Changes in human gut microbiota induced by DOTS treatment in TB patients of Indian cohort

A Major Project dissertation submitted in partial fulfilment of the requirement for the degree of

Master of Technology
In
Bioinformatics

Submitted by

NITIN THUKRAL

(DTU/13/M.TECH/362)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. Yasha Hasija

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE

This is to certify that the M. Tech. dissertation entitled **"Changes in human gut microbiota induced by DOTS treatment in TB patients of Indian cohort"**, submitted by **NITIN THUKRAL (2K13/BIO/12)** in partial fulfilment of the requirement for the award of the degree of Master of Technology, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by him under my guidance.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honouring of any other degree.

Date:

Dr. Yasha Hasija
(Project Mentor)
Department of Biotechnology
Delhi Technological University

Prof D. Kumar
Head of Department
Department of Biotechnology
Delhi Technological University

DECLARATION

I, **Nitin Thukral** hereby declare that the project entitled "**Changes in human gut microbiota induced by DOTS treatment in TB patients of Indian cohort**" is a record of the original work that is done under the guidance of Dr. Y. Singh, Chief Scientist, Allergy & infectious disease (AID), Lab-208, CSIR-Institute of Genomics & Integrative Biology and Dr. Yasha Hasija, Assistant Professor, Department of Biotechnology, Delhi Technological University, Delhi-110042. This report is submitted for the fulfilment of the Major-II. The Introduction, methodology and results that are obtained till now and are embodied in this report have not been submitted to any other University or Institution for the award of any degree or diploma.

Date: Name:

Place: Signature:

ACKNOWLEDGEMENT

I would like to express my profound gratitude and deep regards to various people who have helped and supported me for the building up of this project. I would like to extend my regards and express my greatest appreciation to my guide and project mentor Dr. Yasha Hasija for her belief in my abilities and her exemplary guidance for the project. I am very grateful to her for giving me this opportunity to carry out such a novel and valuable work.

I am highly indebted to Dr. Y. Singh, CSIR-Scientist, IGIB for his faith in my abilities and allowing me to work under his able guidance. This project would not have taken this shape without the guidance, encouragement and valuable suggestions from Abhijit Maji, PhD Scholar, IGIB and Richa Mishra, PhD Scholar, IGIB.

Lastly, I would like to thank faculty and staff of department of Biotechnology, especially Mr. Chhail Bihari Singh and Mr. Jitendra Singh, for their constant support and allowing us to work overtime in the laboratory. Specially thanks to Mr. Rajesh, Mr. Mukesh Singh, Mr. Sandeep and Mr. Anil for additional support in Genome Informatics laboratory.

Nitin Thukral
2K13/BIO/12

Table of Contents

LIST OF FIGURES AND TABLES

Changes in human gut microbiota induced by DOTS treatment in TB patients of Indian cohort

NITIN THUKRAL

Delhi Technological University, Delhi, India

1. Abstract

Objective: Dysbiosis of the gut microbiota can lead to many prolonged maladies such as inflammatory bowel disease, obesity, cancer and autism. It is now also known that antibiotic usage exert strong effects on metabolism of intestinal microbes and thus the human health. But, there is not much systematic characterization of microbiota associated with the Tuberculosis (TB) patients and DOTS. Accordingly, a comparative metagenomics analysis of human gut microbiota of fecal samples taken at several time points from TB patients subjected to DOTS (Direct Observed Treatment, Short-course) was conducted.

Methods: Fecal DNA of 6 different TB patients were collected at three different time points viz. at the time of diagnosis, after 1 week and after 1 month of DOTS and from their healthy family member, was sequenced using Illuminia HiSeq 2000 sequencer at BGI, China. Metagenomics approach was adopted to investigate the taxonomic diversity and functional profile from the data thus obtained, for associated variations in the human gut microbiome of TB patients before, during and after 1 month of DOTS. Further results were compared with the taxonomic diversity and functional profile of healthy human gut microbiota.

Results: We apparently observed oscillatory population dynamics in all the samples. *Prevotella copri* was the most abundant species in all samples except "T" and "U" controls. Also, striking difference in relative abundance of *Prevotella copri* in 'Y' sample was observed where controls displayed its high prevalence in the gut but it remained only a minority in 'YTB'. An inverse relationship was observed between members of phylum *Bacteroidetes* and phylum *Firmicutes*. Many pathways were found to be enriched in TB (0^{th} day) samples. We may conclude that short-term surveillance of TB patients under DOTS pointed towards it minimal effect on the gut microbiota. Also, no fixed trends were observed in case of both taxonomic diversity and functional composition of the gut.

Conclusions: The present study envisioned inter-relationships between taxonomic diversity and functional profile of human gut microbiome and Tuberculosis. The study provides report of human gut microbiota variations to follow-up DOTS. However insights obtained until the midway of this study, suggests that further metagenomic investigations on larger population of TB patients are required to accurately describe the association of the human gut microbiome with TB and DOTS.

2. Introduction

The human body is like a planet inhabited by tiny, not visible to the naked eye, organisms that we refer to as microorganisms. They have now been known to play vital role in virtually all ecosystems ranging from soil, sea and to those in human body environments (Baker *et al.*, 2013; Fang and Evans, 2013; Philippot *et al.*, 2013). The human gut has been the most widely and recently studied ecosystem for deciphering the taxonomy, structure and function of its microbiome. This has been made possible by the use of metagenomic sequencing or culture-independent techniques (De Filippo *et al.*, 2012), which has established the existence of a vast ensemble of such microbes inhabiting the human gut leading to further enrichment of the host`s genetic resource with a new catalogue of genes. Thus, offering complementary metabolic pathways for detoxification, digestion, assimilation of otherwise inaccessible dietary factors, pathogen resistance or protection against pathogen evasion, regulation or development of immune responses and protection against cell injury, production of bioactive compounds and harvesting energy from them. (O'Hara and Shanahan, 2006; Qin *et al.*, 2010; Turnbaugh and Gordon, 2009). Harboring more than 100 trillion microbial cells, the human gut, is one of the leading ecological niche known in the human body.

Recent studies have shown that there exists a unique and relatively stable gut microbiota in each individual, majorly occupied by Firmicutes and *Bacteroidetes*, in addition to minorities like Actinobacteria, Proteobacteria along with some uncharacterized microbes. However, this relatively stable gut microbiota can be easily disrupted by external forces such as diet, migration, nonsteroidal anti-inflammatory drugs, antibiotics, smoking, alcohol consumption and the disease (Jernberg *et al.*, 2007). Perhaps, one of the easiest way by which the human health is affected is through the diet and nutrition which in turn, is significantly influenced by the gut microbiota. Altering the diet from a plant polysaccharide rich, low fat diet to a high fat, high sugar diet (Western diet) led to prominent changes in the gut microbial composition in a single day in human gnotobiotic mice. The members of the phylum Firmicute classes Erysipelotrichi and Bacilli (Enterococcus) were more abundant in comparison to Bacteroidetes, associated with the Western diet (Turnbaugh *et al.*, 2010). The composition of gut microbial communities can be altered by antibiotics therapy as well. Studies have shown that use of antibiotics led to reduction in Bacteroides and Bifidobacterium and enhanced the growth of members of the Firmicutes classes Erysipelotrichi and Bacilli, Campylobacter, genus *Streptococcus* and *Candida Albicans* in the gut (Hill *et al.*, 2010). The liver metabolizing enzymes and the human gut microbiota both have an important role in nutrition, metabolism, immune response and pathogen resistance. Antibiotics administration is believed to disrupt these coevolved interactions, making many individuals susceptible to acute or chronic disease. Several complex human disorders also significantly correlates with the dysbiosis of specific microbial communities (D'Argenio *et al.*, 2013; Gollwitzer and Marsland, 2014; Marchesi *et al.*, 2007). A recent study showed that *Prevotella copri* was carried in their intestinal microbiota by almost 75% of patients with new onset RA. Furthermore, *Prevotella copri* was abundant in the gut of approximately 37.5% of

psoriatic arthritis patients in comparison compared to 21.4% of healthy controls (Scher *et al.*, 2013).

But the important questions still remain, how to recognize which microbes are responsible for the functional contributions; are the most abundant microbes most important; what role does less-abundant microbes have. Therefore, it is important to understand both the composition and the functional of the constituents of the human gut microbiota in diseased state to decipher their role in influencing the diagnosis, disease progression and treatment. It requires fast and accurate bioinformatics tools to analyze the enormous amount of metagenomics data. Evaluating the biodiversity, the structure, the richness of the microbial community and their role in a given environment requires an accurate taxonomic assignment (Ribeca and Valiente, 2011).

Tuberculosis, as per WHO reports, is second to HIV/AIDS as the greatest killer worldwide, caused by *Mycobacterium tuberculosis*. A commonly used treatment regimen for Tuberculosis is a DOTS with isoniazid, rifampicin, pyrazinamide and ethambutol. DOTS is a prolonged treatment in which treatment lasts from six to nine month. In current scenario, no survey has been conducted to study the gastro-intestinal disturbances on the human subjects under the patho-physiological diseased condition caused by *Mycobacterium tuberculosis*. Our study intends to evaluate the taxonomic diversity and functional contributions of the human gut microbiota in the TB patients. In the present study, we applied a combination of molecular biology and computational approaches to assess the impact of DOTS on the gut microbiota of TB patients. By using whole genome shotgun sequencing we aimed to determine the shifts in the microbial community structure in these samples.

3. Review of Literature

Our microbes makes us genetically unique. "The Economist" magazine has cited humans as superorganism. The co-evolution of a vast and diverse microbial ecosystem with mammalian species has profoundly influenced the human health and thus our existence (Bengmark, 1998; Gill *et al.*, 2006). Next generation 'omics' technologies and systems biology have it possible to describe the gut microbiome at transcriptomic, proteomic and metabolic levels, throwing light on the importance of the gut microbiome in human health. Recent studies have shown that gut microbiome not only influences dietary calorific bioavailability (Hooper and Gordon, 2001), drug metabolism (Clayton *et al.*, 2006), immune system maturation, conditioning and response (Macpherson *et al.*, 2000; Mazmanian *et al.*, 2005), post-surgical recovery (Kinross *et al.*, 2011), phenotypic development but also has direct implications in the etiopathogenesis of diverse pathological states such as obesity (Ppatil *et al.*, 2012; Turnbaugh *et al.*, 2006), autism (Finegold, 2008), inflammatory bowel disease (Greenblum *et al.*, 2012; Marchesi *et al.*, 2007), cardiovascular disorders (Holmes *et al.*, 2008) and cancer (Schwabe and Jobin, 2013).

It is also possible to map human gut microbiome variability between species, individuals and populations. The results from multidimensional scaling and principal component analysis (PCoA) of American, Japanese, Danish and Spanish (European) individuals revealed that according to their similarity in composition, all individual samples clustered around three robust points. Clustering was driven independent of age, sex, nationality, and continent or body mass index. "Enterotypes" was the term used to refer to these clusters. *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2), and *Ruminococcus* (enterotype 3) form 3 different enterotypes based in the variations of the above mentioned genera. The existence of Enterotypes suggested that only limited number of harmonious host-microbial symbiotic relationships could be present whereas the discreteness of these balanced states suggested existence of important interactions between gut microbiota and obligate relationships with other species (Arumugam *et al.*, 2011).

The important gut bacteria in the order of numerical importance are Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, followed by Verrucomicrobia and Fusobacteria (Hermann-Bank *et al.*, 2013). The phyla Bacteroidetes, is composed of three classes of gram-negative bacteria and the most studied genus is *Bacteriodes.* Both the phylum Firmicutes and Actinobacteria, includes gram-positive bacteria but with high and low G + C content, respectively. Firmicutes includes the class of Clostridia and the lactic acid producing bacteria whereas *Colinsella* and *Bifidobacterium* spp. are the members of Actinobacteria phylum. Some of them are the first colonizers, right from the birth and some colonize the gut at different age. As mentioned earlier, gut bacteria contributes to the host`s gut defense system and aids in maintaining its normal and balanced functioning but several factors can have variable effect on it composition.

3.1 Gut microbes and gut immunity

The mechanical barrier (particular layer of intestinal epithelial cells, mucus and the enterocytes) and the immune barrier (Immunoglobulins, macrophages, Peyer`s plaques, macrophages, NK cells, neutrophils and mesenteric lymph node) helps the gut in resistance towards pathogens. Gut bacteria maintain resistance against attacking pathogens by competing for nutrients and attachment sites with the pathogenic bacteria, a phenomenon known as colonization resistance (Stecher and Hardt, 2008). Commensal bacteria prevent the invasion of pathogens either by reducing intestinal pH by producing lactate and short-chain fatty acids (SCFAs) (Guarner and Malagelada, 2003) or by producing toxins or carcinogenic metabolites together with volatile fatty acids.

Several factors significantly influence the gut microbiota (Fig 1). The great shift in the structure, representation of the metabolic pathways and gene expression of the gut microbiota was observed when gnotobiotic mice were feed with a high-fat/high sugar diet instead of a usual low fat/plant polysaccharide rich diet. A notable observation was increase in members of Firmicute classes Erysipelotrichi and Bacilli and reduction in Bacteroidetes (known to be associated with the Western diet). Another notable observation in humanized mice fed the Western diet was a significant increase in adiposity as compared to those feeding on the agrarian diet. Thus demonstrating that the gut microbiome could be liable to drastic changes over a very short period of time (Turnbaugh *et al.*, 2010).

Long term and short-term antibiotics therapy have shown to prominently effect the gut microbiota (Jakobsson *et al.*, 2010; Jernberg *et al.*, 2007). Studies have found relative decrease in the members of genera *Bacteroides, Clostridium* and *Bifidobacterium* and promotion in the growth of *Campylobacter, Streptococcus*, *Leuconostoc*, or fungi such as *Candida Albicans* in the gut (Hill *et al.*, 2010).

