics versus metatrancriptomics, metagenomics versus MetaRibo-Seq, metatranscriptomics versus MetaRibo-Seq, and metagenomics versus translation efficiency (TE – ratio of MetaRibo-Seq and metatranscriptomics).

File S1

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We provide sequences for every protein in Samples A, B, C, and D that are translationally regulated in the gut microbiota. Translationally regulated proteins (.faa) and 70 percent identity clustering (.cltsr) are provided for Samples A, B, C, and D. The representative sequences for clusters (.faa) with more than 5 sequences for samples are also provided. The sequence name itself denotes which sample the sequence is found in. Any sequence that begins with a specific identifier can be linked to a sample: HDALDHFB = Sample A, HENMDNCI =Sample B, PJJNKMKO = Sample C, GPBGFMPE = Sample D. Blast2GO[40] results for the representative sequences of consistent clusters are provided.

File S2

We contribute sequences of small proteins we identified. Small protein predictions using prodigal[38] with lower size threshold (60 bp) for Samples A, B, C, and D individually (.faa) are provided. Small protein predictions with MetaRibo-Seq RPKM above 0.5 for the four samples individually (.faa) are given. Combined clustering of these small proteins with translational evidence are provided (.clstr). Those translated clusters with more than 3 sequences and with RNAcode p values < 0.001 are provided in the "smallproteinsequences" folder named by the cluster they belong to.

Bibliography

1. Klassen, J. L. Defining microbiome function. *Nat. Microbiol.* **3**, 864–869 (2018).
2. Petriz, B. A. & Franco, O. L. Metaproteomics as a Complementary Approach to Gut Microbiota in Health and Disease. *Front. Chem.* **5**, 4 (2017).
3. Wilmes, P., Heintz-Buschart, A. & Bond, P. L. A decade of metaproteomics: Where we stand and what the future holds. *Proteomics* **15**, 3409–3417 (2015).
4. Lichtman, J. S., Sonnenburg, J. L. & Elias, J. E. Monitoring host responses to the gut microbiota. *ISME J.* **9**, 1908–1915 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5. Hyatt, D., LoCascio, P. F., Hauser, L. J. & Uberbacher, E. C. Gene and translation initiation site prediction in metagenomic sequences. *Bioinforma. Oxf. Engl.* **28**, 2223–2230 (2012).

6. Franzosa, E. A. *et al.* Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci.* **111**, E2329–E2338 (2014).

7. Mehta, R. S. *et al.* Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* **3**, 347–355 (2018).

8. Schirmer, M. *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* **3**, 337–346 (2018).

9. Metatranscriptome of human faecal microbial communities in a cohort of adult men | Nature Microbiology. Available at: https://www.nature.com/articles/s41564-017-0084-4. (Accessed: 17th April 2018)

10. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).

11. Ingolia, N. T. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell* **165**, 22–33 (2016).

12. Cenik, C. *et al.* Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* **25**, 1610–1621 (2015).

13. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

14. Mohammad, F., Woolstenhulme, C. J., Green, R. & Buskirk, A. R. Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Rep.* **14**, 686–694 (2016).

15. Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).

16. Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-Precision Analysis of Translational Pausing by Ribosome Profiling in Bacteria Lacking EFP. *Cell Rep.* **11**, 13–21 (2015).

17. Lalanne, J.-B. *et al*. Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry. *Cell* **173**, 749-761.e38 (2018).

18. Meyer, M. M. The role of mRNA structure in bacterial translational regulation. *Wiley Interdiscip. Rev. RNA* **8**, (2017).

19. Zengel, J. M. & Lindahl, L. Diverse Mechanisms for Regulating Ribosomal Protein Synthesis in Escherichia coli. in *Progress in Nucleic Acid Research and Molecular Biology* (eds. Cohn, W. E. & Moldave, K.) **47**, 331–370 (Academic Press, 1994).

20. Lindahl, L., Jaskunas, S. R., Dennis, P. P. & Nomura, M. Cluster of genes in Escherichia coli for ribosomal proteins, ribosomal RNA, and RNA polymerase subunits. *Proc. Natl. Acad. Sci.* **72**, 2743–2747 (1975).

21. Fallon, A. M., Jinks, C. S., Strycharz, G. D. & Nomura, M. Regulation of ribosomal protein synthesis in Escherichia coli by selective mRNA inactivation. *Proc. Natl. Acad. Sci.* **76**, 3411–3415 (1979).

22. Dean, D. & Nomura, M. Feedback regulation of ribosomal protein gene expression in Escherichia coli. *Proc. Natl. Acad. Sci.* **77**, 3590–3594 (1980).

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. Bucca, G. *et al.* Translational control plays an important role in the adaptive heat-shock response of Streptomyces coelicolor. *Nucleic Acids Res.* **46**, 5692–5703 (2018).

24. Sawyer, E. B., Grabowska, A. D. & Cortes, T. Translational regulation in mycobacteria and its implications for pathogenicity. *Nucleic Acids Res.* doi:10.1093/nar/gky574

25. C. Taylor, R. *et al.* Changes in translational efficiency is a dominant regulatory mechanism in the environmental response of bacteria. *Integr. Biol.* **5**, 1393–1406 (2013).

26. Morita, Y., Gilmour, C., Metcalf, D. & Poole, K. Translational Control of the Antibiotic Inducibility of the PA5471 Gene Required for mexXY Multidrug Efflux Gene Expression in Pseudomonas aeruginosa. *J. Bacteriol.* **191**, 4966–4975 (2009).

27. Starosta, A. L., Lassak, J., Jung, K. & Wilson, D. N. The bacterial translation stress response. *FEMS Microbiol. Rev.* **38**, 1172–1201 (2014).

28. Jeong, Y. *et al.* The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nat. Commun.* **7**, 11605 (2016).

29. Mathieu, A. *et al.* Discovery and Function of a General Core Hormetic Stress Response in E. coli Induced by Sublethal Concentrations of Antibiotics. *Cell Rep.* **17**, 46–57 (2016).

30. Bullwinkle, T. J. & Ibba, M. Translation quality control is critical for bacterial responses to amino acid stress. *Proc. Natl. Acad. Sci.* **113**, 2252–2257 (2016).

31. Oh, E. *et al.* Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. *Cell* **147**, 1295–1308 (2011).

32. Reck, M. *et al.* Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics* **16**, 494 (2015).

33. Latif, H. *et al.* A streamlined ribosome profiling protocol for the characterization of microorganisms. *BioTechniques* **58**, 329–332 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34. Bag, S. *et al.* An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. *Sci. Rep.* **6**, 26775 (2016).

35. Conditions for ethanol precipitation of active 50 S ribosomal subunits from Escherichia coli - ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/0003269780901372. (Accessed: 15th April 2018)

36. Tamburini, F. B. *et al.* Precision Identification of Diverse Bloodstream Pathogens from the Gut Microbiome. *bioRxiv* 310441 (2018). doi:10.1101/310441

37. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009).

38. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

39. Washietl, S. *et al.* RNAcode: Robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* **17**, 578–594 (2011).

40. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).

41. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**, 2068–2069 (2014).

42. Ndah, E. *et al.* REPARATION: Ribosome Profiling Assisted (Re-)Annotation of Bacterial genomes. *bioRxiv* 113530 (2017). doi:10.1101/113530

43. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

822 44. Tanca, A. *et al.* A straightforward and efficient analytical pipeline for metaproteome
823 characterization. *Microbiome* **2**, 49 (2014).
824 45. Bern, M., Kil, Y. J. & Becker, C. Byonic: Advanced Peptide and Protein Identification
825 Software. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **CHAPTER**,
826 Unit13.20 (2012).
827 46. Nurk, S. *et al.* Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads.
828 in *Research in Computational Molecular Biology* 158–170 (Springer, Berlin, Heidelberg,
829 2013). doi:10.1007/978-3-642-37195-0_13
830 47. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
831 *EMBnet.journal* **17**, 10–12 (2011).
832 48. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient
833 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
834 49. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
835 50. Madden, T. *The BLAST Sequence Analysis Tool*. (National Center for Biotechnology
836 Information (US), 2003).
837 51. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes
838 in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
839 52. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity
840 searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
841 53. riboSeqR. *Bioconductor* Available at: http://bioconductor.org/packages/riboSeqR/.
842 (Accessed: 28th February 2017)
843 54. Minot, S. S., Krumm, N. & Greenfield, N. B. One Codex: A Sensitive and Accurate Data
844 Platform for Genomic Microbial Identification. *bioRxiv* 027607 (2015). doi:10.1101/027607

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

55. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* **26**, 841–842 (2010).

