"Omics" of the mammalian gut – New insights into function

Regina Lamendella[1], Nathan VerBerkmoes[2], and Janet K. Jansson[*1,3,4]

[1] Lawrence Berkley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 92597, USA.

[2] Oak Ridge National Laboratory, Oak Ridge, 1 Bethel Valley Road Oak Ridge, TN, 37830, USA

[3] Joint Genome Institute, Walnut Creek, CA, 2800 Mitchell Drive Walnut Creek, CA 94598

[4] Joint BioEnergy Institute, 1 Cyclotron Rd. MS: 978-4121 Berkeley, CA 94720

[*] Corresponding Author: Janet K. Jansson

JRJansson@lbl.gov

Phone: 510-486-7487

Fax: 510-486-7152

Email address of other authors: RLamendella@lbl.gov (first author)

verberkmoesn@ornl.gov

Abstract:

To understand the role of gut microbes in host health, it is imperative to probe their genetic potential, expression, and ecological status. The current high-throughput sequencing revolution, in addition to advances in mass spectrometry-based proteomics, have recently enabled deep access to these complex environments, and are revealing great insights into the roles of the gastrointestinal (GI) microbiota in host physiology and health. This review discusses examples of how the integration of cutting-edge 'meta-omics' technologies are providing new knowledge about the relationships between host health status in mammals, with a focus on human hosts, and the microbes inhabiting the GI tract. In addition, we address some promises that these techniques hold for future therapeutics and diagnostic applications.

Introduction:

The mammalian gastrointestinal (GI) tract harbors a complex microbial community and its composition reflects the constant co-evolution of these microorganisms with their host environment. As one of the densest microbial habitats on earth, with over $10^{10\text{-}11}$ cells/gram, the gut microbiota is responsible for several key functions including food digestion, provision of substrates to the gut epithelial cells, and host immune responses [1]. Uncovering the taxonomic composition and functional capacity within the mammalian gut microbiota is thus of great importance towards gaining a better understanding of the roles of specific members of the community in host physiology and health.

Current advances in next generation, high-throughput sequencing technologies are now at our disposal to enable deep sequencing of these complex environments, though great

computational challenges still remain. Simultaneously, there has been a revolution in other technologies, including sequencing of total RNA, total proteins, and total metabolites. Collectively, these "meta-omics" techniques provide a means to explore and gain an understanding of the systems biology of the gut environment at different levels of expression. The first step in the "omics" pipeline is microbial community fingerprinting or "microbiomics", in line with other "omics" terms in use. The next "omics" level is metagenomics, or sequencing of total community DNA, including both phylogenetic and functional genes. Community transcriptomics, or metatranscriptomics, is the next step in the pipeline that relies on sequencing of RNA to assess genes that are expressed in a community. Community proteomics, or "metaproteomics" provides information about synthesized proteins and their abundances. Finally, metabolomics, or "meta-metabolomics" describes the end products, or metabolites present in a sample. Here we will discuss some recent examples of the use of these meta-omics approaches in the mammalian gut system and highlight some promising biotechnological applications.

2nd generation 16S rRNA gene fingerprinting: Microbiomics

Traditionally microorganisms in the gut habitat were studied after isolation. However, through the advent of 16S rRNA gene-based studies, we now know that the majority of the microorganisms inhabiting the gut have not yet been cultivated and little is known about their functional role in the gut. Recently, deep sequencing of 16S rRNA genes has provided a more complete picture of the composition of microbial inhabitants of the gut than was previously possible and we have considerable knowledge about the identities of the most dominant members of the community. For example, in most mammals, members of the Firmicutes and

Bacteroidetes are prevalent in the gut environment. However, the relative distributions of species within these phyla vary considerably on an individual basis. Within an individual the gut microbiota is thought to be relatively invariable over time and provides a characteristic fingerprint for each person, although a recent long term study found that the gut microbiota does exhibit considerable daily variations, with few consistent core species [2**]. The normal microbiota can also be disrupted by external factors including antibiotic usage [3], diet [1,4,5], [6,7] and disease [8-13].

Several large funding initiatives have recently been launched to study microorganisms associated with the human host. One of these, the NIH Human Microbiome Project (http://commonfund.nih.gov/hmp/), aims to establish the inter-individual variability of human-associated microbial communities using 2nd generation sequencing approaches with a focus on different body sites, including the gut. It is now apparent that the gut microbiota is distinct in composition when compared to other body sites, such as the oral cavity and skin [2**]. These studies and others have helped to generate hypotheses and gain a better understanding about how the gut microbial community structure relates to host physiology, genetics, immunology, behavior and succession. For example, sequencing of the gut microbiome has revealed that placental mammals birthed vaginally, obtain their first inoculum from the vagina [14,15*]. After this primary inoculation event, the host is exposed to varying environmental exposures that further influence the community composition in the gut. At the same time, the host contributes towards shaping the gut microbial community [1,16,17] and there is a strong genetic influence on the community composition in the gut. Kinship effects have been shown in mouse models and humans, as related individuals share more similar microbial community structures than unrelated

individuals. However, the extent to which genotype affects the gut microbiota remains unknown.

The environmental conditions that are experienced by the host can result in selective pressures that help shape the evolutionary trajectory of host-associated microbial ecosystems. For example, dietary habits can have an influence on the gut microbiota. Mammalian gastrointestinal physiology and diet are in fact two strong predictors of microbial community composition. When the gut microbial profiles of different mammals were compared, they clustered by diet: herbivore, carnivore, omnivores, and sub-clustered by host physiology (hindgut versus foregut fermenters) [1]. Many recent studies have highlighted how diet influences differences in microbial communities that co-diversified with human populations [18-20]. A comparison of the gut microbial communities of Asian, African, and Western humans, revealed that Bacteroidetes were much more abundant in the gut microbiomes of African children (reference). The Bacteroidetes phylotypes isolated from African children were very different from those found in the other countries, comprising bacteria from the genus *Prevotella* and *Xylanibacter*, known to contain genes for cellulose and xylan hydrolysis, and were thus better adapted at extracting energy from plant-rich diets [18]. The effects of diet and host physiology on the gut microbiome span across host lineages, suggesting these are both important factors in shaping the gut microbial community.

While 16S rRNA approaches are extremely valuable for gaining information about microbial community composition and diversity in the GI tract, there are many limitations associated with this approach. For example, nucleotide composition bias within the 16S rRNA gene can lead to incorrect phylogenetic characterizations. Additionally, the 16S rRNA gene represents only a small region of a microbial genome and inferring phylogenies from one gene

alone confers certain risks of misclassification. Thus, confirmation of the identities of microorganisms within complex communities by sequencing of 16S rRNA genes should be corroborated by using several other phylogenetically conserved markers, or by metagenome sequencing as described below. Second generation sequencing technologies have also revealed how targeting different variable regions of the 16S rRNA gene can result in taxonomic biases. For example, the V6-V9 variable region of the 16S rRNA gene is apparently an outlier, yielding the lowest proportion of calls at the genus level [21]. While a majority of the studies have relied on the Roche 454 pyrosequencing platform, other sequencing platforms, such as the Illumina platform, are gaining popularity. For example, the Illumina technology is now generating reproducible taxonomic recovery and diversity patterns at a depth of millions of reads per sample [2**]. Efforts such as the Earth Microbiome Project (www.earthmicrobiome.org) are demonstrating the potential of conducting large-scale studies, in which thousands of samples can be analyzed simultaneously, to survey microbial communities at unprecedented resolutions [22]. The Earth Microbiome Project aims to include thousands of samples from mammalian hosts that will be compared to other microbial communities from diverse habitats on Earth, ranging from marine to soil, thus enabling elucidation of patterns leading to host-associated life styles.