The bidirectional communication between the gut and the brain is the result of the gut-brain axis where neural, immunological and endocrine mechanisms are used by brain to monitor and modulates gastrointestinal system functions. It is believed that intestinal microbiome influences the growth and functioning of the enteric nervous system. In adults, chronic stress is correlated with the increase of class Bacteroides species and Clostridium species combined with increased levels of IL-6 indicating immune activation (Gilbert *et al.*, 2011). Furthermore, investigations found that gut becomes leaky and level of circulating LPS increases as a result of chronic stress.

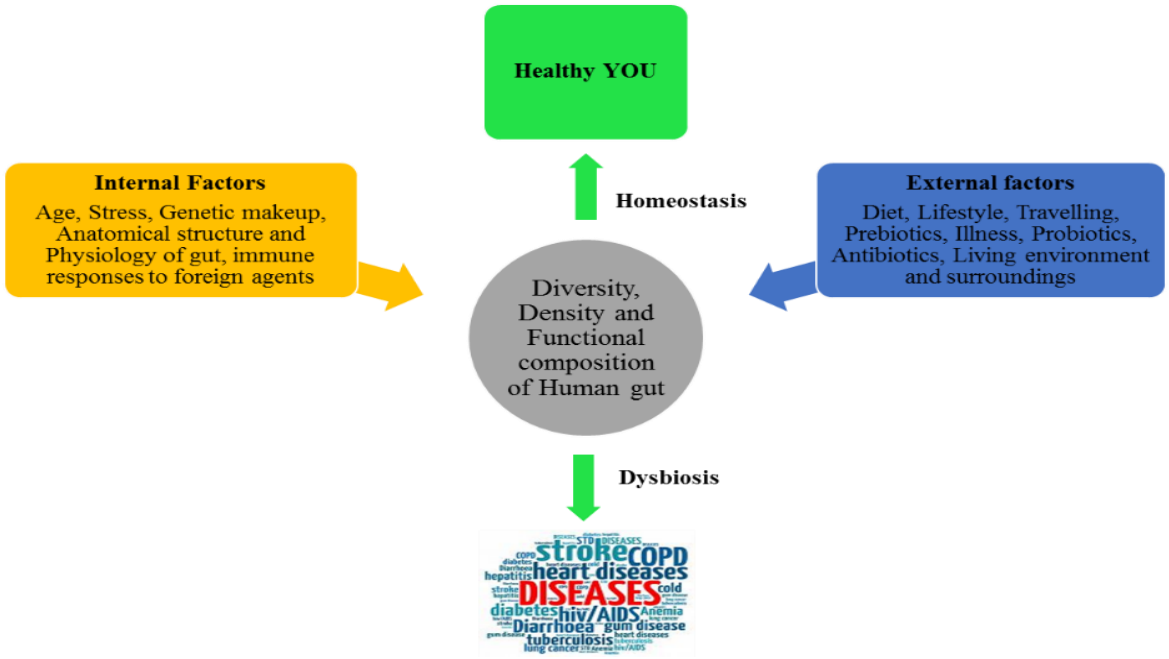


Figure 1: Representing different factors influencing the human gut microbiome

3.2 Dysbiosis

A healthy gut is characterized by the abundance of the gut microbiota. The disruption of the balance between pro-inflammatory and anti-inflammatory microbial species is often associated with low richness of gut microbiota and may trigger host inflammation. The relative abundance of microbial species often uses microbial richness as an accurate biomarker, and further can be used to measure the presence or absence of genes for both known and unknown microbial species in the sample. Interestingly, individuals exhibiting microbial gene counts < 480.000 are characterized by noticeable increase in overall insulin resistance, adiposity, leptin resistance, dyslipidemia, and a more prominent inflammatory phenotype in comparison to high-gene-count individuals. It was reported from a functional perspective that low diversity corresponded to increase in hydrogen sulfide formation and reduction in butyrate-production, reduced production of methane and hydrogen and increased mucus degrading potential. The less healthy state thus can be attributed to gene-poor microbiota (Le Chatelier *et al.*, 2013).

Usually, gut microbiota has a commensal relationship with the host but its imbalance can cause many diseases under abnormal conditions. For instance, pseudomembranous colitis can occur following antibiotics therapy and surgical procedures due to toxin produced by *Clostridium difficile* (Moss and LaMont, 2003). Also, diarrhea was found to be associated with decline in members of genus *Streptococcus*, particularly in *S. alactolyticus* (Hermann-Bank *et al.*, 2013).

Diabetes has been a widespread problem affecting many all around the world. Recently it has been shown to be considerably linked with gut bacteria. The early intestinal microbial colonization at birth impacts the occurrence of diabetes which, in turn is affected by feeding ways, delivery method for birth and Body-Mass Index (BMI) (Burcelin *et al.*, 2011). Type-1 Diabetes (T1D) is an autoimmune disease resulting from the interactions between genetic and environmental factors leading to the destruction of islet beta-cells and thus the ability to produce insulin. The manipulation of gut bacteria controlled mucosal oxidative stress and pro-inflammatory balance leading to restoration mucosal barrier function of intestine, thus providing protection against T1D (Musso *et al.*, 2011). There exists cross-talk between gut microbiota and the innate immunity which may be involved in islets of Langerhans destruction. This means that aberrant gut bacteria could be the potential contributor to pathogenesis of T1D (Hara *et al.*, 2013).

Chan and colleagues used metagenomic approach using DNA from mummified body of Terézia Hausmann to described two *Mycobacterium tuberculosis* genotypes infecting the mummy. This would not have been possible with the use of routine culture techniques. This study shows the power of metagenomics in studying historical specimens and also points to the amount of information we miss while diagnosis individuals using culture-based techniques (Chan *et al.*, 2013). Patil et al. (2012) reported a comparative analysis and quantification of human gut microbiota of Indian cohort (lean, normal, obese and surgically treated-obese individuals) using 16s rRNA sequencing of fecal matter. Though they observed no trend in the distribution of bacteria among different individuals but short chain fatty acids (SCFAs) content was found to be elevated in obese individuals. The treated-obese individuals show reduced levels SCFAs along with reduced *Bacteroides* and archeal counts (Ppatil *et al.*, 2012).

The imbalance in the native gut microbiota contributes significantly to the expansion of different types of tumors such as prostrate, colon and gastroinstestinal cancer. There is a visible difference between gut microbial composition among colon cancer patients and healthy individuals. Two of the *Prevotella* species were completely absent, several bacterial genera producing butyrate were reduced, in case of colon cancer patients. On the other hand, in colorectal cancer fecal samples, *Citrobacter farmer* and *Akkermansia muciniphila* (mucin-degrading species) were over-represented bacteria species (Weir *et al.*, 2013). It is also believed that the gut microbiome may be modulated for the benefit of the host organism. A significant suppression in the incidence of colon cancer, tumor volume and tumor multiplicity was observed upon dietary administration of *Bifidobacterium longum* (Singh *et al.*, 1997). In addition, ingestion of *B. longum* led to considerably inhibition in Azoxymethane-induced cell proliferation, expression of ras-p21 oncoprotein activity and ornithine decarboxylase activity.

It has also been hypothesized that microbial alterations in the human gut has a significant connection with HIV infection pathogenesis (Gori *et al.*, 2011). Such a hypothesis is evident

from the fact that certain gut bacteria such as *Proteus mirabilis, Citrobacter freundii, E. coli,* some *Staphylococcus species* and *Enterobacter aerogenes* found in the HIV-patient`s gut held 90% sequence similarity with HIV-1 (Zajac *et al.,* 2007). HIV infection cause disruption of gut homeostasis, leading to increased concentration of bacterial compounds such as Lipopolysaccharide, peptidoglycan and bacterial DNA, in the circulation further stimulating the immune activation, thus contributing to viral replication and causing further disease progression (Gori *et al.,* 2011). It is believed that disproportionation of the intestinal immune barrier, translocation of immunostimulatory microbial products, and chronic form of systemic inflammation, drive HIV infection to AIDS. A study conducted by Vujkovic-Cvijin *et al.* found enrichment of 579 taxa and depletion of 45 taxa in viremic untreated HIV-infected individuals in comparison to healthy individuals. The significantly enriched taxa were *Erysipelotrichaceae* (known to be associated with obesity and increased cardiovascular disorders), members of phylum Proteobacteria, and genera such as *Salmonella, Shigella, Escherichia, Serratia and Klebsiella* species. Additionally, highly opportunistic pathogens and sources of bacteremia were abundant in mucosae of viremic untreated HIV subjects. However, the relative abundances of Clostridia and Bacteroidia exhibited a decline in case of viremic untreated HIV subjects (Vujkovic-Cvijin *et al.*, 2013).

A higher abundance of pathogenic microbes such as *Candida albicans* and *Pseudomonas aeruginosa* and dramatic reduction of *Bifidobacteria* and *Lactobacilli* was observed in patients living with HIV. Probiotics display potential in significantly assisting the defense system of people living with HIV infection, for instance, *Lactobacillus rhamnosus* (Hemsworth *et al.*, 2011). Gori *et al.* significantly reported the inversely proportional relationship between *bifidobacteria* and *Clostridium coccoides, Clostridium lituseburense, Eubacterium rectale* and *Clostridium histolyticum* under probiotic intake condition (Gori *et al.*, 2011).

Gut microbiota is also related to several other maladies and diseases such as rheumatic arthritis (RA) and disorders affecting kidney (Kotanko *et al.,* 2006; Smith and Macfarlane, 1996). RA patients exhibited significantly reduced levels of *Bifidobacteria* and *Bacteroides fragilis* in their gut. Obesity is also thought to be a risk factor for RA. It is postulated that enhanced systemic exposure to lipopolysaccharide in obese mice models could increase the risk for RA on the observation of increased uptake of LPS through gut lumen to other tissues (Yeoh *et al.*, 2013). In a recent study, intestinal microbiota showed expansion of *Prevotella copri* in about 75% of patients with new onset RA (NORA), thus hinting its association with increased susceptibility to RA. Furthermore, *Prevotella copri* was also found in the gut of 37.5% of psoriatic arthritis patients as compared to 21.5% of healthy controls (Scher *et al.*, 2013).

3.3 Functional screening of gut microbiota

Functional analysis relies on sequencing and analyzing all the additional genetic material apart from the host in the sample, including taxonomically unknown microorganisms, and finally predicting their function based on homology or similarity. Recent investigations provide interesting information related to the functions performed by representative microbial communities within the human gut. It is estimated that this extensive non-redundant catalogue of microbial genes, in the intestinal habitat, translates into proteins involved in up to 20,000 biological functions (Qin *et al.,* 2010). Some of them are important protein complexes (RNA polymerase, ATP synthase, secretory proteins) and some are the ones common to free-living bacteria such as central carbon metabolism, synthesis of amino acid, alanine metabolism and nucleotide sugar metabolism. Some other gene clusters encode proteins important for microbial sustenance such as those involved in linkage to the host or in garnering energy from glycolipids secreted by epithelial cells.

Despite numerous studies revealing highly differing taxonomic compositions of human gut microbiota, interestingly functional compositions are rather similar in healthy individuals. (Methé *et al.,* 2012). Such studies points to the fact that there exists functional redundancy across differing taxonomic compositions, and different microbial species may have similar pathways. Thus, implying that functional profiling may eventually become an optimal approach for defining a "healthy" human gut and segregating it from "diseased" human gut ecosystems.

Owing to the decreasing cost and faster processing, DNA sequencing has come a long way from being just limited to 16s rRNA sequencing. As a result, it is now feasible to analyze and characterize whole genome with reasonable coverage. Importantly, whole-genome sequencing and metagenomic analysis of human fecal samples is capable of describing collective and non-redundant genetic composition of the community from where we could infer important functional and metabolic networks. Apart from providing information about non-bacterial members in the gut microbiota, it can describe viruses, yeasts and protists (Li *et al.,* 2014). Though many recent studies have often been able to link the changes in the taxonomic composition of the gut due to pathologies such as inflammatory bowel diseases, obesity, type 2 diabetes, *Clostridium difficile*–associated diarrhea, advanced chronic liver disease, and others (Cho and Blaser, 2012) but consistency remains still poor among such studies. This can be attributed to the lack of standardized methodology. Also, such associations have not been really able to significantly assign causal or consequential role to microbes in the pathogenesis of a disease. This calls for follow-up studies and, particularly, intervention studies aiming at recovering the structure of the gut microbiota to the baseline.

3.4 Characterizing the shotgun sequencing metagenomic data

As described above, the benefit of using metagenomic sequencing is that it can yield in-depth information about diversity and functional potential of the gut microbiota. This functional profile makes it possible to visualize community dynamics and metabolic properties of the gut. The sequences obtained from each sample for the entire community metagenome, marks the beginning of the whole-genome shotgun analysis. These can be analyzed in different ways for taxonomic composition: either by binning or assembling into contigs or partially assembling into ORFeomes or by simply annotating at the read level. Several freely available software packages are present that compare the metagenomic reads to a database of whole genome sequences or to a database of marker genes using BLAST to determine the phylogeny of microbes in the gut microbiome. The most identical match is then likely used to determine phylogeny of the sequence. Once ORFs in the metagenome are defined, they can be either BLASTed to the reference database or analyzed using codon frequencies or Markov models. Then, the frequencies of the gene/gene products can be linked with the pathways. This allows to define the overall metabolic profile of the sample thus, paving way for inferencing diagnostic and potential explanatory functional biomarkers.

4. Material and method

4.1 Human subject and extraction and purification of DNA

The use of human subjects was approved by IGIB, all participants signed informed consent. In order to study whether prolonged antibiotic treatment disrupts the coevolved interactions between host and gut microbiome, we collected the stool samples from 6 patients before and during the tuberculosis treatment. A commonly used treatment regimen for *Mycobacterium tuberculosis* is a DOTS. The samples were taken at three different time points: at the time of diagnosis (0^{th} day (0)) and during treatment (After 1^{st} week (W) and 1^{st} month (M) of antibiotic DOTS) and their individual healthy family members acted as controls.