56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

57. Warnes, G. R. *et al. gplots: Various R Programming Tools for Plotting Data*. (2016).

58. Jr, F. E. H. & others, with contributions from C. D. and many. *Hmisc: Harrell Miscellaneous*. (2018).

59. Wickham, H., Chang, W. & RStudio. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. (2016).

60. Diedenhofen, B. *cocor: Comparing Correlations*. (2016).

61. Zou, G. Y. Toward using confidence intervals to compare correlations. *Psychol. Methods* **12**, 399–413 (2007).

62. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

63. Bateman, A. *et al*. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/482430; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

866 66. Dhariwal, A. *et al.* MicrobiomeAnalyst: a web-based tool for comprehensive statistical,
867 visual and meta-analysis of microbiome data. *Nucleic Acids Res.* (2017).
868 doi:10.1093/nar/gkx295
869

Figure 1

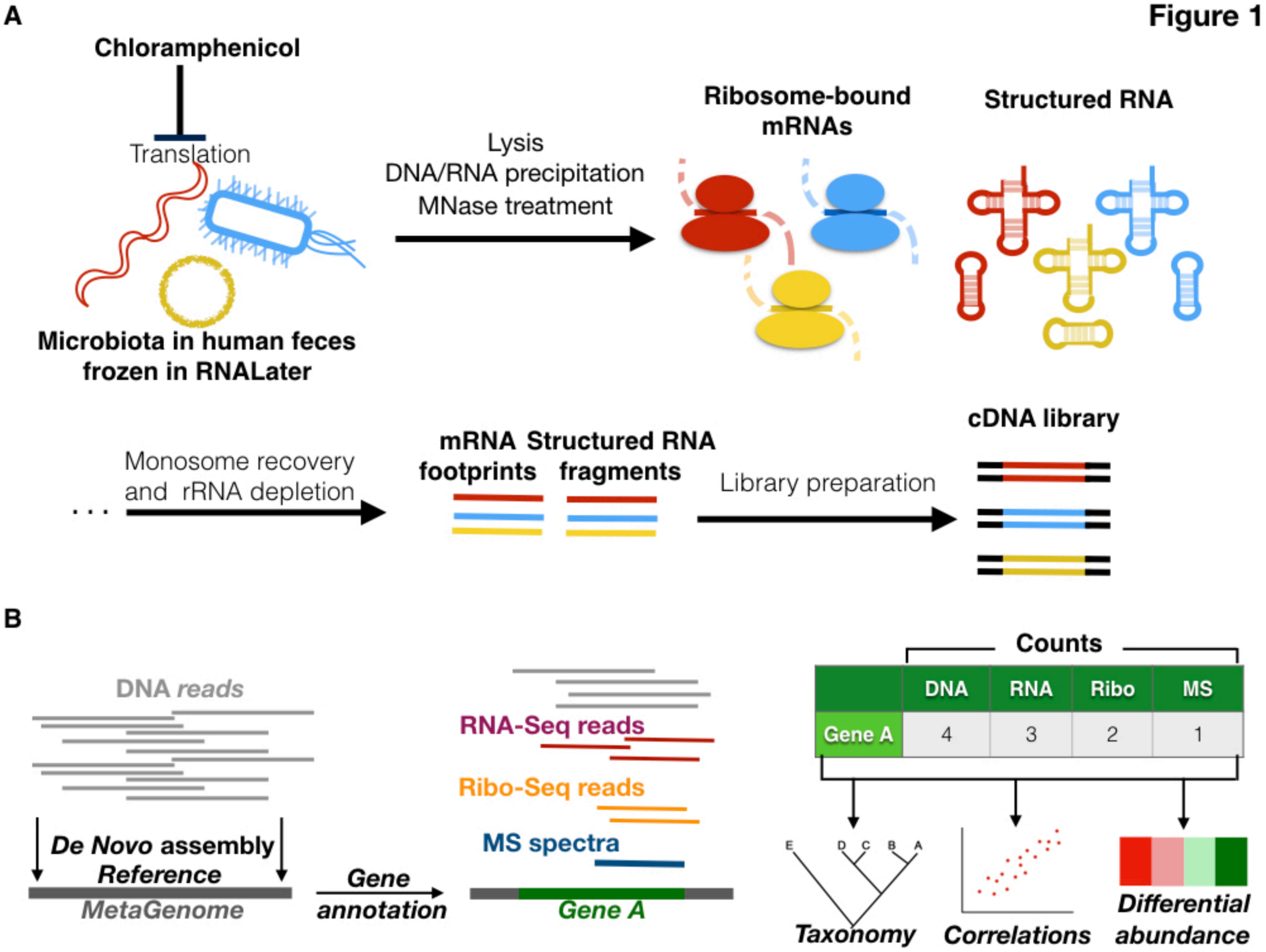

A
Chloramphenicol
Translation
Microbiota in human feces
frozen in RNALater
Lysis
DNA/RNA precipitation
MNase treatment
Ribosome-bound
mRNAs
Structured RNA
Monosome recovery
and rRNA depletion
mRNA
footprints
Structured RNA
fragments
Library preparation
cDNA library
B
DNA reads
De Novo assembly
Reference
MetaGenome
Gene
annotation
RNA-Seq reads
Ribo-Seq reads
MS spectra
Gene A
Counts
DNA
RNA
Ribo
MS
Gene A
4
3
2
1
E
D
C
B
A
Taxonomy
Correlations
Differential
abundance