Lessons learned from metagenomic surveys of mammalian gastrointestinal environments.

The microbiomics studies discussed above provide information about species identities, but not about the functional potential of members of the gut microbial community. By contrast, metagenomic sequencing of total community DNA provides information about both the phylogenetic representation, as well as functional genes [4,5,23-25**]. By comparison of several existing metagenome databases, we can clearly see that the mammalian gut metagenomes

are different from those inhabiting other environments (Figure 1). The gut represents a unique functional environment endowed with a cache of genes encoding processes required for food digestion and host immunity. Consistent host-mediated selective pressures compounded with, a microbially dense, high nutrient environment has most likely created a hotspot for microbial evolution and adaptation in the mammalian gut that is distinct from free-living environments.

Metagenomic studies of mammalian gastrointestinal environments have largely focused on gene-centric approaches to unveil functionality encoded by the microbiome, as a result of its vast phylogenetic and functional repertoires. Comparative metagenomics studies have provided insights into inter-individual host differences, as well as differences between different mammalian gut physiologies. Kurokowa et al. (2008) obtained metagenomes from fecal samples collected from 13 healthy Japanese individuals of different ages and found that adult and infant gut microbiomes have different enriched gene families, encoding core functions that generally do not overlap [24]. In addition, conjugative transposons were found to be enriched in the gut, which has highlighted the important role of horizontal gene transfer in shaping gut microbial communities [3,26]. Metagenomics has also enabled study of the existence of functions potentially carried out by low-abundance species. For example, Arumugam et al. (2011) found that low abundance *Escherichia* species had high representation of genes encoding proteins involved in bacterial pilus assembly in human gut metagenomes [27**]. Currently, there is still a need to further increase the number of reference genomes in databases to facilitate taxonomic assignments from complex metagenomic datasets and to enable the detection of more low-abundant species [27**].

Recently, a large consortium of European and Chinese scientists, "MetaHIT", used the Illumina sequencing platform to define a catalogue of 3.3 million, non-redundant gut microbial genes, representing 100 times more genes than are encoded by the human host genome [28**]. This study also demonstrated how metagenomic approaches could enable the study of associations between bacterial genes and phenotypes. New massive metagenomic sequencing efforts are at such a high depth of coverage, that it is possible to assemble whole genomes from a soup of millions and even billions of sequences. For example, in a deep metagenomic survey of the bovine rumen, a total of 268 gigabases of DNA was sequenced and this was sufficient to enable assembly of 15 previously uncultured isolates with novel lignocellulose degradation genes [25**].

In another recent report from the MetaHIT consortium, metagenomic mining of the human gut revealed three distinctly robust "enterotypes" that spanned across geography, age, gender, and disease [27**] (Figure 2). The three enterotypes were strongly driven by species composition (*Bacteroides*, *Prevotella*, and *Blautia*/unclassified *Lachnospiraceae,* respectively). The finding that the human gut microbiota can be classified into different enterotypes could have important ramifications for predicting individual responses to therapeutics.

While next-generation sequencing technologies are generating gigabase and approaching terabase scale metagenome datasets, new challenges have arisen. For example, it remains challenging to tie novel functional capacities to specific microorganisms. This challenge should be facilitated by continued generation of whole genomes from metagenomic data, in addition to single cell genome sequencing, and sequencing of more cultured isolates. Another challenge is the increased amount of data that is represented by increasingly shorter read lengths, often with

questionable sequence quality. This is demanding increasingly complex computation, data storage, and handling requirements. Additionally, more sophisticated algorithms are necessary to help improve sequence assemblies in complex microbial communities such as the gut. Specifically this problem must be solved to facilitate screening of functional genes in their genomic context (i.e. operons) and for application of some of the other "omics" technologies, such as metaproteomics that relies on accurately assembled and annotated metagenome databases for peptide predictions as described below.

Other "Omics": Metatranscriptomics, Metaproteomics, and Metametabolomics: Activity Hallmarks

Although metagenomics provides information about functional gene content in the gut, it is not known whether the genes are actively expressed and have any functional role in a given sample. When extracting total DNA, it is not possible to know if it originated from intact, viable cells, or not, and is therefore not an ideal proxy for assessing functions carried out by the community at a specific point in time. In order to obtain information about active functions, it is necessary to look at the expression profile (metatranscriptomics) or the protein products (metaproteomics). These two technologies are still technically demanding and have only recently begun to be applied for the study of microbial communities.

Sequencing of RNA has recently been used to obtain information about gene expression in the gut under different conditions. For example, metatranscriptomics was used to compare cecal contents from piglets fed with sow's milk compared to artificial formula, with the finding that several pathways for amino acid metabolism and oxidative stress varied in expression

profiles between the two groups [29]. In another study a metatranscriptome approach was used to determine expression of genes involved in carbohydrate metabolism in the gut [30], with the finding that different genes were specifically expressed in different individuals at any given time. Metatranscriptomics of a healthy human twin pair revealed that the gut had a high representation of expressed genes for carbohydrate metabolism, energy production and synthesis of cellular components [31]. In addition, small RNAs were expressed in the gut, which have important roles as regulatory elements involved in bacterial physiology and pathogenicity [31].

While metatranscriptomics is probing the expression dynamics associated with complex microbial communities, this new technology must deal with the inherent biases associated with the need for subtraction of ribosomal RNA (rRNA) that is normally the dominant RNA species extracted, usually comprising over 90% of the total RNA. This problem may be overcome in the future by deep sequencing of total RNA, including rRNA, since current depth of coverage would still be sufficient for obtaining considerable mRNA transcripts (i.e. one lane of Illumina HiSeq generates approximately 50 Gb of sequence). If 90% of that is rRNA, that still leaves 5 Gb of other RNA, including billions of transcript reads). In addition, RNA is simply more difficult to prepare and preserve compared to DNA due to its chemical nature.

The next step in the "omics" information pipeline is to examine expressed and translated protein products, using a metaproteomics approach. Use of shotgun metaproteomics for analysis of the gut microbiota was first demonstrated for a pair of healthy twins [32]. Briefly the microbial community was enriched from feces by density centrifugation and the entire metaproteome, without selection, was denatured and digested into peptides with trypsin. The peptides were injected onto a two-dimensional liquid chromatography (LC) system coupled with

a rapid scanning high-resolution mass spectrometer (MS) and the resulting high accuracy peptide masses were searched against a database containing gut metagenomes and bacterial reference genomes. By comparing clusters of orthologous groups (COGs) of the metaproteomes to metagenomes, it was apparent that the metaproteomes had a higher representation of proteins in COGs for translation, carbohydrate metabolism and energy production than was predicted from the metagenomes [32]. Similar to the metatranscriptome study mentioned above [31], these data highlighted key functions that were more active in the human gut that could not be predicted from metagenomes.

Although the use of this shotgun proteomics approach based on high performance mass spectrometry (MS) was a significant step forward, it also demonstrated that large technological advances are still needed. The most pressing current bottleneck for metaproteomics of complex environments, such as the gut, is the incompleteness and poor validation of the predicted protein databases used for peptide matching to the millions of tandem mass spectra (MS/MS) obtained in a typical study. The best case scenario is to have matched metagenome databases from the same samples to increase the number of proteins identified [33,34]. Also, we can now take advantage of the increase in available human gut microbiome isolate genomes being produced through the NIH Human Microbiome Project to augment databases with well annotated genes (http://www.hmpdacc-resources.org/cgi-bin/img_hmp/main.cgi). However, both of these types of databases still contain considerable redundancy of functional genes that can complicate interpretation of resulting metaproteome data [33]. Furthermore protein quantitation can be challenging if a different metagenome is used for each metaproteome sample.