4.2 DNA extraction, metagenomic whole genome shotgun sequencing and quality filtering

The DNA was isolated using QiaAmp Stool DNA kit using manufacturer`s guidelines. The concentration and purity of DNA was observed using Nano Drop spectrophotometer (NanoDrop ND-2000) and gel electrophoresis, respectively. DNA samples were later pre-processed by using DNA sample preparation kit protocol (Illumina. San Diego, CA, USA). Then, the samples were subjected to HiSeq sequencing as per the manufacturer's (Illumina) protocol. Illumina Hiseq GAII technology was used to generate 2 x 90bp paired end read with average insert size of 350bp. Base calling was performed using Illumina Genome Analyzer Pipeline software version 1.5.1. Individual metagenome reads were trimmed using following parameters: Q_{20} quality cut-off, a minimum read length of 85bp and allowing no ambiguous nucleotides. All the patients included in the present study were diagnosed with the TB-infection for the first time and tested negative for HIV, all the subjects were non-diabetic and non-smokers. The serum levels were also measured for the patient. Sequence reads obtained were of high quality but still their quality control was performed using NGSQC Toolkit v2.3.3 (Patel and Jain, 2012).

4.3 Taxonomic classification

Taxonomic classification was obtained using MetaPhlAn (Segata *et al.*, 2012) with raw reads as input. According to the 'paired end' criterion, relative abundance was calculated.

4.4 Metagenomic assembly and annotation

Quality filtered reads of all the samples (n=24) of gut microbiome were assembled by using Velvet 1.2.10 (Zerbino and Birney, 2008) set at different k-mer length ranging from 41 to 61, insertion length of 350 bp, an expected coverage of 200 and coverage cut-off of 5 for all samples. The assembled sequences have the average assembly size of 80.34Mb.

4.5 Functional Analysis and pathway reconstruction

For the functional assessment, ORFs were predicted for all the assembled metagenome samples i.e. AB, T, U, V, Y, and Z using FragGeneScan (Rho *et al.*, 2010). Two-step approach was followed for pathway reconstruction, first the KEGG orthologs were assigned to the contigs using KEGG automatic annotation server (KAAS) (Moriya *et al.*, 2007) which was followed by final pathway mapping using Minimal set of pathways (MinPath) (Ye and Doak, 2011). MinPath uses parsimony approach to eliminate noticeably spurious pathways (redundancy) annotated using naïve approaches (such as KAAS and RAST). Furthermore, the reconstructed pathways were normalized by calculating relative abundance with respect to total hits obtained for all the pathways deciphered for each sample. Comparative metabolic potential was further studied both intra-sample (i.e. between control, zero, week and month) and also inter-sample (i.e. between AB, Y, Z, T, U, V).

The assessment is still underway for investigating the metagenomics data of distal gut bacterial communities obtained from the fecal samples of TB patients.

5. Result and Analysis

To investigate the influence of TB and DOTS on the human gut microbiota, we analyzed the human gut microbiome using whole genome shotgun sequencing method followed by metagenomic analysis of the fecal DNA. The samples were obtained from TB patients at three time points viz. at the time of diagnosis (0^{th} day), after one week of DOTS (7^{th} day) and then after one month of DOTS (30^{th} day) and their healthy family members represented the controls. The control and patient`s metadata is represented in the Supplementary Table 1 (**Appendix**) and its corresponding summarized information is provided in the Table 1.

	AGE		**GENDER**		**DIET**		**WEIGHT**		**BMI**	
	Range	**Mean**	**Male**	**Female**	**Veg**	**Non-Veg**	**Range**	**Mean**	**Range**	**Mean**
Controls	19-35 years	26.5	6	0	1	5	58-79 kg	70.17	25.4-35.1	31.01
TB patients	14-47 years	30	3	3	1	5	36-60 kg	46.5	18.07-28.06	21.28

Table 1: Representation of the range and mean of age, weight and BMI along with gender and type of diet in the control and patients

The quality of the short reads obtained following whole genome shotgun sequencing was evaluated using NGSQC Toolkit 2.3.3 (Patel and Jain, 2012). The statistics of the quality control analysis is given in the Supplementary Table 2 (**Appendix**). This was followed by estimating taxonomic composition of each sample using MetaPhlAn 1.7.8. MetaPhlAn 1.7.8 (Segata *et al.*, 2012) identifies population of microbes in a microbial community and their relative abundances for large data sets. It uses clade-specific marker genes database for unambiguously and accurately assigning reads to microbial clades and then measuring their proportions in a given sample. The samples were mapped to phylum, classes, order, family, genus and species level.

5.1 Microbial diversity in Controls

The healthy family member of the TB patients, who have not taken antibiotics from past 6 months acted as controls. All of them have high BMI and all were on non-vegetarian diet except sample "T". Upon taxonomic analysis, we found the controls to be rich in microbial diversity with kingdom Bacteria as the most abundant kingdom Archaea as less abundant. At the phylum level, Bacteroidetes which included class members of Bacteroidia were most abundant followed by Firmicutes, which included class members of Clostridia and Negativicutes and Actinobacteria. The phyla such as Spirochaetes (Class Spirochaeta), Synergistetes (Class

Synergistia), Cyanobacteria, Acidobacteria were almost absent in all the controls whereas phyla Proteobacteria, Thermi, Chlymadiae, Chloroflexi, Fusobacteria, Euryarchaeota and Lentisphaerae were among the rare phyla. Prevotellaceae, Veillonellaceae and Eubacteriaceaei represented the most abundant families whereas Actinomycetaceae, Aeromonadaceae, Camplylobacteraceae, Cardiobacteriaceae, Carnobacteriaceae and Fusobacteriaceae were the absent families in the control samples. Genera such as *Prevotella* (specifically *Prevotella copri*), *Eubacterium* (specifically *Eubacterium eligens*, *Eubacterium rectale*), *Faecalibacterium* (specifically *Faecalibacterium prausnitzii*), *Bacteriodes* and *Megamonas* were relatively more abundant in controls. *Prevotella copri* was the most prominently abundant species in all the controls (figure 3).

5.2 Microbial diversity in TB patients

All of them had low BMI as compared to their healthy counterparts and all were on non-vegetarian diet except sample "T". In TB patient samples (0^{th} day), taxonomic analysis found kingdom Bacteria to be most abundant and kingdom Archaea (present only in VZ, YZ, ZZ and was absent in ABZ and TZ) to be in minority only in comparison to its healthy counterparts. Kingdom *Archaea* disappeared during treatment in samples 'Y' and 'U', rest all samples only showed up-down fluctuations in its relative abundance in comparison to healthy controls, who had not consumed antibiotics from past 6 months.

Phyla Bacteroidetes and Firmicutes were most abundant in TB patients. An inverse relationship was observed between phyla Bacteroidetes and Firmicutes i.e. as the abundance of Bacteroidetes decline, the abundance of Firmicutes increases. Also, phylum Proteobacteria showed significant increase in its abundance alongside Firmicutes. The relative abundance of class Actinobacteria was observed to swings on both sides and the final value after 1 month of DOTS was close to that observed in normal healthy controls. Negative correlation between phyla Bacteroidetes and Firmicutes was reciprocated in their respective classes as well (Class Bacteroides and Class Clostridia) (figure 5). Class Negativicutes belonging to phylum Firmicutes, which represents only gram-negative bacteria, was also abundant in some samples.

Genus *Bacteroides* shows variable behavior among different samples. It was observed to be most prominent in sample 'Y' with very high relative abundance. Genus *Escherichia* showed high abundance in TB patients not undergoing any treatment (untreated TB patients) and it also shows a decreasing trend with antibiotics dosage except in one sample "Y", where it goes on increasing. Genus *Eubacterium* and *Faecalibacterium* were found to be positive correlated. Genera such as *Lactococcus* and *Mycobacterium* showed decrease in relative enrichment. Our samples were mostly enriched with genus *Prevotella*. No clear trend was found across all samples and even after 1 month of DOTS. Though gut was highly enriched with genus *Prevotella* in all healthy individuals and in almost all TB samples. Though high fluctuations were observed in two

samples, “Y” and “T” which calls for increasing the sample size for more visible and fruitful results (figure 3).

5.3 Comparative taxonomic diversity analysis

This was followed by comparative analysis to assess the effect of disease and treatment (antibiotics dosage in this case) on the human gut microbiome. Inter-individual (controls versus TB patients (0^{th} day)) analysis and intra-individual analysis (TB patients on 0^{th} day, TB patients on 7^{th} day, TB patients on 30^{th} day) was performed using the relative abundance values obtained using MetaPhlAn. The mean relative abundance of genus *Prevotella* was lower in TB patients as compared to their healthy counterparts. It suddenly showed a jump in abundance after 1 week of DOTS and continued the same trend at the end of 1 month of DOTS. However, the behavior shown by sample ‘Y’ in terms of genus *Prevotella* was strikingly different which was evident from it being reduced to only a minority genus in TB patients as compared to its healthy counterpart and other TB samples. Even after 1 week of DOTS, the abundance was very much below the average level and this was even visible after 1 month of DOTS. Genus *Mycobacterium* had same abundance in “T” control and “T” sample obtained after following 1 month of DOTS. In case of sample “AB”, it increased with TB and then declined with DOTS. The untreated TB samples and samples obtained following 1 month of DOTS were rich in *Lactobacillus ruminis*, specifically in sample “T”. *Eubacterium rectale* showed high abundance in TB (0^{th} day) samples. After 1 month of DOTS, in TB patients’ genus *Ruminococcus* was always found to be less abundant but its relative abundance was close to that of their controls. Genus *Escherichia* was relatively more abundant in TB patients (0^{th} day). They also showed a decline after 1 week of DOTS and after 1 month of DOTS except in ‘Y’ sample where the its abundance increased at the end of 1 month of DOTS.

Some species were found to be absent in all the samples irrespective of the disease or antibiotics intake, which include *Fusobacterium mortiferum, Gardnerella vaginalis, Gemella haemolysis, Gemella moribillum, Gemella unclassified, Gramicatella adiacens* and *Haemophilus influenzae*.

Samples with BMI >= 25 (UTB and all controls) had mean relative abundance of 72.18, 2.38, 4.0477 for genera *Prevotella, Bacteroides* and *Eubacterium* respectively whereas samples with BMI <25 (all TB patients except UTB) had mean relative abundance of 47.45, 11.23 and 14.86 respectively for all the above genera.

As medication, in particular DOTS, can affect the gut microbiota but we found only minor effects of antibiotics within TB patient and among comparison of TB patients and healthy individuals. An interesting generalizable finding is that the genera *Prevotella*, in particular *Prevotella copri* is most abundant in healthy as well as TB patients.

The summarized results of comparison of few abundant taxa in controls and TB patients (0^{th} day) is represented in the figure below up to the family level of hierarchical classification (figure 2)

whereas the heatmap presenting top 25 species with horizontal and vertical clustering on a 'log' scale of all the 24 samples was plotted using MetaPhlAn_hclust_heatmap.py script of MetaPhlAn 1.7.8 (figure 4).