Figure 2

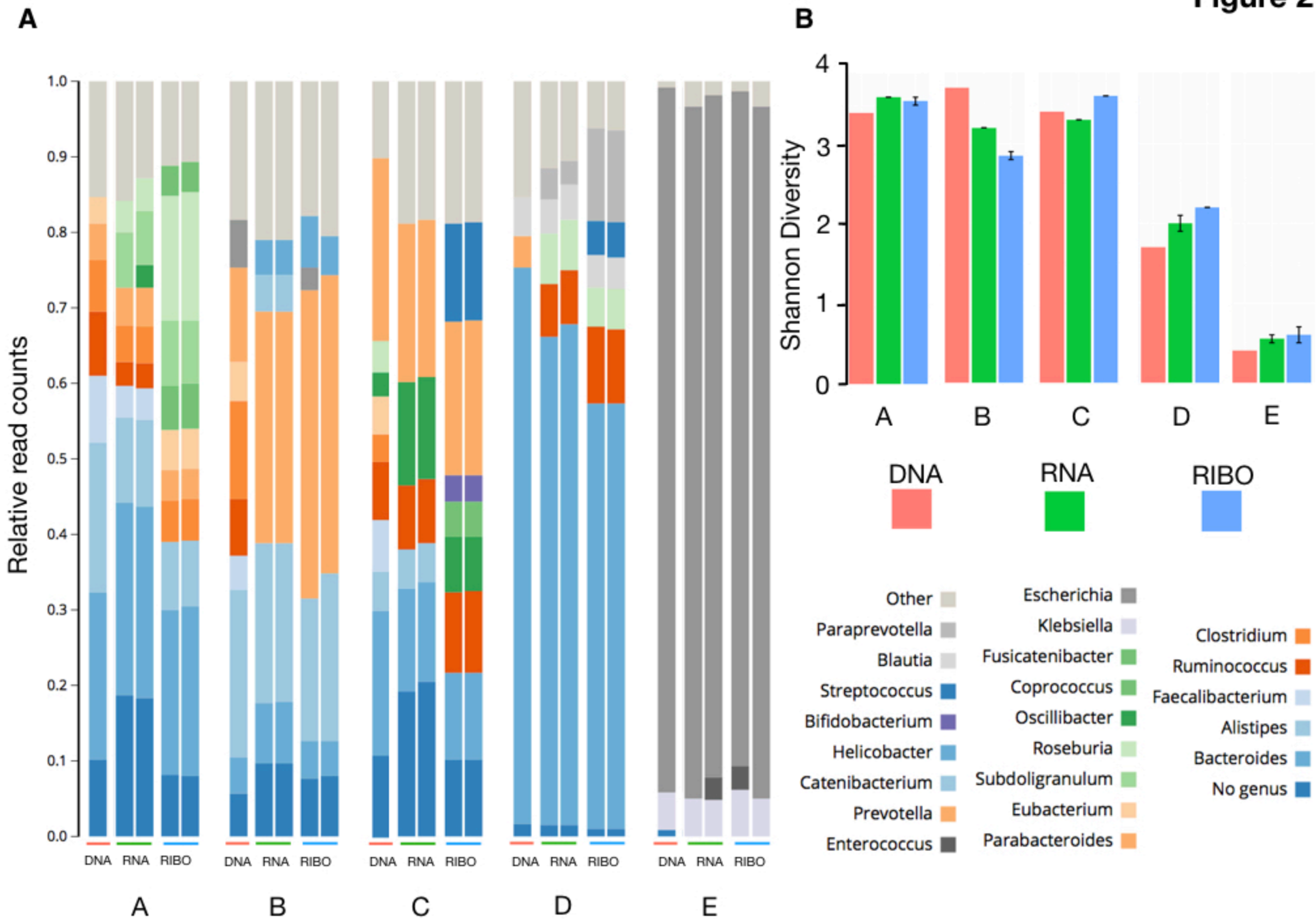
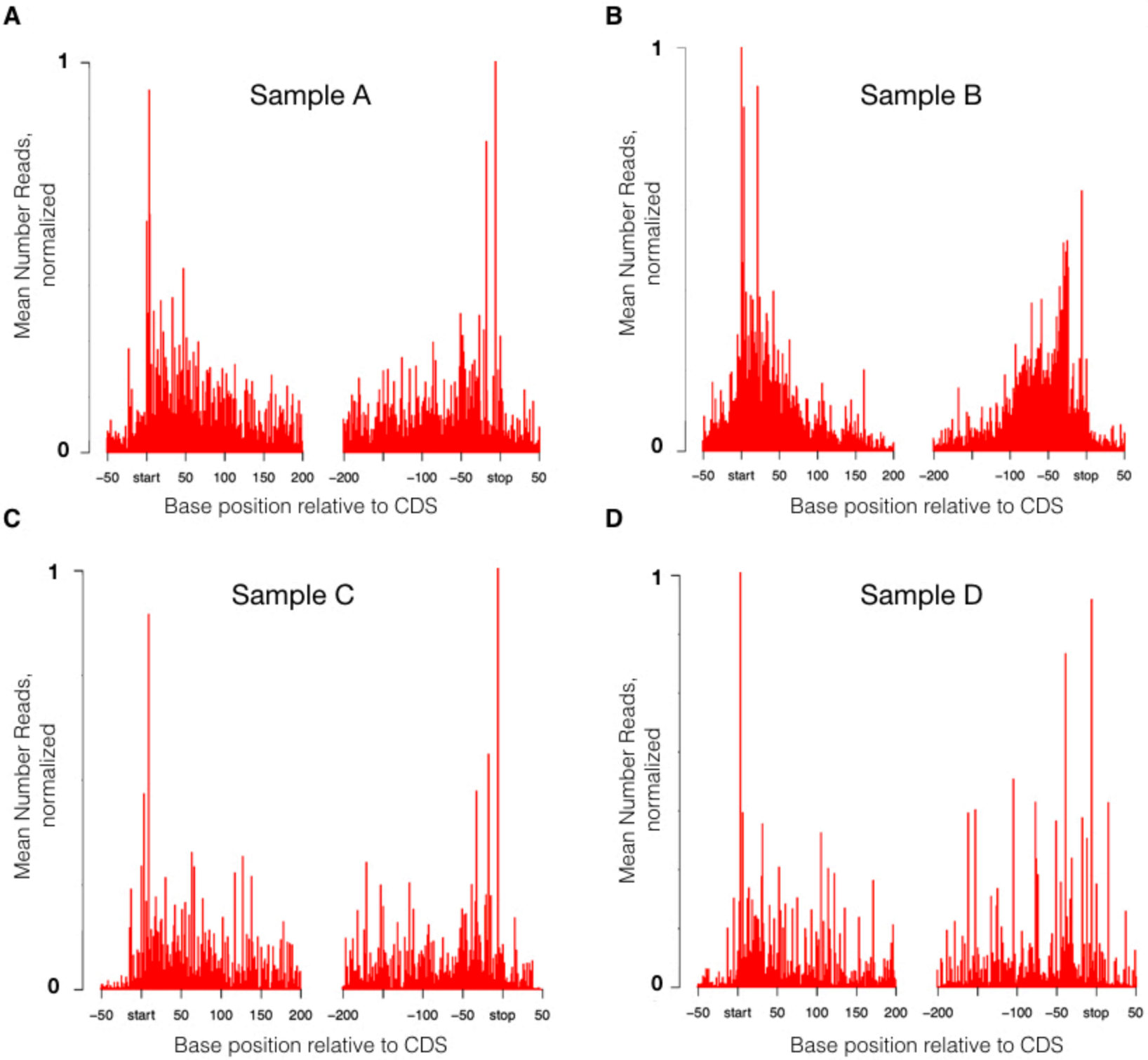

A
B
Relative read counts
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
DNA RNA RIBO
A
B
C
D
E
Shannon Diversity
4
3
2
1
0
DNA
RNA
RIBO
Other
Paraprevotella
Blautia
Streptococcus
Bifidobacterium
Helicobacter
Catenibacterium
Prevotella
Enterococcus
Escherichia
Klebsiella
Fusicatenibacter
Coprococcus
Oscillibacter
Roseburia
Subdoligranulum
Eubacterium
Parabacteroides
Clostridium
Ruminococcus
Faecalibacterium
Alistipes
Bacteroides
No genus

Figure 3


A
Sample A
Mean Number Reads, normalized
1
0
−50 start 50 100 150 200 −200 −100 −50 stop 50
Base position relative to CDS
B
Sample B
Mean Number Reads, normalized
1
0
−50 start 50 100 150 200 −200 −100 −50 stop 50
Base position relative to CDS
C
Sample C
Mean Number Reads, normalized
1
0
−50 start 50 100 150 200 −200 −100 −50 stop 50
Base position relative to CDS
D
Sample D
Mean Number Reads, normalized
1
0
−50 start 50 100 150 200 −200 −100 −50 stop 50
Base position relative to CDS