Currently, large-scale metaproteomics studies are still desperately needed to build reference proteome databases of both host and microbial components of the mammalian gut,

including representatives of different ages, sex, ethnicity and geographic locations. Such a database would be extremely valuable for future studies of disease, development of vaccines, and other diet/obesity studies. In addition, there is a need to determine the dynamics of the microbiota and how the host and microbial proteins change over time. It is not known if the gut metaproteome is relatively stable or if it experiences dramatic dynamic fluctuations. Recently a new approach combining a novel method of lavage sampling, high throughput shotgun proteomics and bioinformatics/statistics revealed the biogeographic distribution of several proteins in different mucosal regions of the human colon, thus identifying proteins at the site of host-microbe interactions [35].

As these respective "omics" technologies continue to be developed, usually in separate laboratory studies, the outlook for combining technologies for a systems biology analysis of the mammalian gut is promising. One recent study combined several "omics" techniques such as DNA sequencing, transcriptomics and proteomics in a global effort to characterize the cecum from gnotiobiotic mice with *E. rectale* and/or *Bacteroides thetaiotaomicron* inoculums [36]. The study showed that *B. thetaiotaomicron* adapts to the presence of *E. rectale* by the up-regulation of specific polysaccharide utilization loci, whereas *E. rectale* adapted to the *B. thetaiotaomicron* by decreasing production of its glycan-degrading enzymes. This study is a good example, albeit a simplified model, of using a systems biology approach to better understand niche specialization and functional redundancy of representatives of two common gut bacteria and is currently being expanded to more complex but controlled systems.

The final proof that metabolic processes have occurred in the gut is evident by examination of metabolic products. The field of (meta)metabolomics, also called metabonomics [37] has yielded interesting insights into gut function. A comparison of metabolites in fecal

extracts of humans, mice and rats by NMR spectroscopy revealed that while some metabolites were common to all species, such as short chain fatty acids, each host species had a unique metabolic profile [38]. Recent research has also demonstrated the large impact the gut microbiome has on mammalian blood metabolites, suggesting a major interplay between bacterial and mammalian metabolism [39*].When the metabolic profiles of different portions of mouse intestines were compared, the biochemical composition of each topographical intestinal region possessed a specific metabolic profile [40]. The metabolomics field is currently in need of transformational advances to more easily handle the plethora of small molecule metabolites produced by the microbiota and more complete metabolite databases to enable accurate identification of metabolites.

Omics insights to the role of the gut microbiota in health

Impact of Diet

Recent "omics" studies have emphasized the importance of diet on shaping the structure of the mammalian gut microbiota. A comparative metagenomic survey of 33 mammalian species, illustrated how the gut microbiota adapts to diet, and how these adaptation events span across different mammalian lineages, supporting the theory of convergence of microbial communities within gastrointestinal systems [41*]. Comparative studies of the cow rumen and termite hindgut microbiome, showed fundamental differences in glycoside hydrolase and cellulosome functional genes, most likely due to the different diets of the respective hosts as their resident microbes coevolved to digest different plant cell wall types [5]. In another recent comparative metagenomics study, similar functional gene profiles were found among swine gut, chicken cecum and cow rumen. This finding may seem counterintuitive considering the vast

differences in gut anatomy and physiology of these hosts, but in this case, the similarities were attributed to similar animal husbandry practices [26]. In a metagenomic study of Japanese individuals, a gene encoding a microbial β-porphyranase from a marine microbe, *Zobellia galactanivorans*, was detected in DNA sequences from the gut microbe, *Bacteroides plebeius* [42*]. This finding suggests that these novel carbohydrate active enzymes were acquired by horizontal gene transfer in the gut metagenome of Japanese individuals that include seaweed in their diet. Furthermore, in a controlled feeding study, gut microbial enterotypes were strongly associated with diet, with *Bacteroides* correlating to diets high in protein and animal fat, while diets high in carbohydrates were associated with a prevalence of *Prevotella* [43].

Several microbial profiling studies by Jeff Gordon and colleagues revealed an increase in the ratio of Firmicutes/Bacteroides in obese humans compared to healthy, which coincided with an increased capacity for polysaccharide fermentation [7,23]. In addition, gnotobiotic mice inoculated with a human gut microbial community and fed different a "Western diet" showed an increase in Firmicutes that coincided with increased fat deposition and this trait was transmissible via microbiota transplantation [7]. In another study, specific metabolites, produced by gut microbes, were correlated to either lean or obese animals [40,44]. Mestdagh et al. (2011), recently demonstrated that the gut microbiota modulated the lipid metabolism of brown adipose tissue, suggesting an important role of the microbiome in the fight against obesity [45].

Impact of Disease

Several different studies of inflammatory bowel disease (IBD) have demonstrated a gut microbiota that is disturbed compared to a normal healthy state. This "dysbiosis" of the gut microbiota has been highlighted in studies of individuals with the specific IBD disease, Crohn's

Disease (CD), in particular those individuals with inflammation in the small intestine, or ileum. These individuals have a lower fecal microbial diversity in general, as well as a reduction in several members of the Firmicutes phylum [10,13,46]. One example of a particular microbe that was less abundant in ileal CD was *Faecalibacterium prausnitzii,* a bacterium that is considered to be beneficial due to its action as a producer of butyrate, a substrate for gut epithelial cells (Figure 3). In addition, to shifts in the microbial community, there were thousands of metabolites that differed in amounts in fecal water extracts of healthy individuals and those with Crohn's disease when using a metabolomics screening approach [12]. Some other diseases, such as necrotizing enterocolitis in infants, have also been correlated to dramatic perturbations in the fecal microbiota; in this case comprising only members of the Proteobacteria and Firmicutes [47]. Further research is necessary to untangle whether the microbiota has a causative role in disease, or whether dysbiosis is a by-product of the disease

Promise of "Omics" for Diagnostic and Therapeutic Biotechnological applications

These advances in "omics" approaches are promising for several biotechnological applications, including biomarker discovery for novel diagnostics and new drug targets, as well as therapeutic strategies. The importance of the gut microbiota for therapeutic applications was dramatically revealed in a study using bacteriotherapy by way of fecal transplantation, to successfully treat *Clostridium difficile*-associated disease and reestablish the normal colonic microbiota [11**]. Additionally, use of beneficial bacteria as probiotics is a promising therapy to improve symptoms of inflammatory bowel diseases such as ulcerative colitis and Crohn's disease [48,49]. Exploiting the gut metagenome for the development of novel therapeutics, including

antimicrobial peptides, immunoregulatory molecules and stress tolerance genes is another emerging area of research with hopes to increase the technological and physiological efficacy of probiotic bacterial strains. Additionally, advances in genomics, transcriptomics, and metabolomics are facilitating a deep understanding of the modes of action of specific probiotics.

"Omics" approaches have now advanced the study of the human-microbiome interface, allowing for more in-depth understanding of disease processes involving the gut mucosal interface [35,50]. For example, recent studies have focused on the ability of the gut microbiota to communicate with the brain and thus modulate behavior [51]. In addition, there is evidence that low-grade inflammation might be directly linked to the gut microbiota through metabolic endotoxemia, and is involved in the control of the gut barrier [52]. Recent research is even suggesting that bacteria from the oral cavity and the gut, may be involved with other diseases, such as atherosclerosis [53].