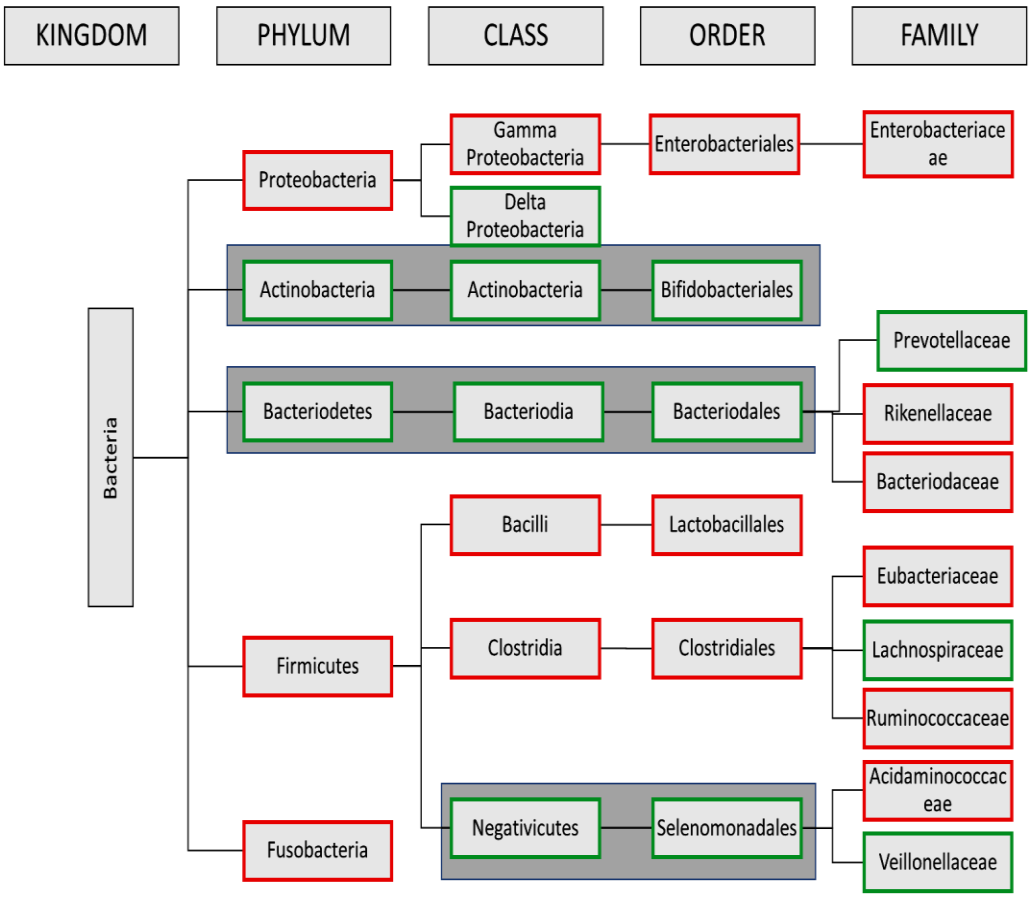
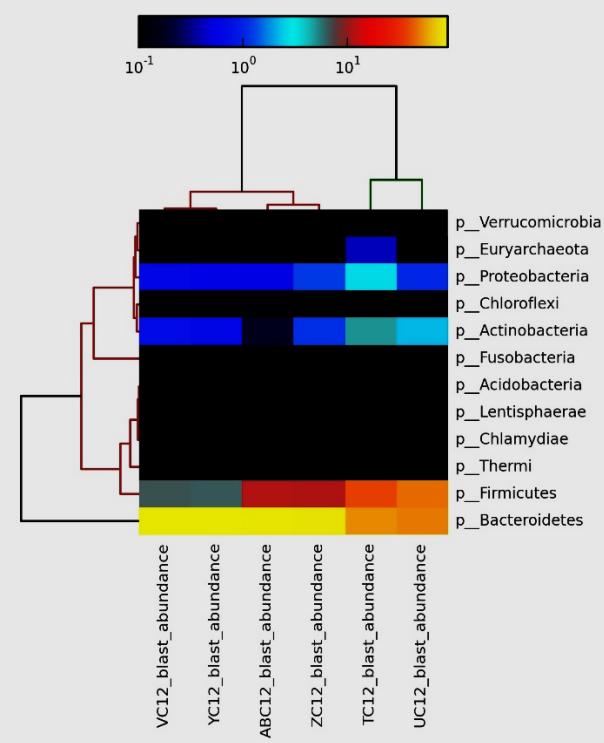
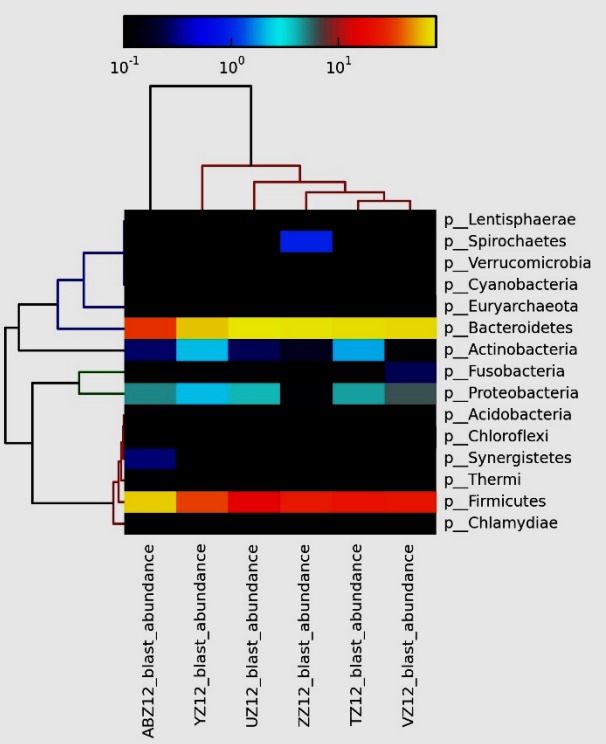


Figure 2: Schematic diagram indicating taxa/lineages average overabundance in TB patients and healthy people gut data sets. Area shaded in grey: lineages/taxas observed to be overabundant in healthy controls. Green boxes: taxa abundant in healthy people and red boxes: taxa abundant in TB patients (zero days)

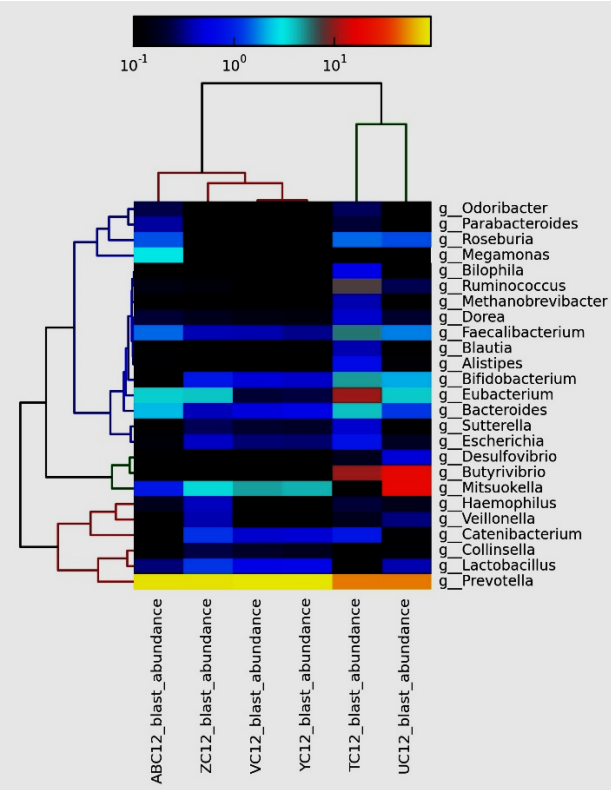
A


10-1
100
101
p__Verrucomicrobia
p__Euryarchaeota
p__Proteobacteria
p__Chloroflexi
p__Actinobacteria
p__Fusobacteria
p__Acidobacteria
p__Lentisphaerae
p__Chlamydiae
p__Thermi
p__Firmicutes
p__Bacteroidetes
VC12_blast_abundance
YC12_blast_abundance
ABC12_blast_abundance
ZC12_blast_abundance
TC12_blast_abundance
UC12_blast_abundance

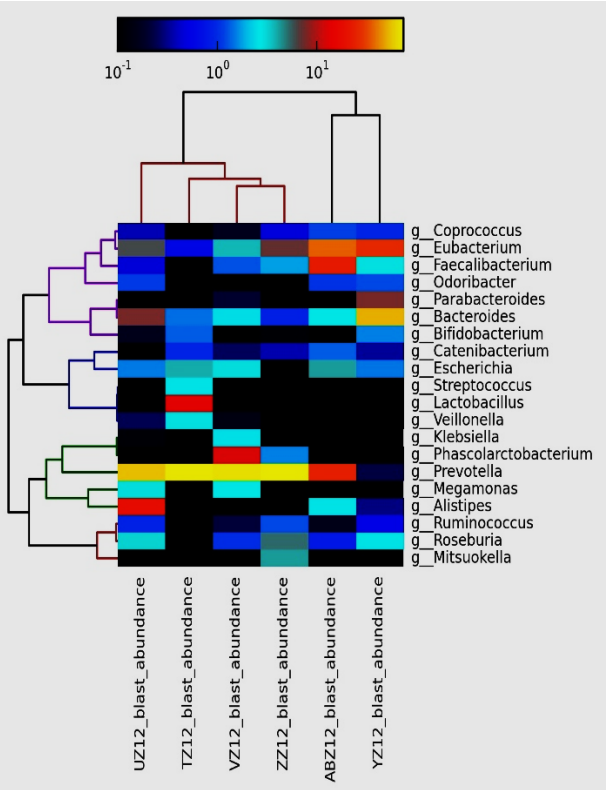
B


10-1
100
101
p__Lentisphaerae
p__Spirochaetes
p__Verrucomicrobia
p__Cyanobacteria
p__Euryarchaeota
p__Bacteroidetes
p__Actinobacteria
p__Fusobacteria
p__Proteobacteria
p__Acidobacteria
p__Chloroflexi
p__Synergistetes
p__Thermi
p__Firmicutes
p__Chlamydiae
ABZ12_blast_abundance
YZ12_blast_abundance
UZ12_blast_abundance
ZZ12_blast_abundance
TZ12_blast_abundance
VZ12_blast_abundance

C


10-1
100
101
g__Odoribacter
g__Parabacteroides
g__Roseburia
g__Megamonas
g__Bilophila
g__Ruminococcus
g__Methanobrevibacter
g__Dorea
g__Faecalibacterium
g__Blautia
g__Alistipes
g__Bifidobacterium
g__Eubacterium
g__Bacteroides
g__Sutterella
g__Escherichia
g__Desulfovibrio
g__Butyrivibrio
g__Mitsuokella
g__Haemophilus
g__Veillonella
g__Catenibacterium
g__Collinsella
g__Lactobacillus
g__Prevotella
ABC12_blast_abundance
ZC12_blast_abundance
VC12_blast_abundance
YC12_blast_abundance
TC12_blast_abundance
UC12_blast_abundance

D


10-1
100
101
g__Coprococcus
g__Eubacterium
g__Faecalibacterium
g__Odoribacter
g__Parabacteroides
g__Bacteroides
g__Bifidobacterium
g__Catenibacterium
g__Escherichia
g__Streptococcus
g__Lactobacillus
g__Veillonella
g__Klebsiella
g__Phascolarctobacterium
g__Prevotella
g__Megamonas
g__Alistipes
g__Ruminococcus
g__Roseburia
g__Mitsuokella
UZ12_blast_abundance
TZ12_blast_abundance
VZ12_blast_abundance
ZZ12_blast_abundance
ABZ12_blast_abundance
YZ12_blast_abundance

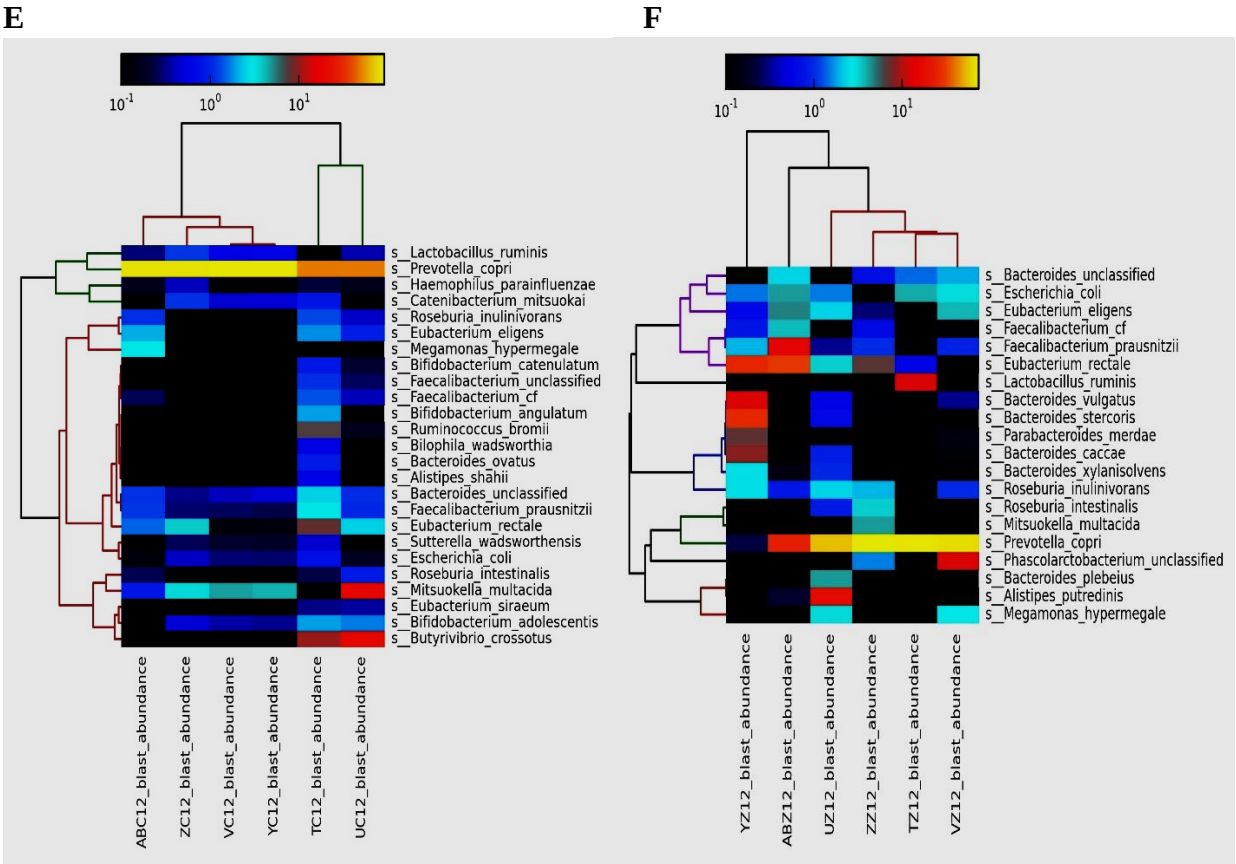


Figure 3: Heatmaps drawn using MetaPhlAn_hclust_heatmap.py script of MetaPhlAn 1.7.8 displaying top 20 results along with horizontal and vertical clustering at different hierarchical levels. **A.** Heatmap of all the controls at phylum level showing phylum Bacteroidetes to be most abundant in all samples and followed by phylum Firmicutes. **B.** Heatmap of all TB patients (0^{th} day) at phylum level phylum Bacteroidetes to be most abundant in all samples and followed by phylum Firmicutes. **C.** Heatmap of all the controls at genus level where genus *Prevotella* appeared to be most abundant in all samples and followed by genera *Bifidobacterium* and the *Eubacterium* **D.** Heatmap of all the TB patients at genus level where genus *Prevotella* appeared to be most abundant in all samples with an exception in sample "Y" and followed by genera *Eubacterium*, *Faecalibacterium* and *Bacteroides*. **E.** Heatmap of all the controls at species level where *Prevotella copri* shows higher relative abundance **F.** Heatmap of all the TB patients at species level where *Prevotella copri* shows higher relative abundance except the sample "Y".