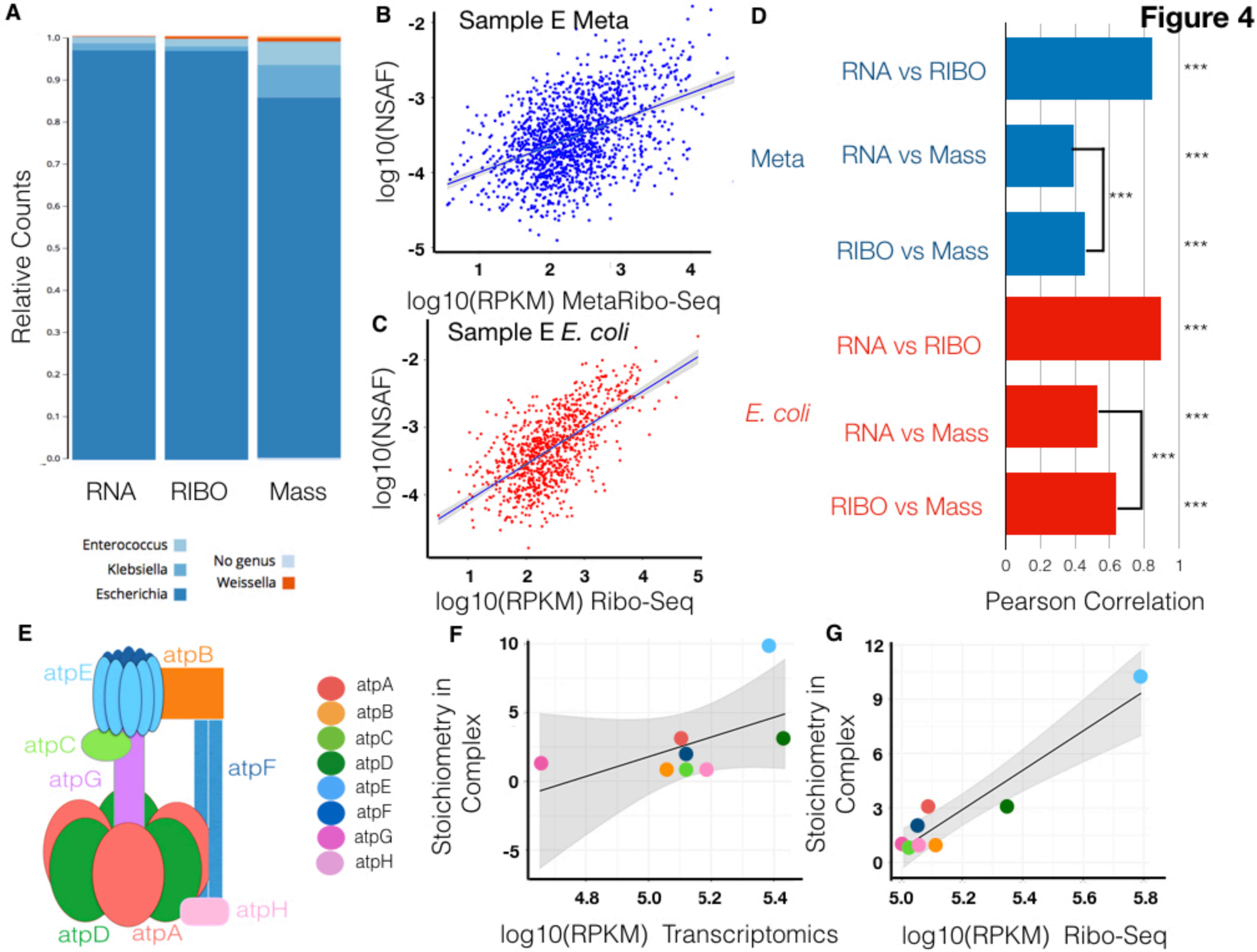


Figure 4
A
Relative Counts
RNA
RIBO
Mass
Enterococcus
Klebsiella
Escherichia
No genus
Weissella
B
Sample E Meta
log10(NSAF)
log10(RPKM) MetaRibo-Seq
C
Sample E E. coli
log10(NSAF)
log10(RPKM) Ribo-Seq
D
Meta
RNA vs RIBO
RNA vs Mass
RIBO vs Mass
E. coli
RNA vs RIBO
RNA vs Mass
RIBO vs Mass
***
Pearson Correlation
E
atpA
atpB
atpC
atpD
atpE
atpF
atpG
atpH
F
Stoichiometry in Complex
log10(RPKM) Transcriptomics
G
Stoichiometry in Complex
log10(RPKM) Ribo-Seq

Figure 5

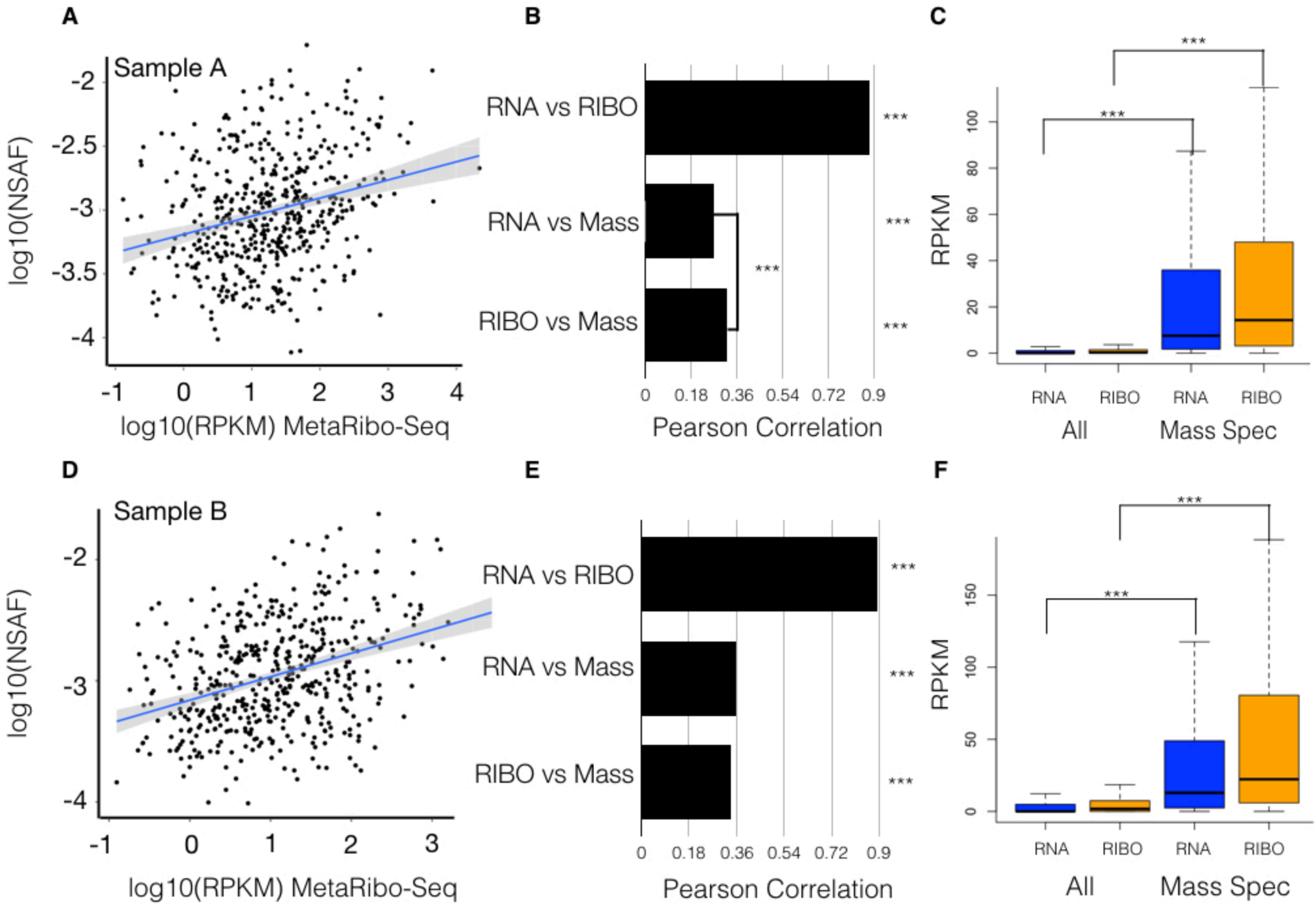

A
Sample A
log10(NSAF)
log10(RPKM) MetaRibo-Seq
B
RNA vs RIBO
RNA vs Mass
RIBO vs Mass
***
Pearson Correlation
C
RPKM
RNA
RIBO
All
Mass Spec
D
Sample B
E
F