Biomarkers

Specific microbes, expressed genes, proteins, and metabolites hold promise as potential disease biomarkers. For example, the finding that certain species, such as *Faecalibacterium prausnitzii,* are significantly reduced in patients with ileal Crohn's disease (ICD), suggests that low amounts of this bacterium in feces is a biomarker of ICD [10,13]. Additionally, several gut metabolites are potential biomarkers of the ICD phenotype [12]. Metabolic reactions initiated by gut microbes can lead to signature metabolites in other body fluids, including blood and urine. For example, some metabolites identified in serum are potential biomarkers of the onset of obesity [54]. Future needs for biomarker research include large population studies to validate their correlation to specific disease phenotypes.

Recent high-throughput sequencing approaches are promising to revolutionize personalized medicine because of the potential for rapid screening of specific genes as biomarkers. Although much emphasis has been made on disease biomarkers in the human genome, there remains much to be discovered about disease biomarkers in the human second genome, or microbiome. Application of 2nd generation sequencing technologies should lead towards identification of new gene targets as well as facilitate future reverse engineering of the disease networks that underlie individual patient phenotypes. To achieve that goal, more emphasis needs to be placed on developing robust strategies to model and deconvolute these complex datasets. Ultimately, these advances can lead towards more efficient therapies to restore and promote human health.

Conclusions:

Use of "omics" approaches has recently enabled the charicterization of the phylogenetic and functional repertoire of microbial communities inhabiting the mammalian gut. Inevitably deeper mining of meta-omics datasets will reveal novel aspects of microbial symbioses with mammalian hosts. Although we now have a general consensus about the normal microbial community composition in the mammalian gut, the impact of environmental factors, such as diet, and disease on the gut community dynamics over time are still being explored. Meta-omics approaches are also starting to reveal functional markers that will be important for future diagnostic and therapeutic tools for several diseases.

Acknowledgements

This study was funded in part by a grant from the NIH Human Microbiome Program (UH2DK083991) and E.O. Lawrence Fellowship to R.L. (USDOE301931). This work was also partially supported by the U.S. Department of Energy and LBNL under Contract No. DE-AC02-05CH11231.

Figure Captions:

Figure 1. Non-metric multidimensional scaling of host and non-host associated metagenomes. A matrix of relative abundance of COG functions per metagenome was generated using the 'compare general functions' tool within the Integrated Microbial Genome's (IMG) online software tool. [57]. COG abundances from the five mouse metagenomes were combined and normalized, as these metagenomes had very low sequencing coverage. Non-metric multidimensional scaling was performed in PCORD v5 sofware (MJM softwares), using the Bray-Curtis distance measure. Squares denote the 28 different microbial metagenomes used in the analysis. The following metagenomes were used from IMG [Taxon Object ID]. Endophytic microbiome from Rice [2010549000], Fossil microbial community from Whale Fall at Santa Cruz Basin of the Pacific Ocean Sample #1 [2001200002], Fossil microbial community from Whale Fall at Santa Cruz Basin of the Pacific Ocean Sample #2 [2001200003], Fossil microbial community from Whale Fall at Santa Cruz Basin of the Pacific Ocean Sample #3 [2001200004],Guerrero Negro salt ponds hypersaline mat 01(G) [2004247000], Guerrero Negro salt ponds hypersaline mat 02(H) [2004247001], Guerrero Negro salt ponds hypersaline mat 03(I) [2004247002], Human Gut Community Subject 7 [72004002000], Human Gut Community Subject 8?[82004002001], Mountain Pine Beetle microbial communities from Grand Prairie, Alberta, sample from Hybrid pine (MPB hybrid beetle) [2032320009] Mountain Pine Beetle

microbial communities from McBride, British Columbia, Canada, sample from Lodgepole pine (Lodgepole pine) [2035918003], Mouse Gut Community lean1 [2004230001], Mouse Gut Community lean2 [2004230004], Mouse Gut Community lean3 [2004230000], Mouse Gut Community ob1 [2004230003], Mouse Gut Community ob2 [2004230002], Sample 266 [2018540003], Sample 267 [2018540002], Sludge/Australian, Phrap Assembly [2000000001], Sludge/US, Jazz Assembly [2001000000], Sludge/US, Phrap Assembly [2000000000], Soil microbial communities from sample at FACE Site North Carolina NCD_ElevF (NCD_ElevF) [2124908001], Soil microbial communities from sample at FACE Site 1 Maryland Estuary CO2+ (Maryland Estuary elevated) [2035918006], Soil microbial communities from sample at FACE Site 1 Maryland Estuary CO2- (Maryland Estuary ambient) [2032320004], Termite Protist Endosymbiont Community [2021593003], Xyleborus affinis microbiome from Bern, Switzerland, sample of adult community (Ambrosia beetle adult) [2043231000], Sample 1 [2019105001], sample 2 [2019105002].

Figure 2. Comparison of "-omics" datasets. Exploratory data analysis of two healthy twins using (A) 16S rRNA community structure (B) metaproteomic profiles (C) metabolomic profiles. Each taxa, protein, or metabolite was plotted as its difference in relative abundance. Blue and orange taxa, proteins, or metabolites have higher relative abundance in Patient 6A and Patient 6B, respectively. For visualization purposes, taxa with relative abundances greater that 30%, metabolites greater than 2%, and proteins greater than 0.6% were removed. (D) Network of KEGG pathways plotted as the log fold change difference in expression of pathways from patient 6A to patient 6B. Red=higher pathway expression in patient 6B; Green= higher pathway expression in patient 6A; brown= more equivalent pathway expression in 6A and 6B. Pathway

nodes are sized by the average pathway expression for both patients. Grey nodes denote pathways not captured in the metaproteomes.

Figure 3. Relative abundance of *Fecalibacterium prausnitzii*, in fecal and biopsy samples from healthy (green), ileal Crohn's disease (red), and colonic Crohn's disease (blue) patients.

Figures

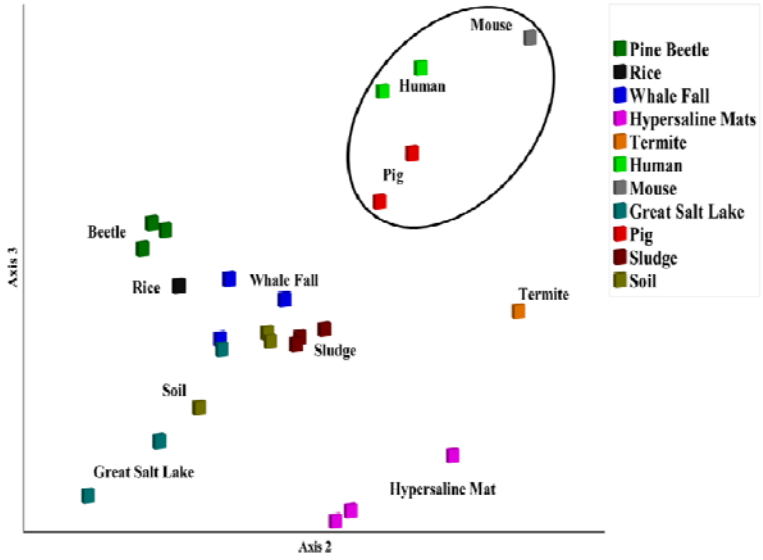

Figure 1.