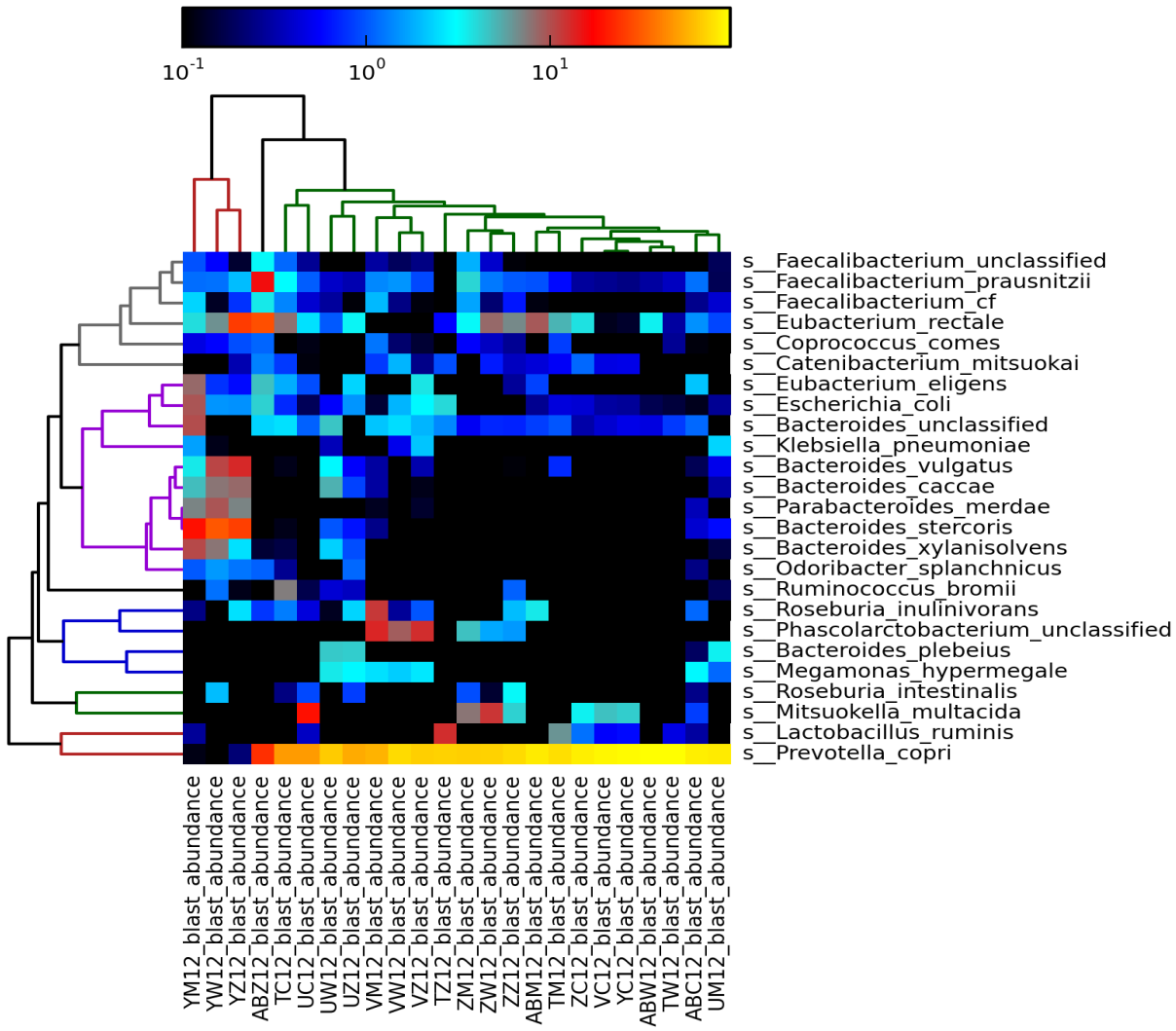


Figure 4: Heatmap showing comparative relative abundances at species level in all the 24 samples on a log scale with species and sample clustering.

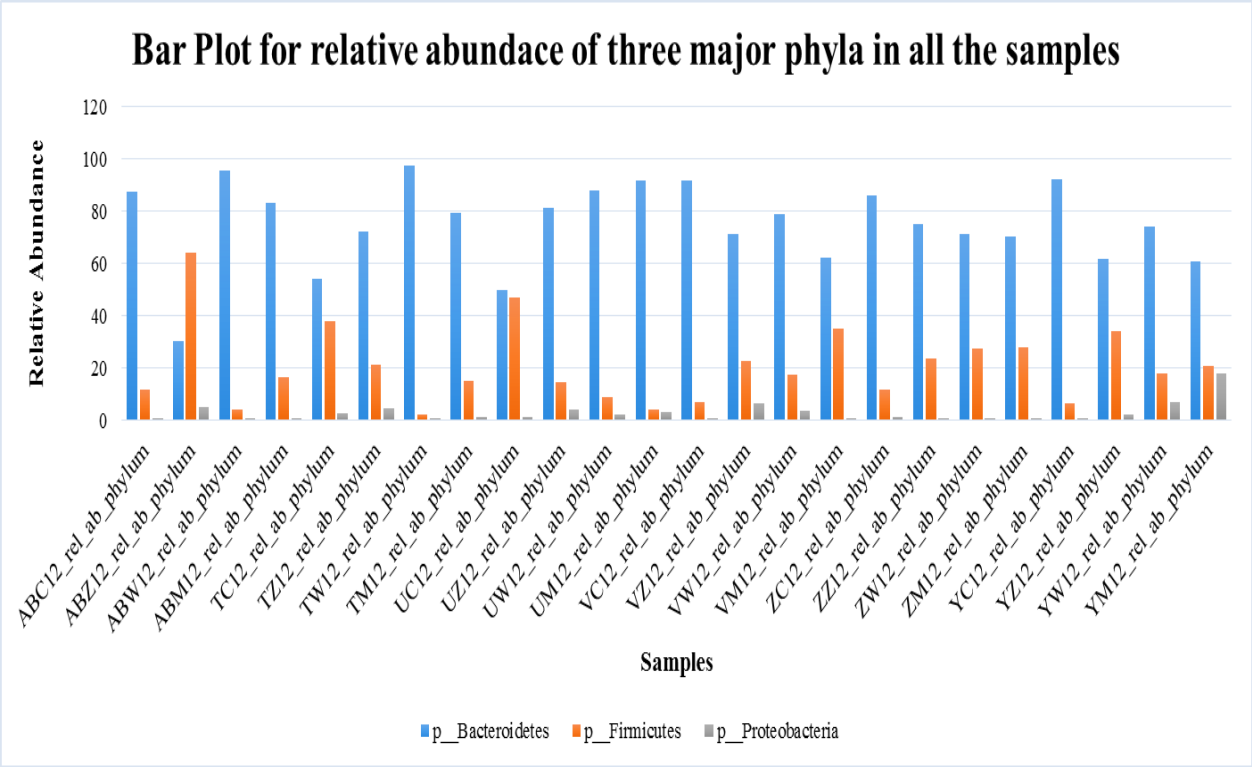


Figure 5: Barplot for relative abundance of phyla Bacteroidetes, Firmicutes and Proteobacteria. Phylum Bacteroidetes and phylum Firmicutes display an inverse relationship in almost all the samples, with phylum Bacteroidetes being the most abundant in all the samples.

Table 2: Table depicting abundant genera and their reported functions

Reported Genus in Control	**Reported Functions**	**Reported Genus in TB Patients**	**Reported Functions**
A: Bifidobacterium	Used as Probiotic	A: Bifidobacterium	Used as Probiotic
B: Collinsella	Core Microbiome Community	B: Coprococcus	Core microbiome Community
C: Blautia	Assimilation of Nutrients	C: Roseburia	Production of Butyrate

D: Roseburia	Production of Butyrate	D: Ruminococcus	Cellulose Metabolism
E: Butyrivibrio	Degradation of Plant fibers	E: Faecalibacterium	Potential Probiotic for IBD
F: Coprococcus	Core Microbiome Community	F: Eubacterium	Production of Fatty acids
G: Dorea	Not Reported	G: Phascolarctobacterium	Not Reported
H: Ruminococcus	Cellulose Metabolism	H: Veillonella	Reported to cause Irritable Bowel Syndrome
I: Faecalibacterium	Potential Probiotic for IBD	I: Mitsuokella	Prevention and Treatment of Hyperlactation
J: Eubacterium	Production of Fatty acids	J: Megamonas	Associated with Diabetes Mellitus Type II
K: Veillonella	Reported to cause Irritable Bowel Syndrome	K: Catenibacterium	Improvement in Immunity
L: Mitsuokella	Prevention and Treatment of Hyperlactation	L: Lactobacillus	Native member of Gut Microbiota
M: Megamonas	Associated with Diabetes Melliatus Type II	M: Bacteroides	Lean Body Profile
N: Catenibacterium	Improvement in Immunity	N: Alistipes	Not Reported

O: Lactobacillus	Native member of Gut Microbiota	O: Odoribacter	Reported to cause Colon Cancer
P: Streptococcus	Pathogenic	P: Parabacteroides	Reduction in the severity of Intestinal Inflammation
Q: Bacteroides	Lean Body Profile	Q: Prevotella	Pathogenic
R: Prevotella	Pathogenic	R: Escherichia	Native member of Gut Microbiota
S: Alistipes	Not Reported	S: Klebsiella	Facultative Pathogen in Immuno-compromised patients
T: Escherichia	Native member of Gut Microbiota		
U: Haemophilus	Pathogenic		
V: Desulfovibrio	Pathogenic		
W: Sutterella	Associated with autism		

5.4 Functional Analysis

Comparative metabolic potential analysis revealed interesting results providing useful insights on the gut microbiota between the control and zero day TB patient samples. In congruence with all the control as well as zero day samples, pathways like tetracycline biosynthesis, polyketide sugar unit biosynthesis, D-alanine metabolism and biosynthesis of vancomycin group antibiotics were revealed to be the among the most enriched. These pathways irrespective of the physiological state were found to be generally enriched contributing majorly to the gut metabolism. Although, opposed to this pattern sample "Y" showed variations; Y control revealed all the above pathways to be abundant but its TB counterpart (0^{th} day) sample showed enrichment of selenoamino acid metabolism, Glycolysis/Gluconeogenesis, Galactose metabolism and benzoate degradation pathway. Here, we can hypothesize the individual sample variation to be the cause of such striking difference. Similarly, sample AB also showed small variation in the overall pattern with the control sample showing enrichment of tetracycline biosynthesis, biosynthesis of vancomycin group antibiotics , D-Alanine metabolism and polyketide sugar unit biosynthesis whereas in case of the zero day patient sample pathways like methionine metabolism, streptomycin biosynthesis, inositol phosphate metabolism, and nucleotide excision repair were more abundant.

Pathways abundant only in TB (0^{th} day) samples: ABC transporters-General, Aminosugar metabolism except in "T" sample, Bacterial chemotaxis, Beta-Lactam resistance with exception in "T" sample, Butanoate metabolism with exception in 'T' sample, Fatty acid metabolism with exception in "Z" sample, Gamma Hexachlorohexane degradation, Glutamate metabolism, Glycerophospholipid metabolism, Glutathione metabolism, Glycine, Serine and Threonine metabolism, with exception in 'T' sample, Glycolysis/Gluconeogenesis with exception in 'T' sample, Glyoxylate and dicarboxylate metabolism, peptidoglycan biosynthesis, Phenylalanine metabolism, Purine metabolism, Pyruvate metabolism, Selenoamino acid metabolism with exception in 'T' sample, Sphingolipid metabolism with exception in 'T' and 'U' sample where it is 0, Synthesis and degradation of ketone bodies osbserved in AB, U, Y TB patients, Trinitrotoluene degradation with exception in 'T' sample, Two-component system-General, Ubiquinone biosynthesis, Valine, Leucine and isoleucine biosynthesis, Lysine degradation.

Pathways abundant in control samples: Alkaloid biosynthesis-II, Type IV secretion system in most controls.

This has been shown in the form of heatmap in Supplementary Figure 1 (**Appendix**).

6. Discussion

TB is a widespread infectious disease caused by various strains of *Mycobacterium.* Though the main cause of TB is *Mycobacterium* infection but a number of factors such as HIV, smoking, chronic lung diseases (example silicosis), overcrowding and malnutrition make individuals more susceptible to TB. The standard DOTS treatment for TB includes isoniazid (along with pyridoxal phosphate), pyrazinamide, rifampicin and ethambutol for 2 months, then for 4 months isoniazid and rifampicin alone are administered. After six months, the patient is generally observed to be free of *Mycobacterium* infection. The present study with the help of metagenomic whole genome sequence analysis, highlights how the gut microbial composition differs between healthy and TB patients along with the changes it undergo during DOTS. The study was primarily motivated by the observation that the composition of the microbiota tends to change with the administration of antibiotics and disease progression (Ferrer *et al.,* 2013; Hara *et al.,* 2013).

The abundant phyla in TB patients were Bacteroidetes, Firmicutes, Proteobacteria and Actinobacteria but there level of richness is less as in healthy controls. A negative correlation was observed between phylum Bacteroides and phylum Firmicutes in all the samples irrespective of the pathophysiological condition and antibiotics intake with phylum Bacteroidetes being the major phylum (figure 5). This trend was visible up to the subsequent order level. Phylum Bacteroidetes include three classes of gram-negative, non-spore forming, aerobic or anaerobic bacteria. Class Bacteroidia is the most studied class which includes the genus *Bacteroides* (usually abundant in the warm-blooded animals including humans). These are mostly opportunistic pathogens. The capability of Bacteroidetes to feed on carbohydrates, their ancient symbiotic relationship with their hosts (Bäckhed et al., 2005; Xu et al., 2007) and their known vital position in the gut microbiome might have helped to sustain their abundance over time. Phylum Firmicutes includes mostly gram-positive bacteria exhibiting low G+C content in their genome. It has two major classes: class Clostridia and Bacilli. A recent study has demonstrated the existence of mainly three Enterotypes in the gut: *Prevotella, Bacteriodes* and *Ruminococcus*. *Prevotella* strives on carbohydrate and simple sugars, indicating their association with carbohydrate-rich diet typical of agrarian societies whereas *Bacteroides* enterotype is associated with diet components like animal proteins, amino acids and saturated fats, typical of western diet and *Ruminococcus* prevailed in those individuals that consume lot of alcohol and polyunsaturated fatty acids (PUFA). Though it is not clear, the type of microbial flora which is healthier in comparison to the others (Arumugam *et al.,* 2011). Further, phylum Proteobacteria seem to be the early colonizers of the mammalian gut and may, therefore, be less competitive and abundant than well-adapted late colonizers (Trosvik *et al.,* 2010).