Figure 6

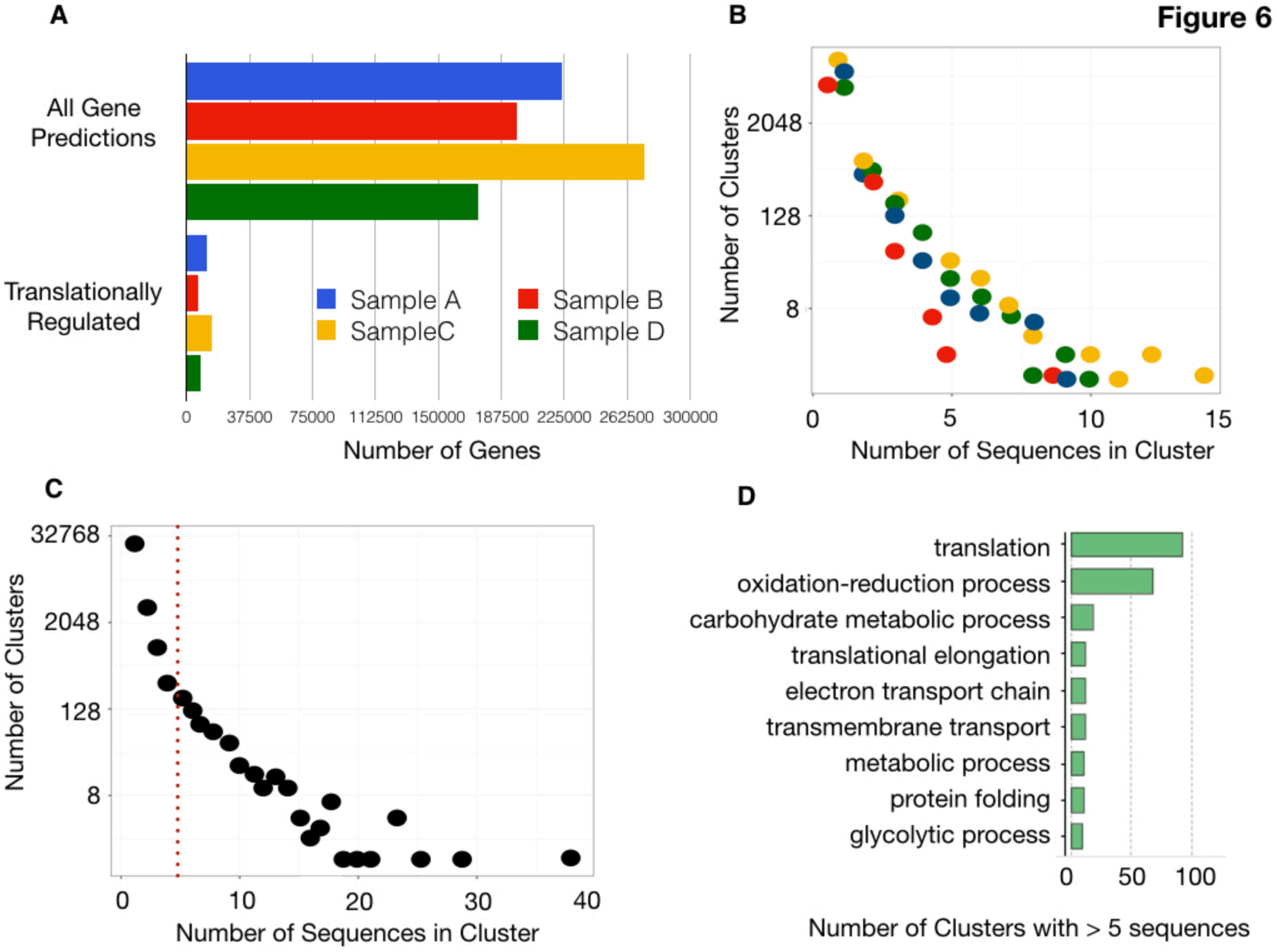

A
All Gene Predictions
Translationally Regulated
Sample A
SampleC
Sample B
Sample D
0
37500
75000
112500
150000
187500
225000
262500
300000
Number of Genes
B
Number of Clusters
2048
128
8
0
5
10
15
Number of Sequences in Cluster
C
Number of Clusters
32768
2048
128
8
0
10
20
30
40
Number of Sequences in Cluster
D
translation
oxidation-reduction process
carbohydrate metabolic process
translational elongation
electron transport chain
transmembrane transport
metabolic process
protein folding
glycolytic process
0
50
100
Number of Clusters with > 5 sequences

Figure S1

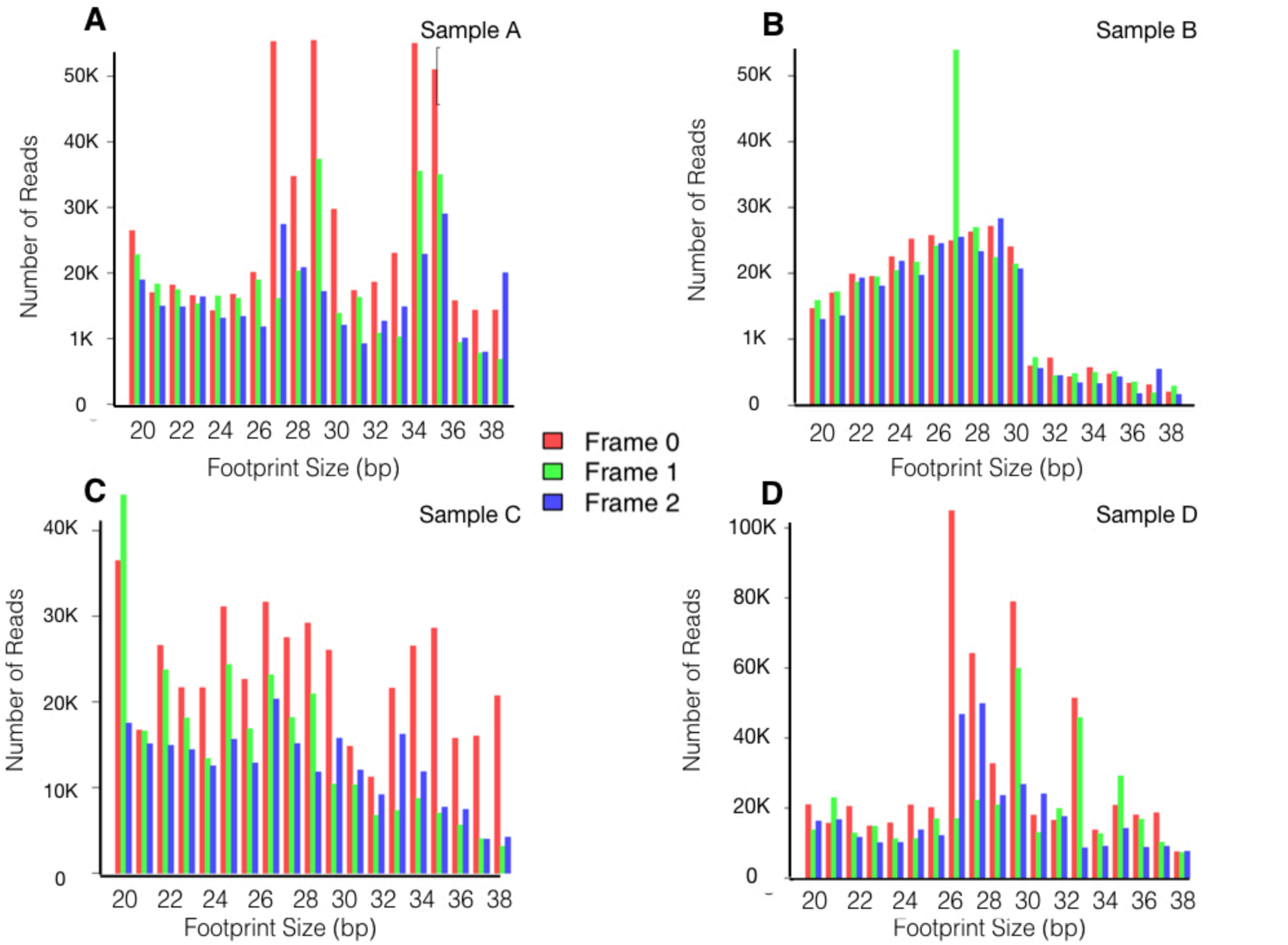

A
Sample A
Number of Reads
50K
40K
30K
20K
1K
0
20 22 24 26 28 30 32 34 36 38
Footprint Size (bp)
B
Sample B
Number of Reads
50K
40K
30K
20K
1K
0
20 22 24 26 28 30 32 34 36 38
Footprint Size (bp)
Frame 0
Frame 1
Frame 2
C
Sample C
Number of Reads
40K
30K
20K
10K
0
20 22 24 26 28 30 32 34 36 38
Footprint Size (bp)
D
Sample D
Number of Reads
100K
80K
60K
40K
20K
0
20 22 24 26 28 30 32 34 36 38
Footprint Size (bp)