A

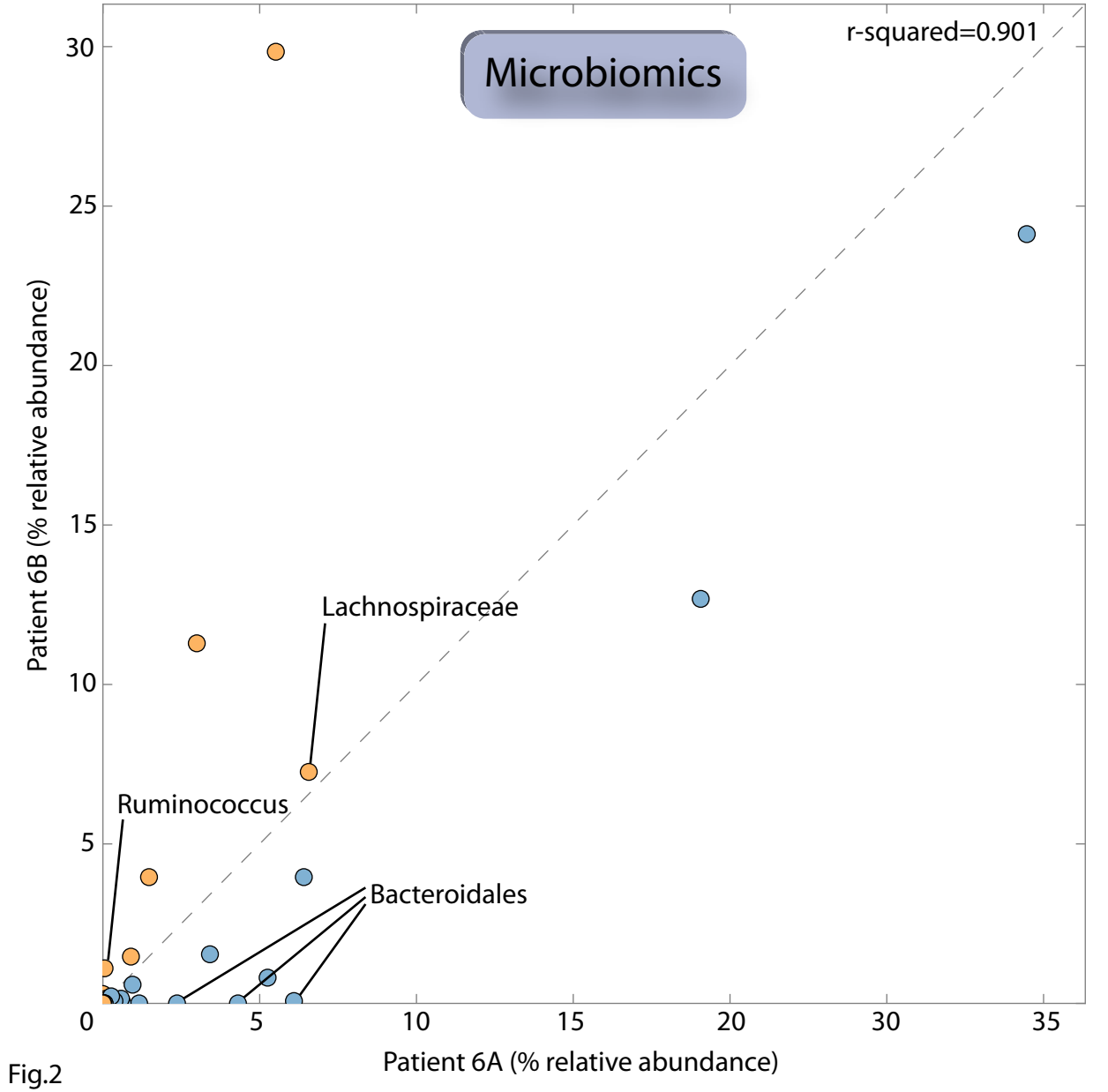
Microbiomics
r-squared=0.901
Patient 6B (% relative abundance)
Patient 6A (% relative abundance)
Lachnospiraceae
Ruminococcus
Bacteroidales
0
5
10
15
20
25
30
35

Fig.2

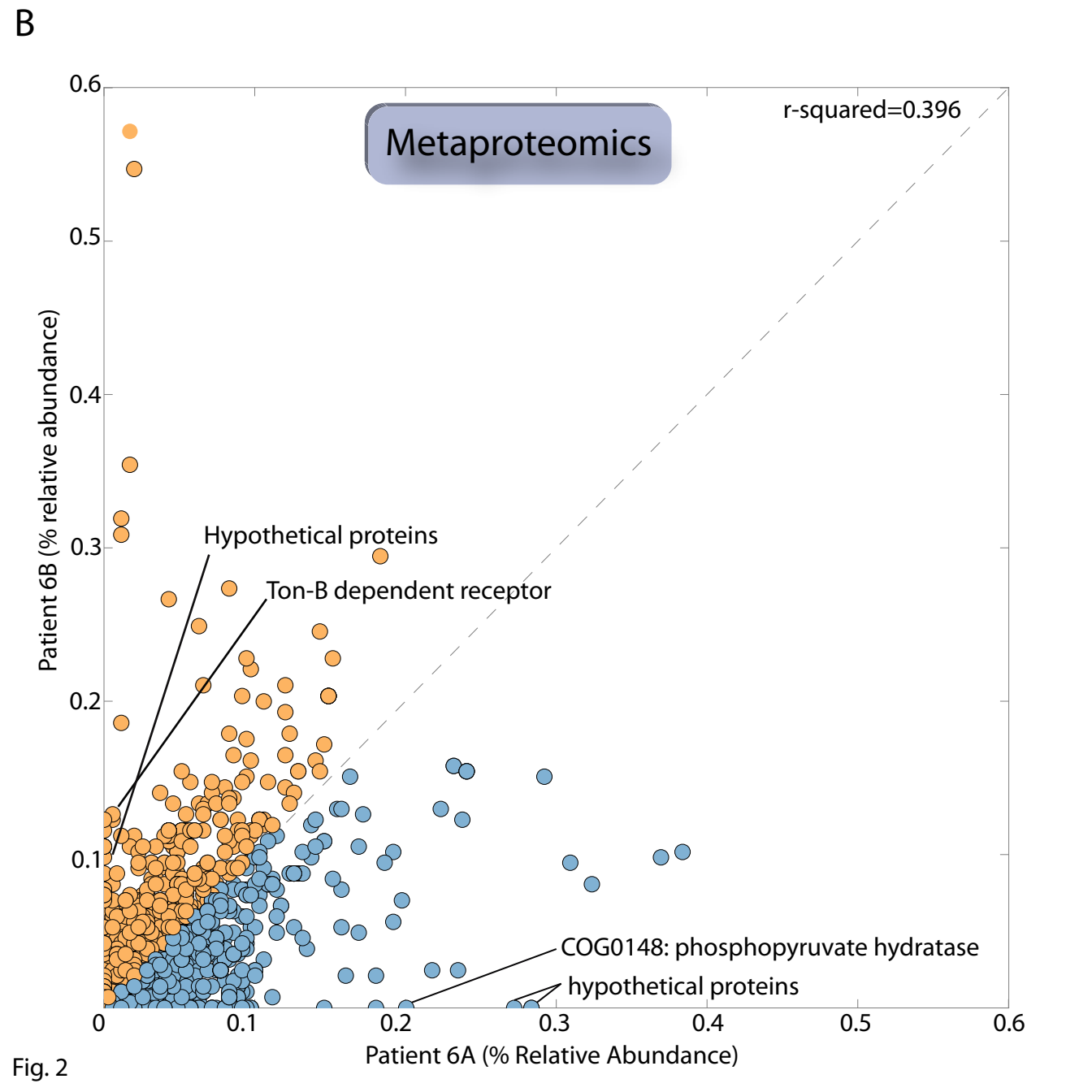
B
Metaproteomics
r-squared=0.396
Patient 6B (% relative abundance)
Patient 6A (% Relative Abundance)
0
0.1
0.2
0.3
0.4
0.5
0.6
Hypothetical proteins
Ton-B dependent receptor
COG0148: phosphopyruvate hydratase
hypothetical proteins