There were lot of exceptional changes both from the point of diversity and functional composition, specifically in sample "T" and sample "Y". *Prevotella copri* was found to be highly abundance in healthy controls, particularly in "AB", "V", "Y" and "Z". *Prevotella copri*`s abundance increased after 1 week of DOTS and then sublimes a bit after 1 month of DOTS. 'T' and 'U' samples showed a dissimilar trend. In sample "U", its abundance increased with the treatment progression and was less abundance in control samples. Genus *Eubacterium* also showed an increase in abundance in TB patients except in 'T' where it declined. This can certainly be attributed to the diet or BMI as only sample "T" was consuming a vegetarian diet. Genus *Flavobacterium* was absent in all samples irrespective of the physiological state. This means it is unaffected by diet, age, TB or DOTS (antibiotics). Also genera *Stenutrophomonas* and *Solobacterium* were also found to be absent from all the samples. Additionally, as mentioned in the result section, contrary to the trends observed in other samples many pathways were less or more abundant in "T" samples (0^{th} day) in comparison to controls. This could be attributed to the type of diet consumed, as only sample "T" was known to live on vegetarian diet.

Some genera and species showed only slight fluctuations in their relative abundances in all the samples. They didn`t seem to follow any logical trend. We hypothesized that this behavior may be attributed: to the fact these genera or species fluctuations is a natural phenomenon even in healthy individuals, or some genetically and functionally similar genera or species tries to cover up for such fluctuations i.e. they might have negative correlation with any other genera or species. Relevant to this, we could say that antibiotics might not have much effect on these genera or species. Though, we have summarized the reported functions of the abundant genera in our controls and TB patient samples, for a quick reference (Table 2).

Some studies have reported the recovery of gut microbiota towards baseline in model animals after short-term antibiotic treatment therapy (Robinson and Young, 2010) but several weeks after the withdrawal of antibiotics for instance, quinolones (Dethlefsen *et al.*, 2008) and cefoparazone (Robinson and Young, 2010) evident changes were observed. Another study has revealed the effect of long-term antibiotic treatment with rifampicin on termite gut microbiota, leading to irreversible changes in microbial community composition along with effecting weight, longevity and fecundity (Rosengaus *et al.*, 2011).

The isoniazid, pyrazinamide, ethambutol are MTb specific drugs whereas Rifampicin is a broad-spectrum antibiotic. There seems to be minimal effect of short-term DOTS treatment on the human gut microbiota. However, complete DOTS treatment could have significant impact on the gut microbiota.

While relative importance of these results may not be able to disentangle the wires linking the gut microbiota and TB/DOTS treatment but we could definitely state certain conclusive results from them. The patterns of microbial abundance observed in our study seems to be relevant if not conclusive, for diagnostic and /or therapeutic approaches targeting the human gut microbiome. The results call for more extensive study which includes following up healthy

controls and patients for longer time and extending the study to metatranscriptomic and metaproteomic levels.

From the literature survey and based on the present study, we could additionally state few more things with quite a lot of certainty. In order to decipher links between gut microbiota and physiological or disease states, one theme that seemed common to the both taxonomic and functional analysis is that it would be more revealing to use larger sample size than to sequence each sample at greater sequencing depth. However, far deeper sequencing would be required for certain associations involving minority species and rare genes. In such cases, it might require direct gut mucosal sampling rather than relying on stool samples as a proxy. Relevant to this issue, it is also possible that certain undetectable genes or species might have disproportionately larger impact on the microbiota relative to its abundance. To this context, it becomes really important to distinguish the presence or abundance of which genes and species is affected due to a particular physiological states and due to which physiological states is the result of presence or abundance of certain genes and species. Though, it has been possible to establish causality in a few cases such as the one involving, stool transplantations between different humans leading to cure persistent *Clostridium difficile* infections, involving transfer of microbial communities that are collectively associated with a physiological state between different mice or even between humans and mouse. However, several ELSI (Ethical, Legal and Social Issues) prevent the manipulation of the human microbiome.

Our study also has some drawbacks: small sample size, less prolonged treatment samples, DNA was not extracted for healthy sample at the same time points as for the diseased individual, lack of standardized protocol for taxonomic and functional assignment of sequence reads.

Differentiating between causal and mere side effect associations between microbiota and physiological states from the prospective time series studies, is still an undoable task. Also, the understanding of our microbial counterparts harbored in the human body is still shallow. Numerous research projects in a wide variety of microbes for sequencing their complete genomes are underway. Additionally, it can be challenging to interpret and understand the biological relevance of a differentially represented group of gene functions between two or more samples. Though an increasing number of tools and approaches that allow the identification of considerable differences between this lists are being developed (Kuczynski *et al.*, 2011). All this is thought to have an intense effect on our understanding of the disease-associated strains and pathogenicity.

7. Future Perspective

The results obtained till now from our current study combined with the recent studies reveal that though studying metagenomics using fecal samples gives quite a good results but we need to move on with direct sampling of the gut cells. They also compel us to take the next step and perform systematic studies that involves multiple disorders. We hereby can say that we also have plans to study the human gut microbiota associated with HIV-TB patients. These studies should be guided by standardized pipelines and experts to obtain meaningful results. The results would help address questions pertaining to the loss and recovery of microbial community during the course of treatment. Additionally, our findings emphasize to conduct large scale, well characterized epidemiological studies to detect variations among healthy and diseased individuals.

Our findings have added to the perspective of using metagenomic analysis of whole genome shotgun sequencing of DNA obtained from fecal matter aimed at characterizing disease-associated deviations from a healthy gut microbiota. This study gives us many additional research assignments such as: comparative study of the human gut microbiota among different populations including Indian populations using the whole genome sequencing data, new studies to separate from the fecal DNA virus-like particles and identify CRISPR elements using sequencing data, identifying additionally the antibiotic resistance genes (ARDB) in the available data and thus interpret the effect of antibiotic resistance genes on the variations in human gut microbiota, study long-term effect of DOTS on human gut microbiota of TB patients. Additionally this study provides us the opportunity to standardize the pipeline for analyzing the whole genome metagenomic sequencing data as well as comparing these results with the results obtained using 16s rRNA sequencing data, thus differentiating between them in the terms of efficiency and accuracy.

8. Conclusion

The human gut encompasses an ensemble of microbes mainly bacteria along with virus, fungi and eukaryotes which have been linked to a wide range health phenotypes and pathophysiologies. Molecular approaches have provided spectral insights into the taxonomic diversity and functional profile of the human gut microbiota and allowed tracking changes associated with the disease and associated treatment therapies. Designing novel interventions aimed at improving symbiotic relationships and diseased state is tend to improve with better understanding of the relationship between microbial communities and the host health.

With respect to our understanding and interpretation of the microbial patterns associated with disease and DOTS, we may conclude that short-term surveillance of TB patients under DOTS pointed towards it minimal effect on the gut microbiota. Also, no fixed trends were observed in case of both taxonomic diversity and functional composition of the gut. We may hypothesize that microbial shifts observed may be the result of positive or negative correlations between different genera or species. This also points to the fact that there is functional redundancy across taxonomic diversity, and different microbial species in our sample may have similar pathways. These results also suggests us that we cannot accurate make conclusions on such a small sample size and thus emphasize to conduct large scale, well characterized studies to detect variations among healthy and diseased individuals.

This research will open doors for further investigations for engineering microbiota either directly (using prebiotics, probiotics, drugs) or indirectly (using diet as preventive measures or treatment therapy) for TB. Also, metatranscriptomic and metaproteomic quantification of metagenomic sample is expected to yield fascinating insights into our relationship and that of TB with our gut microbial ecosystem.

References

Arumugam, M; Raes, J; Pelletier, E; Le Paslier, D; Yamada, T; Mende, DR; Fernandes, GR; Tap, J; Bruls, T; Batto, J-M; Bertalan, M; Borruel, N; Casellas, F; Fernandez, L; Gautier, L; Hansen, T; Hattori, M; Hayashi, T; Kleerebezem, M; Kurokawa, K; Leclerc, M; Levenez, F; Manichanh, C; Nielsen, HB; Nielsen, T; Pons, N; Poulain, J; Qin, J; Sicheritz-Ponten, T; Tims, S; Torrents, D; Ugarte, E; Zoetendal, EG; Wang, J; Guarner, F; Pedersen, O; de Vos, WM; Brunak, S; Doré, J; Antolín, M; Artiguenave, F; Blottiere, HM; Almeida, M; Brechot, C; Cara, C; Chervaux, C; Cultrone, A; Delorme, C; Denariaz, G; Dervyn, R; Foerstner, KU; Friss, C; van de Guchte, M; Guedon, E; Haimet, F; Huber, W; van Hylckama-Vlieg, J; Jamet, A; Juste, C; Kaci, G; Knol, J; Lakhdari, O; Layec, S; Le Roux, K; Maguin, E; Mérieux, A; Melo Minardi, R; M'rini, C; Muller, J; Oozeer, R; Parkhill, J; Renault, P; Rescigno, M; Sanchez, N; Sunagawa, S; Torrejon, A; Turner, K; Vandemeulebrouck, G; Varela, E; Winogradsky, Y; Zeller, G; Weissenbach, J; Ehrlich, SD; Bork, P. 2011. Enterotypes of the human gut microbiome. Nature 473, 174–180. doi:10.1038/nature10187

Bäckhed, F; Ley, RE; Sonnenburg, JL; Peterson, DA; Gordon, JI. 2005. Host-bacterial mutualism in the human intestine. Science 307, 1915–1920. doi:10.1126/science.1104816

Baker, BJ; Sheik, CS; Taylor, CA; Jain, S; Bhasi, A; Cavalcoli, JD; Dick, GJ. 2013. Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. ISME J. 7, 1962–73. doi:10.1038/ismej.2013.85

Bengmark, S. 1998. Ecological control of the gastrointestinal tract. The role of probiotic flora. Gut 42, 2–7. doi:10.1136/gut.42.1.2

Burcelin, R; Serino, M; Chabo, C; Blasco-Baque, V; Amar, J. 2011. Gut microbiota and diabetes: From pathogenesis to therapeutic perspective. Acta Diabetol. 48, 257–273. doi:10.1007/s00592-011-0333-6

Chan, JZ-M; Sergeant, MJ; Lee, OY-C; Minnikin, DE; Besra, GS; Pap, I; Spigelman, M; Donoghue, HD; Pallen, MJ. 2013. Metagenomic Analysis of Tuberculosis in a Mummy. Nejm 369, 289–290. doi:10.1056/NEJMc1302295

Cho, I; Blaser, MJ. 2012. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260–270. doi:10.1038/nrg3182

Clayton, TA; Lindon, JC; Cloarec, O; Antti, H; Charuel, C; Hanton, G; Provost, J-P; Le Net, J-L; Baker, D; Walley, RJ; Everett, JR; Nicholson, JK. 2006. Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature 440, 1073–1077. doi:10.1038/nature04648

D'Argenio, V; Precone, V; Casaburi, G; Miele, E; Martinelli, M; Staiano, A; Salvatore, F; Sacchetti, L. 2013. An altered gut microbiome profile in a child affected by Crohn's disease normalized after nutritional therapy. Am. J. Gastroenterol. 108, 851–2. doi:10.1038/ajg.2013.46

De Filippo, C; Ramazzotti, M; Fontana, P; Cavalieri, D. 2012. Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. Brief. Bioinform. 13, 696–710. doi:10.1093/bib/bbs070

Dethlefsen, L; Huse, S; Sogin, ML; Relman, DA. 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol. 6, e280. doi:10.1371/journal.pbio.0060280

Fang, S; Evans, RM. 2013. Wealth management in the gut. Nature 1, 9–10. doi:10.1038/500538a

Ferrer, M; dos Santos, VAPM; Ott, SJ; Moya, A. 2013. Gut microbiota disturbance during antibiotic therapy: A multi-omic approach. Gut Microbes 5, 1591–1601. doi:10.1136/gutjnl-2012-303184

Finegold, SM. 2008. Therapy and epidemiology of autism-clostridial spores as key elements. Med. Hypotheses 70, 508–511. doi:10.1016/j.mehy.2007.07.019

Gilbert, JA; Meyer, F; Bailey, MJ. 2011. The future of microbial metagenomics (or is ignorance bliss?). ISME J. 5, 777–779. doi:10.1038/ismej.2010.178

Gill, SR; Pop, M; Deboy, RT; Eckburg, PB; Turnbaugh, PJ; Samuel, BS; Gordon, JI; Relman, DA; Fraser-Liggett, CM; Nelson, KE. 2006. Metagenomic analysis of the human distal gut microbiome. Science 312, 1355–1359. doi:10.1126/science.1124234

Gollwitzer, ES; Marsland, BJ. 2014. Microbiota abnormalities in inflammatory airway diseases - Potential for therapy. Pharmacol. Ther. 141, 32–39. doi:10.1016/j.pharmthera.2013.08.002