Figure S2

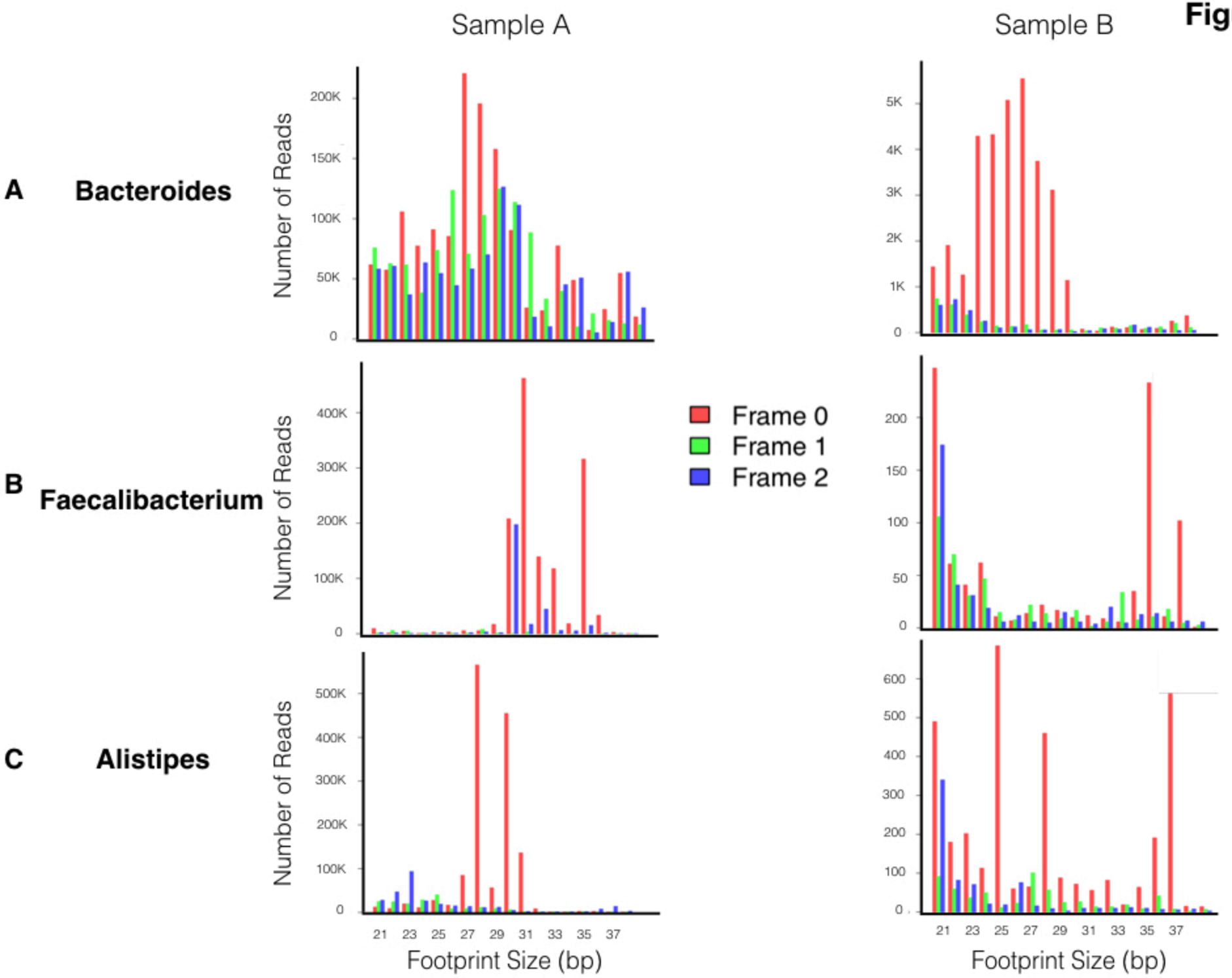

Sample A
Sample B
A Bacteroides
B Faecalibacterium
C Alistipes
Number of Reads
Frame 0
Frame 1
Frame 2
Footprint Size (bp)
200K
150K
100K
50K
0
5K
4K
3K
2K
1K
400K
300K
500K
600
200
150
100
50
21 23 25 27 29 31 33 35 37