Fig. 2

C

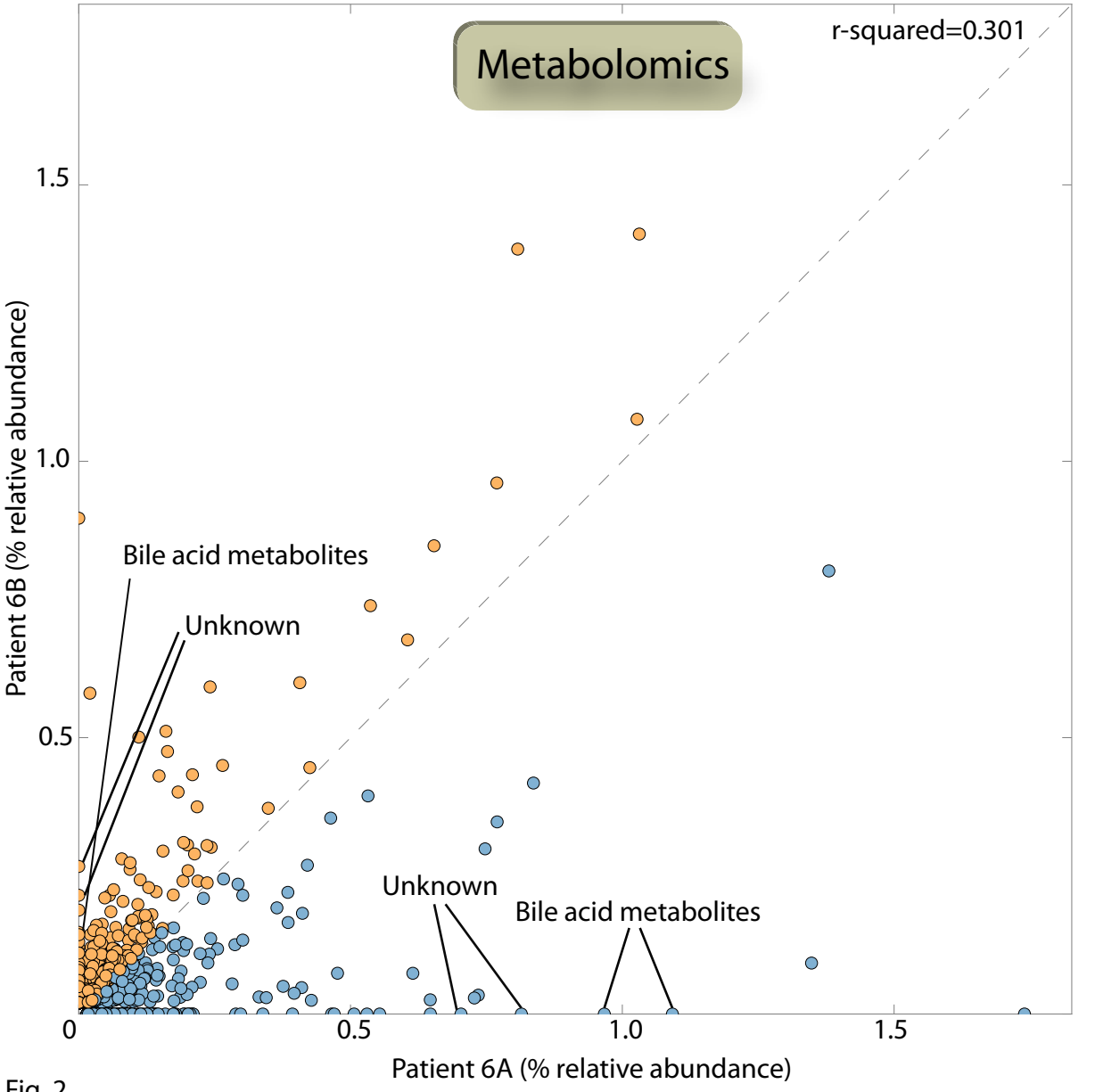
Metabolomics
r-squared=0.301
Patient 6B (% relative abundance)
Patient 6A (% relative abundance)
0
0.5
1.0
1.5
Bile acid metabolites
Unknown
Unknown
Bile acid metabolites

Fig. 2

D

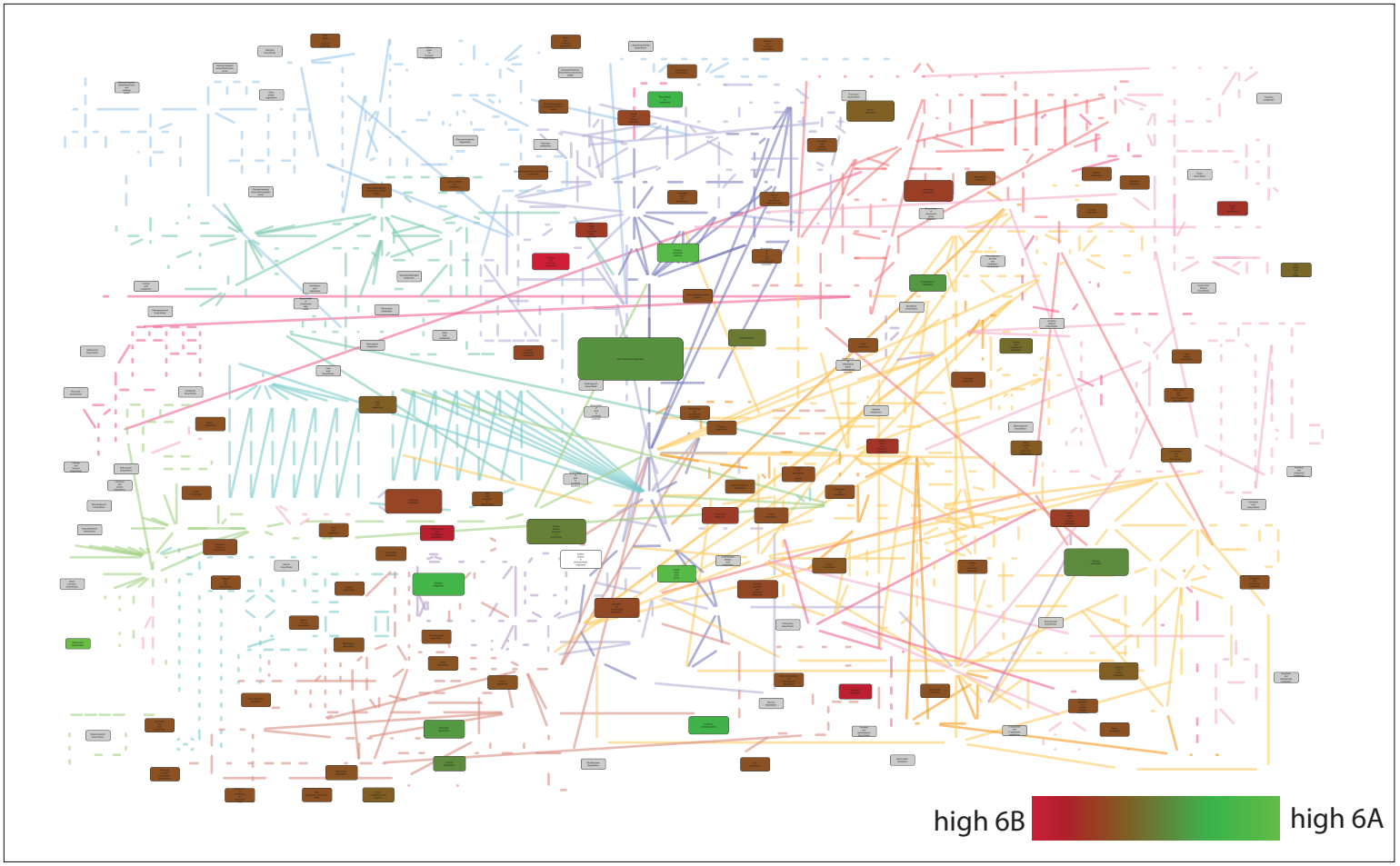

Fig. 2.