Gori, A; Rizzardini, G; Van't Land, B; Amor, KB; van Schaik, J; Torti, C; Quirino, T; Tincati, C; Bandera, A; Knol, J; Benlhassan-Chahour, K; Trabattoni, D; Bray, D; Vriesema, A; Welling, G; Garssen, J; Clerici, M. 2011. Specific prebiotics modulate gut microbiota and immune activation in HAART-naive HIV-infected adults: results of the "COPA" pilot randomized trial. Mucosal Immunol. 4, 554–563. doi:10.1038/mi.2011.15

Greenblum, S; Turnbaugh, PJ; Borenstein, E. 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc. Natl. Acad. Sci. 109, 594–599. doi:10.1073/pnas.1116053109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1116053109

Guarner, F; Malagelada, J-R. 2003. Gut flora in health and disease. Lancet 360, 512–519. doi:10.1016/S0140-6736(03)12489-0

Hara, N; Alkanani, AK; Ir, D; Robertson, CE; Wagner, BD; Frank, DN; Zipris, D. 2013. The role of the intestinal microbiota in type 1 diabetes. Clin. Immunol. 146, 112–119. doi:10.1016/j.clim.2012.12.001

Hemsworth, J; Hekmat, S; Reid, G. 2011. The development of micronutrient supplemented probiotic yogurt for people living with HIV: Laboratory testing and sensory evaluation. Innov. Food Sci. Emerg. Technol. 12, 79–84. doi:10.1016/j.ifset.2010.11.004

Hermann-Bank, ML; Skovgaard, K; Stockmarr, A; Larsen, N; Mølbak, L. 2013. The Gut Microbiotassay: a high-throughput qPCR approach combinable with next generation sequencing to study gut microbial diversity. BMC Genomics 14, 788. doi:10.1186/1471-2164-14-788

Hill, D a; Hoffmann, C; Abt, MC; Du, Y; Kobuley, D; Kirn, TJ; Bushman, FD; Artis, D. 2010. Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. Mucosal Immunol. 3, 148–158. doi:10.1038/mi.2009.132

Holmes, E; Loo, RL; Stamler, J; Bictash, M; Yap, IKS; Chan, Q; Ebbels, T; De Iorio, M; Brown, IJ; Veselkov, K a; Daviglus, ML; Kesteloot, H; Ueshima, H; Zhao, L; Nicholson, JK; Elliott, P. 2008. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature 453, 396–400. doi:10.1038/nature06882

Hooper, L V; Gordon, JI. 2001. Commensal host-bacterial relationships in the gut. Science 292, 1115–1118. doi:10.1126/science.1058709

Jakobsson, HE; Jernberg, C; Andersson, AF; Sjölund-Karlsson, M; Jansson, JK; Engstrand, L. 2010. Short-term antibiotic treatment has differing long- term impacts on the human throat and gut microbiome. PLoS One 5. doi:10.1371/journal.pone.0009836

Jernberg, C; Löfmark, S; Edlund, C; Jansson, JK. 2007. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J. 1, 56–66. doi:10.1038/ismej.2007.3

Kinross, JM; Alkhamesi, N; Barton, RH; Silk, DB; Yap, IKS; Darzi, AW; Elaine, H; Nicholson, JK. 2011. Global metabolic phenotyping in an experimental laparotomy model of surgical trauma. J. Proteome Res. 10, 277–287. doi:10.1021/pr1003278

Kotanko, P; Carter, M; Levin, NW. 2006. Intestinal bacterial microflora--a potential source of chronic inflammation in patients with chronic kidney disease. Nephrol. Dial. Transplant 21, 2057–2060. doi:10.1093/ndt/gfl281

Kuczynski, J; Lauber, CL; Walters, W a.; Parfrey, LW; Clemente, JC; Gevers, D; Knight, R. 2011. Experimental and analytical tools for studying the human microbiome. Nat. Rev. Genet. 13, 47–58. doi:10.1038/nrg3129

Le Chatelier, E; Nielsen, T; Qin, J; Prifti, E; Hildebrand, F; Falony, G; Almeida, M; Arumugam, M; Batto, J-M; Kennedy, S; Leonard, P; Li, J; Burgdorf, K; Grarup, N; Jørgensen, T; Brandslund, I; Nielsen, HB; Juncker, AS; Bertalan, M; Levenez, F; Pons, N; Rasmussen, S; Sunagawa, S; Tap, J; Tims, S; Zoetendal, EG; Brunak, S; Clément, K; Doré, J; Kleerebezem, M; Kristiansen, K; Renault, P; Sicheritz-Ponten, T; de Vos, WM; Zucker, J-D; Raes, J; Hansen, T; Bork, P; Wang, J; Ehrlich, SD; Pedersen, O. 2013. Richness of human gut microbiome correlates with metabolic markers. Nature 500, 541–6. doi:10.1038/nature12506

Li, J; Jia, H; Cai, X; Zhong, H; Feng, Q; Sunagawa, S; Arumugam, M; Kultima, JR; Prifti, E; Nielsen, T; Juncker, AS; Manichanh, C; Chen, B; Zhang, W; Levenez, F; Wang, J; Xu, X; Xiao, L; Liang, S; Zhang, D; Zhang, Z; Chen, W; Zhao, H; Al-Aama, JY; Edris, S; Yang, H; Wang, J; Hansen, T; Nielsen, HB; Brunak, S; Kristiansen, K; Guarner, F; Pedersen, O; Doré, J; Ehrlich, SD; Bork, P; Wang, J. 2014. An integrated catalog of reference genes in the human gut microbiome. Nat. Biotechnol. 32, 834–841. doi:10.1038/nbt.2942

Macpherson, AJ; Gatto, D; Sainsbury, E; Harriman, GR; Hengartner, H; Zinkernagel, RM. 2000. A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. Science 288, 2222–2226. doi:10.1126/science.288.5474.2222

Marchesi, JR; Holmes, E; Khan, F; Kochhar, S; Scanlan, P; Shanahan, F; Wilson, ID; Wang, Y. 2007. Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. J. Proteome Res. 6, 546–551. doi:10.1021/pr060470d

Mazmanian, SK; Cui, HL; Tzianabos, AO; Kasper, DL. 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell 122, 107–118. doi:10.1016/j.cell.2005.05.007

Methé, BA; Nelson, KE; Pop, M; Creasy, HH; Giglio, MG; Huttenhower, C; Gevers, D; Petrosino, JF; Abubucker, S; Badger, JH; Chinwalla, AT; Earl, AM; FitzGerald, MG; Fulton, RS; Hallsworth-Pepin, K; Lobos, EA; Madupu, R; Magrini, V; Martin, JC; Mitreva, M; Muzny, DM; Sodergren, EJ; Versalovic, J; Wollam, AM; Worley, KC; Wortman, JR; Young, SK; Zeng, Q; Aagaard, KM; Abolude, OO; Allen-Vercoe, E; Alm, EJ; Alvarado, L; Andersen, GL; Anderson, S; Appelbaum, E; Arachchi, HM; Armitage, G; Arze, CA; Ayvaz, T; Baker, CC; Begg, L; Belachew, T; Bhonagiri, V; Bihan, M; Blaser, MJ; Bloom, T; Bonazzi, VR; Brooks, P; Buck, GA; Buhay, CJ; Busam, DA; Campbell, JL; Canon, SR; Cantarel, BL; Chain, PS; Chen, I-MA; Chen, L; Chhibba, S; Chu, K; Ciulla, DM; Clemente, JC; Clifton, SW; Conlan, S; Crabtree, J; Cutting, MA; Davidovics, NJ; Davis, CC; DeSantis, TZ; Deal, C; Delehaunty, KD; Dewhirst, FE; Deych, E; Ding, Y; Dooling, DJ;

Dugan, SP; Michael Dunne, W; Scott Durkin, A; Edgar, RC; Erlich, RL; Farmer, CN; Farrell, RM; Faust, K; Feldgarden, M; Felix, VM; Fisher, S; Fodor, AA; Forney, L; Foster, L; Di Francesco, V; Friedman, J; Friedrich, DC; Fronick, CC; Fulton, LL; Gao, H; Garcia, N; Giannoukos, G; Giblin, C; Giovanni, MY; Goldberg, JM; Goll, J; Gonzalez, A; Griggs, A; Gujja, S; Haas, BJ; Hamilton, HA; Harris, EL; Hepburn, TA; Herter, B; Hoffmann, DE; Holder, ME; Howarth, C; Huang, KH; Huse, SM; Izard, J; Jansson, JK; Jiang, H; Jordan, C; Joshi, V; Katancik, JA; Keitel, WA; Kelley, ST; Kells, C; Kinder-Haake, S; King, NB; Knight, R; Knights, D; Kong, HH; Koren, O; Koren, S; Kota, KC; Kovar, CL; Kyrpides, NC; La Rosa, PS; Lee, SL; Lemon, KP; Lennon, N; Lewis, CM; Lewis, L; Ley, RE; Li, K; Liolios, K; Liu, B; Liu, Y; Lo, C-C; Lozupone, CA; Dwayne Lunsford, R; Madden, T; Mahurkar, AA; Mannon, PJ; Mardis, ER; Markowitz, VM; Mavrommatis, K; McCorrison, JM; McDonald, D; McEwen, J; McGuire, AL; McInnes, P; Mehta, T; Mihindukulasuriya, KA; Miller, JR; Minx, PJ; Newsham, I; Nusbaum, C; O'Laughlin, M; Orvis, J; Pagani, I; Palaniappan, K; Patel, SM; Pearson, M; Peterson, J; Podar, M; Pohl, C; Pollard, KS; Priest, ME; Proctor, LM; Qin, X; Raes, J; Ravel, J; Reid, JG; Rho, M; Rhodes, R; Riehle, KP; Rivera, MC; Rodriguez-Mueller, B; Rogers, Y-H; Ross, MC; Russ, C; Sanka, RK; Sankar, P; Fah Sathirapongsasuti, J; Schloss, J a.; Schloss, PD; Schmidt, TM; Scholz, M; Schriml, L; Schubert, AM; Segata, N; Segre, J a.; Shannon, WD; Sharp, RR; Sharpton, TJ; Shenoy, N; Sheth, NU; Simone, GA; Singh, I; Smillie, CS; Sobel, JD; Sommer, DD; Spicer, P; Sutton, GG; Sykes, SM; Tabbaa, DG; Thiagarajan, M; Tomlinson, CM; Torralba, M; Treangen, TJ; Truty, RM; Vishnivetskaya, T a.; Walker, J; Wang, L; Wang, Z; Ward, D V.; Warren, W; Watson, MA; Wellington, C; Wetterstrand, K a.; White, JR; Wilczek-Boney, K; Qing Wu, Y; Wylie, KM; Wylie, T; Yandava, C; Ye, L; Ye, Y; Yooseph, S; Youmans, BP; Zhang, L; Zhou, Y; Zhu, Y; Zoloth, L; Zucker, JD; Birren, BW; Gibbs, R a.; Highlander, SK; Weinstock, GM; Wilson, RK; White, O. 2012. A framework for human microbiome research. Nature 486, 215–221. doi:10.1038/nature11209

Moriya, Y; Itoh, M; Okuda, S; Yoshizawa, AC; Kanehisa, M. 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, 182–185. doi:10.1093/nar/gkm321

Moss, A; LaMont, J. 2003. Clostridium difficile Infection and Pseudomembranous Colitis. Best Pract. Res. Clin. Gastroenterol. 17, 475–493. doi:10.1053/ybega.2003.375

Musso, G; Gambino, R; Cassader, M. 2011. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu. Rev. Med. 62, 361–380. doi:10.1146/annurev-med-012510-175505

O'Hara, AM; Shanahan, F. 2006. The gut flora as a forgotten organ. EMBO Rep. 7, 688–693. doi:10.1038/sj.embor.7400731

Patel, RK; Jain, M. 2012. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. PLoS One 7. doi:10.1371/journal.pone.0030619

Philippot, L; Raaijmakers, JM; Lemanceau, P; van der Putten, WH. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat. Rev. Microbiol. 11, 789–99. doi:10.1038/nrmicro3109

Ppatil, D; Pdhotre, D; Gchavan, S; Sultan, A; Jain, DS; Lanjekar, VB; Gangawani, J; Sshah, P; Stodkar, J; Shah, S; Ranade, DR; Patole, MS; Shouche, YS. 2012. Molecular analysis of gut microbiota in obesity among Indian individuals. J. Biosci. 37, 647–657. doi:10.1007/s12038-012-9244-0

Qin, J; Li, R; Raes, J; Arumugam, M; Burgdorf, KS; Manichanh, C; Nielsen, T; Pons, N; Levenez, F; Yamada, T; Mende, DR; Li, J; Xu, J; Li, S; Li, D; Cao, J; Wang, B; Liang, H; Zheng, H; Xie, Y; Tap, J; Lepage, P; Bertalan, M; Batto, J-M; Hansen, T; Le Paslier, D; Linneberg, A; Nielsen, HB; Pelletier, E; Renault, P; Sicheritz-Ponten, T; Turner, K; Zhu, H; Yu, C; Li, S; Jian, M; Zhou, Y; Li, Y; Zhang, X; Li, S; Qin, N; Yang, H; Wang, J; Brunak, S; Dore, J; Guarner, F; Kristiansen, K; Pedersen, O; Parkhill, J; Weissenbach, J; Bork, P; Ehrlich, SD; Wang, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65.