Figure S3

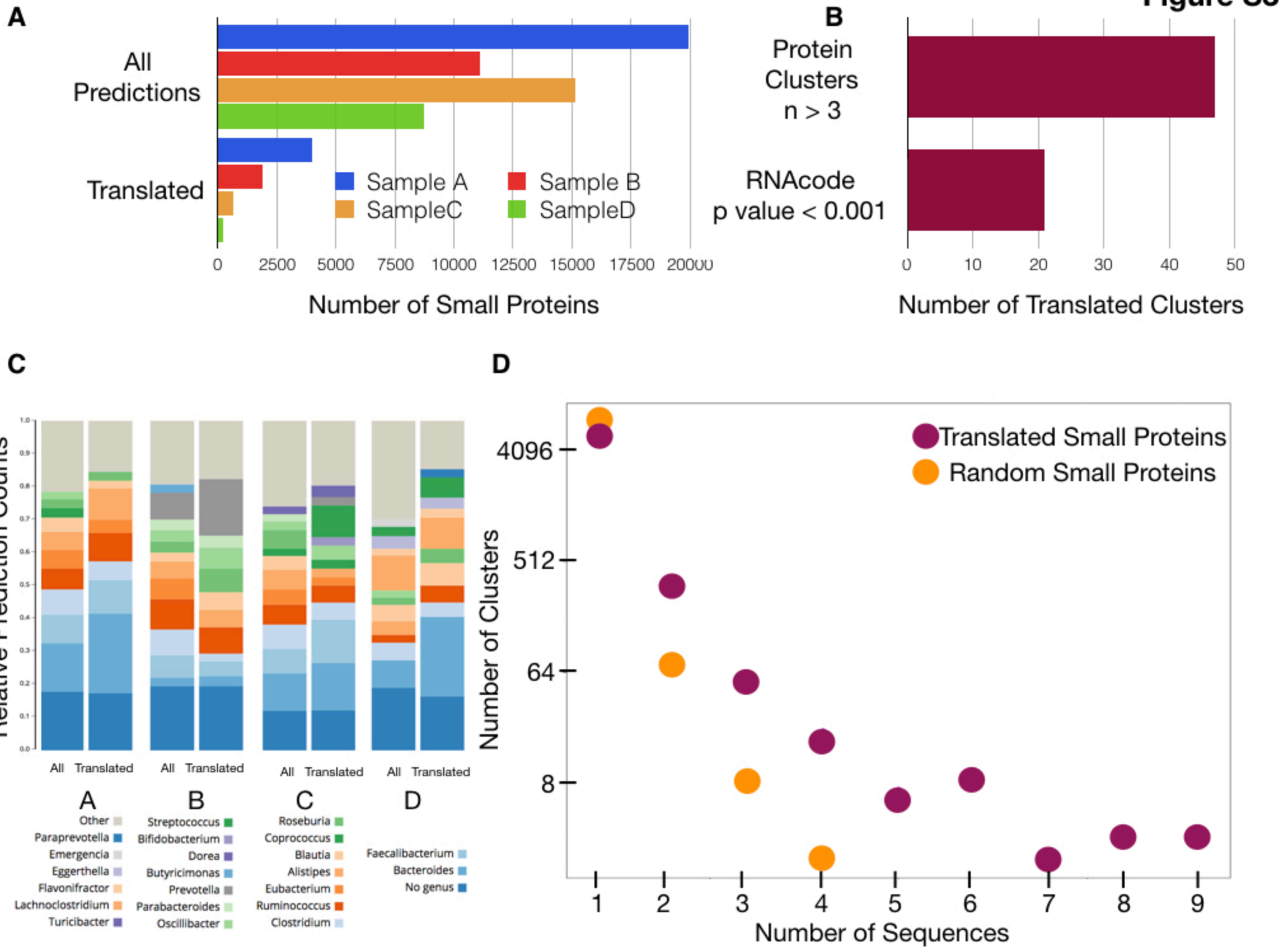

A
All Predictions
Translated
Sample A
Sample B
SampleC
SampleD
0
2500
5000
7500
10000
12500
15000
17500
20000
Number of Small Proteins
B
Protein Clusters n > 3
RNAcode p value < 0.001
0
10
20
30
40
50
Number of Translated Clusters
C
Relative Prediction Counts
All
Translated
A
Other
Paraprevotella
Emergencia
Eggerthella
Flavonifractor
Lachnoclostridium
Turicibacter
B
Streptococcus
Bifidobacterium
Dorea
Butyricimonas
Prevotella
Parabacteroides
Oscillibacter
C
Roseburia
Coprococcus
Blautia
Alistipes
Eubacterium
Ruminococcus
Clostridium
D
Faecalibacterium
Bacteroides
No genus
D
Translated Small Proteins
Random Small Proteins
Number of Clusters
4096
512
64
8
1
2
3
4
5
6
7
8
9
Number of Sequences

Figure S4

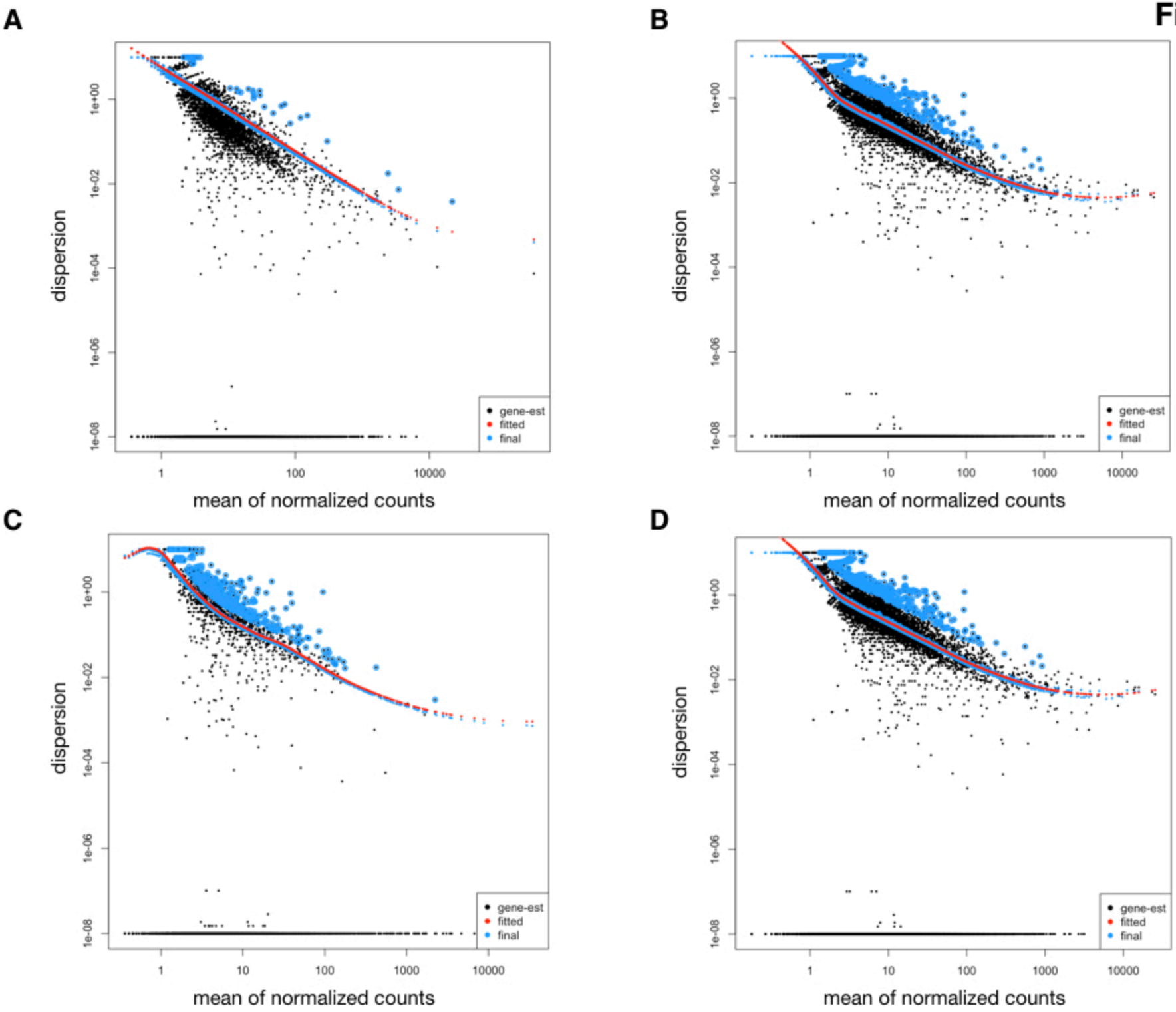

A
B
C
D
dispersion
mean of normalized counts
gene-est
fitted
final
1e+00
1e-02
1e-04
1e-06
1e-08
1
10
100
1000
10000