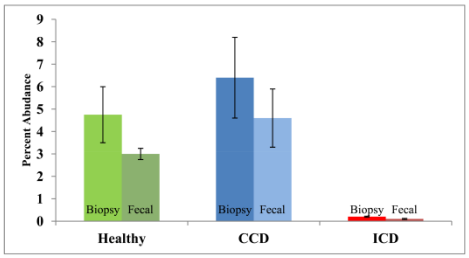

Figure 3.

References

1. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI: **Evolution of mammals and their gut microbes.** *Science* (2008) **320**(5883):1647-1651.

2. **Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R: **Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.** *P Natl Acad Sci USA* (2011) **108**(4516-4522.

This is the first study that demonstrates the utility of high-throughput Illumina sequencing at a depth averaging millions of 16S rRNA gene sequences per sample. The authors show reproducible taxonomies and diversity patterns of environmental and host-associated communities. This study opens up the possibility for conducting large-scale, deep sequencing studies of hundreds to thousands of samples, providing the ability to survey microbial communities at high spatial and temporal resolutions.

3. Jernberg C, Lofmark S, Edlund C, Jansson JK: **Long-term impacts of antibiotic exposure on the human intestinal microbiota.** *Microbiol-Sgm* (2010) **156**(3216-3223.

4. Turnbaugh PJ, Baeckhed F, Fulton L, Gordon JI: **Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome.** *Cell Host Microbe* (2008) **3**(4):213-223.

5. Brulc JM, Antonopoulos DA, Miller MEB, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM *et al*: **Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases.** *P Natl Acad Sci USA* (2009) **106**(6):1948-1953.

6. Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, Cantarel BL, Coutinho PM, Henrissat B, Leclerc M, Dore J *et al*: **Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes.** *Genome Res* (2010) **20**(11):1605-1612.

7. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest.** *Nature* (2006) **444**(7122):1027-1031.

8. Detlefsen B, Boemers TM, Schimke C: **Necrotizing enterocolitis in premature twins with twin-to-twin transfusion syndrome.** *Eur J Pediatr Surg* (2008) **18**(1):50-52.

9. *Williams BL, Hornig M, Buie T, Bauman ML, Paik MC, Wick I, Bennett A, Jabado O, Hirschberg DL, Lipkin WI: **Impaired Carbohydrate Digestion and Transport and Mucosal Dysbiosis in the Intestines of Children with Autism and Gastrointestinal Disturbances.** *Plos One* (2011) **6**(9).

This study compares the mucosal microbiome in intestinal biopsies from children with autism and gastrointestinal disease and those of children with only gastrointestinal disease. Metagenomic analysis of og the microbiota revealed decreases in Bacteroidetes and increases in Betaproteobactia and these were correlated with certain ileal transcripts, indicating a relationship between human intestinal gene expression and microbial community structure in children with gastrointestinal problems and autism.

10. Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, Jarnerot G, Tysk C, Jansson JK, Engstrand L: **A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes.** *Gastroenterology* (2010) **139**(6):1844-U1105.

11. **Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ: **Changes in the Composition of the Human Fecal Microbiome After Bacteriotherapy for Recurrent Clostridium difficile-associated Diarrhea.** *J Clin Gastroenterol* (2010) **44**(5):354-360.

This was the first report of "bacteriotherapy" by way of fecal transplantation to treat *Clostridium difficile*-associated disease (CDAD). Using microbiomics the authors assessed the efficacy of bacteriotherapy in a patient suffering from recurrent CDAD before and after treatment by fecal transplantation from a healthy donor. This study showed how bacteriotherapy resulted in restoration of both the structure and function of the gut microbial communities and was correlated to restoration of health of the recipient.

12. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, Tysk C, Schmitt-Kopplin P: **Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease.** *Plos One* (2009) **4**(7).

13. Willing B, Halfvarson J, Dicksved J, Rosenquist M, Jarnerot G, Engstrand L, Tysk C, Jansson JK: **Twin Studies Reveal Specific Imbalances in the Mucosa-associated Microbiota of Patients with Ileal Crohn's Disease.** *Inflamm Bowel Dis* (2009) **15**(5):653-660.

14. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE: **Succession of microbial consortia in the developing infant gut microbiome.** *P Natl Acad Sci USA* (2011) **108**(4578-4585.

15. *Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R: **Bacterial Community Variation in Human Body Habitats Across Space and Time.** *Science* (2009) **326**(5960):1694-1697.

This study elucidated biogeographical patterns of bacterial communities associtated with 27 different body sites and determined community composition was determined most strongly by body habitat.

16. Ley RE, Peterson DA, Gordon JI: **Ecological and evolutionary forces shaping microbial diversity in the human intestine.** *Cell* (2006) **124**(4):837-848.

17. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI: **Worlds within worlds: evolution of the vertebrate gut microbiota.** *Nat Rev Microbiol* (2008) **6**(10):776-788.

18. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P: **Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.** *P Natl Acad Sci USA* (2010) **107**(33):14691-14696.

19. Thomas F, Hehemann JH, Rebuffet E, Czjzek M, Michel G: **Environmental and gut bacteroidetes: the food connection.** *Front Microbiol* (2011) **2**(93.

20. Wang GZ, Luo HY, Meng K, Wang YR, Huang HQ, Shi PJ, Pan X, Yang PL, Diao QY, Zhang HF, Yao B: **High Genetic Diversity and Different Distributions of Glycosyl Hydrolase Family 10 and 11 Xylanases in the Goat Rumen.** *Plos One* (2011) **6**(2).

21. Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, Hwang J, Chen J, Berkowsky R, Nessel L, Li H *et al*: **Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags.** *BMC Microbiol* (2010) **10**(206.

22. Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J, Ley R, Fierer N, Field D, Kyrpides N, Glockner FO *et al*: **The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6(th) 2010.** *Stand Genomic Sci* (2010) **3**(3):249-253.

23. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M *et al*: **A core gut microbiome in obese and lean twins.** *Nature* (2009) **457**(7228):480-U487.

24. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, Taylor TD *et al*: **Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes.** *DNA Res* (2007) **14**(4):169-181.

25. **Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo SJ, Clark DS, Chen F, Zhang T, Mackie RI *et al*: **Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen.** *Science* (2011) **331**(6016):463-467.

To date this is the largest metagenomic sequencing effort of the cow rumen. In this study the authors sequence 268 Gigabases of metagenomic DNA from the cow rumen. The aithors were able to assemle 15 new microbial genomes from this dataset and found thousands of carbohydrate active genes that were able to degrade recalcitrant celluosic substrates.

26. Lamendella R, Domingo JWS, Ghosh S, Martinson J, Oerther DB: **Comparative fecal metagenomics unveils unique functional capacity of the swine gut.** *Bmc Microbiology* (2011) **11**(

27. **Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M *et al*: **Enterotypes of the human gut microbiome (vol 473, pg 174, 2011).** *Nature* (2011) **474**(7353).

By comparing faecal metagenomes of individuals from different countries the authors discover the identity of three robust clusters or enterotypes within the human population.

The differences in the enterotypes can be explained by species and functional gene composition. The authors also identify marker genes for functional modules that correlated with age, BMI, and gender suggesting the diagnostic potential of microbial markers and the utility oflarge metagenomic surveys.

28. **Qin JJ, Li RQ, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR *et al*: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* (2010) **464**(7285):59-U70.

This was the first seminal metagenomic, characterization of 3.3 million non-redundant microbial genes,from more than 100 individuals from different countries. For the first time, the authors define and describe the minimal gut metagenome and the minimal gut bacterial genome.

29. Poroyko V, White JR, Wang M, Donovan S, Alverdy J, Liu DC, Morowitz MJ: **Gut Microbial Gene Expression in Mother-Fed and Formula-Fed Piglets.** *Plos One* (2010) **5**(8).

30. Booijink CCGM, Boekhorst J, Zoetendal EG, Smidt H, Kleerebezem M, de Vos WM: **Metatranscriptome Analysis of the Human Fecal Microbiota Reveals Subject-Specific Expression Profiles, with Genes Encoding Proteins Involved in Carbohydrate Metabolism Being Dominantly Expressed.** *Appl Environ Microb* (2010) **76**(16):5533-5540.

31. Gosalbes MJ, Durban A, Pignatelli M, Abellan JJ, Jimenez-Hernandez N, Perez-Cobas AE, Latorre A, Moya A: **Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota.** *Plos One* (2011) **6**(3).

32. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK: **Shotgun metaproteomics of the human distal gut microbiota.** *Isme J* (2009) **3**(2):179-189.

33. Cantarel BL, Erickson AR, Verberkmoes NC, Erickson BK, Carey PA, Pan C, Shah M, Mongodin EF, Jansson JK, Fraser-Liggett CM, Hettich RL: **Strategies for metagenomic-guided whole-community proteomics of complex microbial environments.** *Plos One* (2011) **6**(11):e27173.

34. Rooijers K, Kolmeder C, Juste C, Dore J, de Been M, Boeren S, Galan P, Beauvallet C, de Vos WM, Schaap PJ: **An iterative workflow for mining the human intestinal metaproteome.** *Bmc Genomics* (2011) **12.**

35. Li X, LeBlanc J, Truong A, Vuthoori R, Chen SS, Lustgarten JL, Roth B, Allard J, Ippoliti A, Presley LL, Borneman J *et al*: **A metaproteomic approach to study human-microbial ecosystems at the mucosal luminal interface.** *Plos One* (2011) **6**(11):e26542.

36. Mahowald MA, Rey FE, Seedorf H, Turnbaugh PJ, Fulton RS, Wollam A, Shah N, Wang C, Magrini V, Wilson RK, Cantarel BL *et al*: **Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla.** *Proc Natl Acad Sci U S A* (2009) **106**(14):5859-5864.

37. Kinross JM, Darzi AW, Nicholson JK: **Gut microbiome-host interactions in health and disease.** *Genome Med* (2011) **3**(3):14.

38. Saric J, Wang Y, Li J, Coen M, Utzinger J, Marchesi JR, Keiser J, Veselkov K, Lindon JC, Nicholson JK, Holmes E: **Species variation in the fecal metabolome gives insight into differential gastrointestinal function.** *J Proteome Res* (2008) **7**(1):352-360.

39. *Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G: **Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites.** *Proc Natl Acad Sci U S A* (2009) **106**(10):3698-3703.

This study illustrates the large effect of the gut microbiome on mammalian blood metabolites using MS-based metabolomics methods and suggests a significant interaction between bacterial and mammalian metabolism.

40. Martin FP, Wang Y, Yap IK, Sprenger N, Lindon JC, Rezzi S, Kochhar S, Holmes E, Nicholson JK: **Topographical variation in murine intestinal metabolic profiles in relation to microbiome speciation and functional ecological activity.** *J Proteome Res* (2009) **8**(7):3464-3474.

41. *Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI: **Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans.** *Science* (2011) **332**(6032):970-974.

This study demonstrated how targeted sequencing of bacterial 16S ribosomal RNA genes could be used to understand how microbial communities adapt to diet and this adaptation of the microbiota to diet was similar across several different mammalian lineages.

42. *Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G: **Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota.** *Nature* (2010) **464**(7290):908-U123.

This is the first metagenomic study to show how novel carbohydrate active enzymes were acquired in human gut bacteria. The authors characterize the first porphyranases from a

member of the marine Bacteroidetes were transferred to the gut bacterium *Bacteroides plebeius* isolated from Japanese individuals.

43. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R *et al*: **Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes.** *Science* (2011) **333**(6052):105-108.

44. Waldram A, Holmes E, Wang Y, Rantalainen M, Wilson ID, Tuohy KM, McCartney AL, Gibson GR, Nicholson JK: **Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents.** *J Proteome Res* (2009) **8**(5):2361-2375.

45. Mestdagh R, Dumas ME, Rezzi S, Kochhar S, Holmes E, Claus SP, Nicholson JK: **Gut Microbiota Modulate the Metabolism of Brown Adipose Tissue in Mice.** *J Proteome Res* (2011).

46. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Dore J: **Low counts of Faecalibacterium prausnitzii in colitis microbiota.** *Inflamm Bowel Dis* (2009) **15**(8):1183-1189.

47. Wang YW, Hoenig JD, Malin KJ, Qamar S, Petrof EO, Sun J, Antonopoulos DA, Chang EB, Claud EC: **16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis.** *Isme J* (2009) **3**(8):944-954.

48. Cain AM, Karpa KD: **Clinical utility of probiotics in inflammatory bowel disease.** *Altern Ther Health Med* (2011) **17**(1):72-79.

49. Duary RK, Bhausaheb MA, Batish VK, Grover S: **Anti-inflammatory and immunomodulatory efficacy of indigenous probiotic Lactobacillus plantarum Lp91 in colitis mouse model.** *Mol Biol Rep* (2011).

50. Zheng X, Xie G, Zhao A, Zhao L, Yao C, Chiu NH, Zhou Z, Bao Y, Jia W, Nicholson JK: **The footprints of gut microbial-Mammalian co-metabolism.** *J Proteome Res* (2011) **10**(12):5512-5522.

51. Cryan JF, O'Mahony SM: **The microbiome-gut-brain axis: from bowel to behavior.** *Neurogastroenterol Motil* (2011) **23**(3):187-192.

52. Cani PD, Delzenne NM: **Involvement of the gut microbiota in the development of low grade inflammation associated with obesity: focus on this neglected partner.** *Acta Gastroenterol Belg* (2010) **73**(2):267-269.

53. Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE, Backhed F: **Human oral, gut, and plaque microbiota in patients with atherosclerosis.** *Proc Natl Acad Sci U S A* (2011) **108 Suppl 1**(4592-4598.

54. He Q, Ren P, Kong X, Wu Y, Wu G, Li P, Hao F, Tang H, Blachier F, Yin Y: **Comparison of serum metabolite compositions between obese and lean growing pigs using an NMR-based metabonomic approach.** *J Nutr Biochem* (2012) **23**(2):133-139.

DISCLAIMER