Figure S5

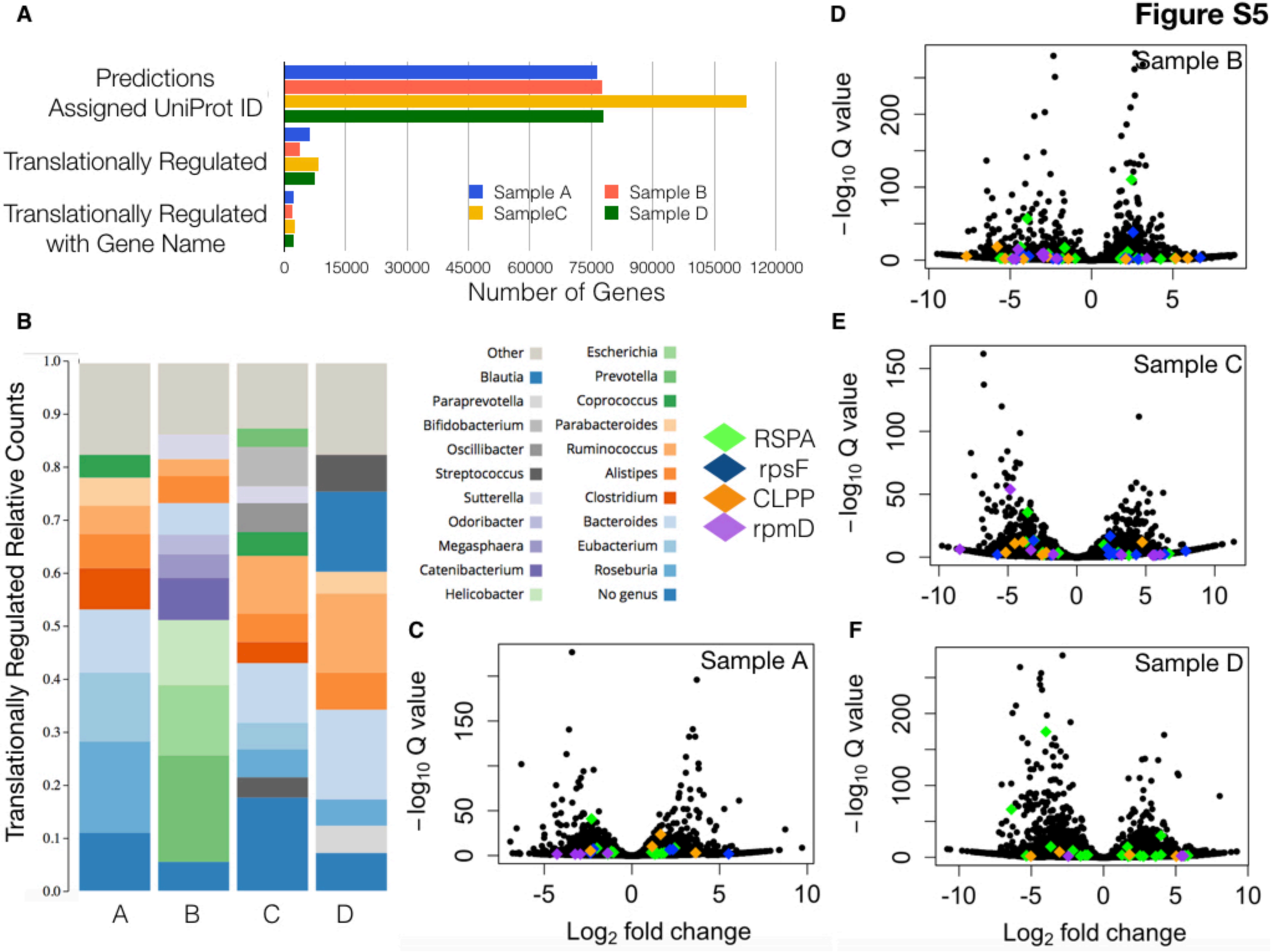
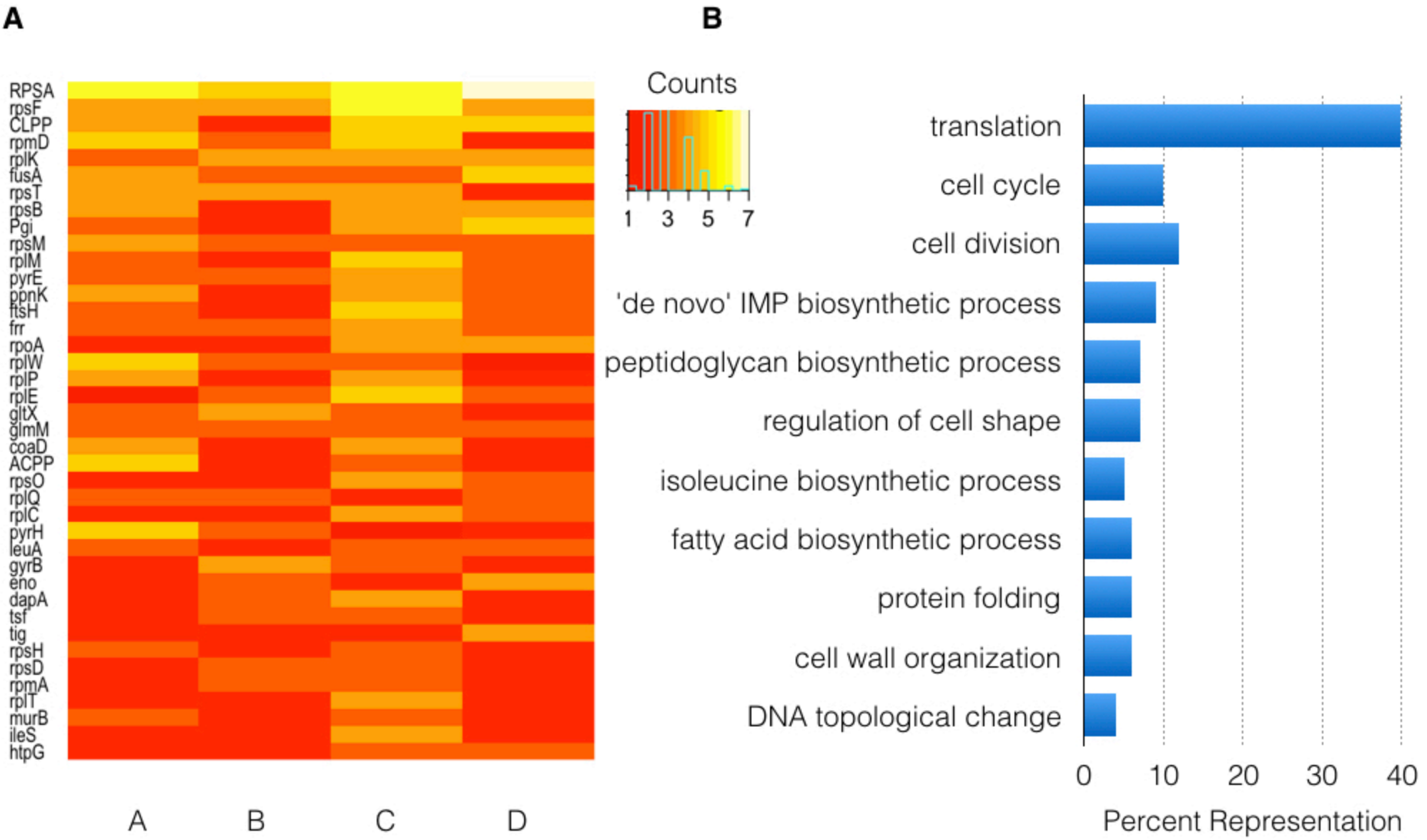

A
Predictions
Assigned UniProt ID
Translationally Regulated
Translationally Regulated
with Gene Name
Sample A
SampleC
Sample B
Sample D
0
15000
30000
45000
60000
75000
90000
105000
120000
Number of Genes
B
Translationally Regulated Relative Counts
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
A
B
C
D
Other
Blautia
Paraprevotella
Bifidobacterium
Oscillibacter
Streptococcus
Sutterella
Odoribacter
Megasphaera
Catenibacterium
Helicobacter
Escherichia
Prevotella
Coprococcus
Parabacteroides
Ruminococcus
Alistipes
Clostridium
Bacteroides
Eubacterium
Roseburia
No genus
RSPA
rpsF
CLPP
rpmD
C
Sample A
$-\log_{10}$ Q value
0
50
150
-5
0
5
10
Log_2 fold change
D
Sample B
$-\log_{10}$ Q value
0
100
200
-10
-5
0
5
E
Sample C
$-\log_{10}$ Q value
0
50
100
150
-10
-5
0
5
10
F
Sample D
$-\log_{10}$ Q value
0
100
200
-10
-5
0
5
10
Log_2 fold change

Figure S6


A
RPSA
rpsF
CLPP
rpmD
rplK
fusA
rpsT
rpsB
Pgi
rpsM
rplM
pyrE
ppnK
ftsH
frr
rpoA
rplW
rplP
rplE
gltX
glmM
coaD
ACPP
rpsO
rplQ
rplC
pyrH
leuA
gyrB
eno
dapA
tsf
tig
rpsH
rpsD
rpmA
rplT
murB
ileS
htpG
A
B
C
D
B
Counts
1
3
5
7
translation
cell cycle
cell division
'de novo' IMP biosynthetic process
peptidoglycan biosynthetic process
regulation of cell shape
isoleucine biosynthetic process
fatty acid biosynthetic process
protein folding
cell wall organization
DNA topological change
0
10
20
30
40
Percent Representation