Rho, M; Tang, H; Ye, Y. 2010. FragGeneScan: Predicting genes in short and error-prone reads. Nucleic Acids Res. 38, 1–12. doi:10.1093/nar/gkq747

Ribeca, P; Valiente, G. 2011. Computational challenges of sequence classification in microbiomic data. Brief. Bioinform. 12, 614–625. doi:10.1093/bib/bbr019

Robinson, CJ; Young, VB. 2010. Antibiotic administration alters the community strcuture of the Gastrointestinal Micobiota. Gut Microbes 279–284. doi:10.1128/IAI.01520-08.e

Rosengaus, RB; Zecher, CN; Schultheis, KF; Brucker, RM; Bordenstein, SR. 2011. Disruption of the termite gut microbiota and its prolonged consequences for fitness. Appl. Environ. Microbiol. 77, 4303–4312. doi:10.1128/AEM.01886-10

Scher, JU; Sczesnak, A; Longman, RS; Segata, N; Ubeda, C; Bielski, C; Rostron, T; Cerundolo, V; Pamer, EG; Abramson, SB; Huttenhower, C; Littman, DR. 2013. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife 2013, 1–20. doi:10.7554/eLife.01202.001

Schwabe, RF; Jobin, C. 2013. The microbiome and cancer. Nat. Rev. Cancer 13, 800–12. doi:10.1038/nrc3610

Segata, N; Waldron, L; Ballarini, A; Narasimhan, V; Jousson, O; Huttenhower, C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods 9, 811–814. doi:10.1038/nmeth.2066

Singh, J; Rivenson, A; Tomita, M; Shimamura, S; Ishibashi, N; Reddy, BS. 1997. Bifidobacterium longum, a lactic acid-producing intestinal bacterium inhibits colon cancer and modulates the intermediate biomarkers of colon carcinogenesis. Carcinogenesis 18, 833–841. doi:10.1093/carcin/18.4.833

Smith, EA; Macfarlane, GT. 1996. Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. J. Appl. Bacteriol. 81, 288–302.

Stecher, B; Hardt, WD. 2008. The role of microbiota in infectious disease. Trends Microbiol. 16, 107–114. doi:10.1016/j.tim.2007.12.008

Trosvik, P; Stenseth, NC; Rudi, K. 2010. Convergent temporal dynamics of the human infant gut microbiota. ISME J. 4, 151–158. doi:10.1038/ismej.2009.96

Turnbaugh, PJ; Gordon, JI. 2009. The core gut microbiome, energy balance and obesity. J. Physiol. 587, 4153–4158. doi:10.1113/jphysiol.2009.174136

Turnbaugh, PJ; Ley, RE; Mahowald, MA; Magrini, V; Mardis, ER; Gordon, JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–1031. doi:10.1038/nature05414

Turnbaugh, PJ; Ridaura, VK; Faith, JJ; Rey, FE; Gordon, JI. 2009. Metagenomic Analysis in Humanized Gnotobiotic Mice. Sci Transl Med., 1(6): 6ra14. doi:10.1126/scitranslmed.3000322.

Vujkovic-Cvijin, I; Dunham, RM; Iwai, S; Maher, MC; Albright, RG; Broadhurst, MJ; Hernandez, RD; Lederman, MM; Huang, Y; Somsouk, M; Deeks, SG; Hunt, PW; Lynch, S V; McCune, JM. 2013. Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. Sci. Transl. Med. 5, 193. doi:10.1126/scitranslmed.3006438

Weir, TL; Manter, DK; Sheflin, AM; Barnett, B a.; Heuberger, AL; Ryan, EP. 2013. Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults. PLoS One 8. doi:10.1371/journal.pone.0070803

Xu, J; Mahowald, MA; Ley, RE; Lozupone, CA; Hamady, M; Martens, EC; Henrissat, B; Coutinho, PM; Minx, P; Latreille, P; Cordum, H; Van Brunt, A; Kim, K; Fulton, RS; Fulton, LA; Clifton, SW; Wilson, RK; Knight, RD; Gordon, JI. 2007. Evolution of Symbiotic Bacteria in the Distal Human Intestine. PLoS Biol. 5, 1574–1586. doi:10.1371/journal.pbio.0050156

Ye, Y; Doak, TG. 2011. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Metagenomes. Handb. Mol. Microb. Ecol. I Metagenomics Complement. Approaches 5, 453–460. doi:10.1002/9781118010518.ch52

Yeoh, N; Burton, JP; Suppiah, P; Reid, G; Stebbings, S. 2013. The role of the microbiome in rheumatic diseases. Curr. Rheumatol. Rep. 15, 314. doi: 10.1007/s11926-012-0314-y

Zajac, V; Ševurková, V; Mátelová, L; Ujházy, E. 2015. Detection of HIV-1 sequences in intestinal bacteria of HIV / AIDS patients. Neuroendocrinology Letters 28, 591-595

Zerbino, DR; Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829. doi:10.1101/gr.074492.107

Appendix

Supplementary Table 1: Table showing patient identities with the corresponding controls along with metadata (Age, Sex, Height, Diet pattern & BMI)

PATIENT_ID	AGE (yrs.)	SEX	HEIGHT (cm)	Wt.(Kg)	Diet	BMI
ABTB	14	F	145	38	NON-VEG	18.07
VTB	20	F	145	40	NON-VEG	19.02
TTB	47	M	156	60	VEG	24.65
YTB	22	M	140	36	NON-VEG	18.37
ZTB	35	M	160	50	NON-VEG	19.53
UTB	40	F	140	55	NON-VEG	28.06
CONTROL (corresponding)	**AGE (yrs.)**	**SEX**	**HEIGHT(cm)**	**Wt.(Kg)**	**Diet**	**BMI**
ABC	23	M	148	64	NON-VEG	29.2
VC	19	M	151	58	NON-VEG	25.4
TC	27	M	153	69	VEG	29.4
YC	31	M	149	76	NON-VEG	34.2
ZC	35	M	150	79	NON-VEG	35.1
UC	24	M	151	75	NON-VEG	32.8

NGSQC Toolkit

A toolkit for the quality control (QC) of next generation sequencing (NGS) data. The toolkit comprises of user-friendly standalone tools for quality control of the sequence data generated using Illumina and Roche 454 platforms with detailed results in the form of tables and graphs, and filtering of high-quality sequence data. It also includes few other tools, which are helpful in NGS data quality control and analysis.

Supplementary Table 2: NGSQC Toolkit 2.3.3 quality control statistics for each file of paired end reads.

Sample Name	Raw Reads	Trimmed Reads	QC Filtered	%Trimmed	%QC Filtered
ABC_1	38839511	38678053	38678053	**0.415705543**	**100**
ABC_2	38839511	38678053	38678053	**0.415705543**	**100**
ABZ_1	43567167	43420857	43420857	**0.335826289**	**100**
ABZ_2	43567167	43420857	43420857	**0.335826289**	**100**
ABW_1	47225694	46989475	46989475	**0.500191696**	**100**
ABW_2	47225694	46989475	46989475	**0.500191696**	**100**
ABM_1	44317517	44117024	44117024	**0.452401248**	**100**
ABM_2	44317517	44117024	44117024	**0.452401248**	**100**
TC_1	38983138	38920741	38920741	**0.160061512**	**100**
TC_2	38983138	38920741	38920741	**0.160061512**	**100**
TZ_1	48448506	48202662	48202662	**0.507433604**	**100**
TZ_2	48448506	48202662	48202662	**0.507433604**	**100**
TW_1	39267846	39090046	39090046	**0.452787759**	**100**
TW_2	39267846	39090046	39090046	**0.452787759**	**100**
TM_1	38044449	37910909	37910909	**0.351010472**	**100**
TM_2	38044449	37910909	37910909	**0.351010472**	**100**
UC_1	39702579	39543025	39543025	**0.401873138**	**100**
UC_2	39702579	39543025	39543025	**0.401873138**	**100**
UZ_1	40242362	40101628	40101628	**0.349716053**	**100**
UZ_2	40242362	40101628	40101628	**0.349716053**	**100**
UW_1	45447473	45372663	45372663	**0.164607612**	**100**
UW_2	45447473	45372663	45372663	**0.164607612**	**100**
UM_1	38719937	38623078	38623078	**0.250152783**	**100**
UM_2	38719937	38623078	38623078	**0.250152783**	**100**
VC_1	43947901	43876691	43876691	**0.162032767**	**100**
VC_2	43947901	43876691	43876691	**0.162032767**	**100**
VZ_1	49826432	49620005	49620005	**0.414292157**	**100**

VZ_2	49826432	49620005	49620005	**0.414292157**	**100**
VW_1	44085385	43868941	43868941	**0.49096543**	**100**
VW_2	44085385	43868941	43868941	**0.49096543**	**100**
VM_1	43236720	43043862	43043862	**0.446051412**	**100**
VM_2	43236720	43043862	43043862	**0.446051412**	**100**
YC_1	40020082	39879681	39879681	**0.350826368**	**100**
YC_2	40020082	39879681	39879681	**0.350826368**	**100**
YZ_1	45777551	45552526	45552526	**0.491561901**	**100**
YZ_2	45777551	45552526	45552526	**0.491561901**	**100**
YW_1	41449279	41263314	41263314	**0.448656779**	**100**
YW_2	41449279	41263314	41263314	**0.448656779**	**100**
YM_1	35599565	35510730	35510730	**0.249539566**	**100**
YM_2	35599565	35510730	35510730	**0.249539566**	**100**
ZC_1	47925728	47846351	47846351	**0.165625027**	**100**
ZC_2	47925728	47846351	47846351	**0.165625027**	**100**
ZZ_1	36989970	36838346	36838346	**0.409905712**	**100**
ZZ_2	36989970	36838346	36838346	**0.409905712**	**100**
ZW_1	34895276	34809655	34809655	**0.24536559**	**100**
ZW_2	34895276	34809655	34809655	**0.24536559**	**100**
ZM_1	32095327	32016761	32016761	**0.24478953**	**100**
ZM_2	32095327	32016761	32016761	**0.24478953**	**100**

Velvet 1.2.10

Velvet is an algorithm package that has been designed to deal with *de novo* genome assembly and short read sequencing alignments. This is achieved through the manipulation of de Bruijn graphs for genomic sequence assembly via the removal of errors and the simplification of repeated regions

$ ./shuffleSequences_fasta.pl filename1.fq filename2.fq > shuffledfilename.fq
[Where filename1.fq and filename2.fq represents the two paired-end short reads

$ ./velveth foldername 51 –fastq –shortPaired shuffledfilename.fq
[51=k-mer, shuffledfilename.fq is the file obtained after running shuffle commands]

$ ./velvetg foldername –exp_cov 200 –cov_cutoff 5 –ins_length 350 –ins_length_sd 20 –min_contig_lgth 500 –scaffolding no

Supplementary Table 3: Table showing finalized Assembly statistics of all the fecal sample DNA ran on different k-mer.

S. No.	Sample_ID	K-mer	Exp_Cov	Cov_cutoff	Nodes	N50	Assembly Size (Mbp)
1	AB Control	45	200	5	98986	5012	46
2	AB Zero	51	200	5	177785	5203	145
3	AB Week	57	200	5	55461	19560	38.8
4	AB month	61	200	5	54500	19041	49
5	T control	45	200	5	114960	7827	105.4
6	T zero	51	200	5	65593	5197	48.4
7	T week	47	200	5	76679	20387	51.9
8	T month	55	200	5	56238	15569	64.3
9	U Control	47	200	5	156426	8036	102.1
10	U Zero	41	200	5	110382	11657	130
11	U week	55	200	5	45285	23578	78.2
12	U month	55	200	5	31641	11073	55.2
13	V control	57	200	5	81060	10447	46.5
14	V Zero	55	200	5	69342	8968	64.8
15	V Week	51	200	5	37281	12279	49.6
16	V Month	41	200	5	113572	11496	100.2
17	Y Control	59	200	5	75368	10438	47.4
18	Y Zero	51	200	5	41199	33351	100
19	Y week	49	200	5	56315	20581	115.8
20	Y month	41	200	5	86644	14758	114
21	Z Control	57	200	5	102671	7421	54
22	Z Zero	49	200	5	133792	11575	110.3
23	Z week	45	200	5	164375	6603	91.5
24	Z month	45	200	5	170111	7073	119.8

FASTX Toolkit

The FASTX-Toolkit version 0.0.6 is a collection of command line tools for Short-Reads FASTA/FASTQ files preprocessing. Next-Generation sequencing machines usually produce FASTA or FASTQ files, containing multiple short-reads sequences (possibly with quality information). It is sometimes more productive to preprocess the FASTA/FASTQ files before mapping the sequences to the genome - manipulating the sequences to produce better mapping results. It was used to convert FASTQ files into FASTA files.

$ ./fastq_to_fasta –v –i ABC12.fq -o ABC12.fasta

[-v] = Verbose - report number of sequences.

[-i INFILE] = FASTA/Q input file. default is STDIN.

[-o OUTFILE] = FASTA output file. default is STDOUT.

MetaPhlAn

MetaPhlAn is a computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data. MetaPhlAn relies on unique clade-specific marker genes identified from 3,000 reference genomes, allowing:

- up to 25,000 reads-per-second (on one CPU) analysis speed (orders of magnitude faster compared to existing methods);
- unambiguous taxonomic assignments as the MetaPhlAn markers are clade-specific;
- accurate estimation of organismal relative abundance (in terms of number of cells rather than fraction of reads);
- species-level resolution for bacterial and archaeal organisms;
- extensive validation of the profiling accuracy on several synthetic datasets and on thousands of real metagenomes

For generating blast and abundance files

$ ./MetaPhlAn.py –input_type multifasta mergedfiles.fasta –t rel_ab –blastdb blastsdb/mpa –blastout Abundances/blastoutput_filename –nproc 8 >Abundances/samplename_abundance.txt

For extracting relative abundances at individual taxonomic levels

$ ./metaphlan.py -t rel_ab –tax_lev g –input_type blastout ABC12_Blastout.txt > ABC12_Abundance_genus.txt

Merging output files for different samples

$utils/merge_MetaPhlAn_tables.py Abundances/*.txt > output/merged_abundances.txt

The resulting table contains relative abundances with microbial clades as rows and samples as columns.

Plotting heatmaps

$ plotting_scripts/metaphlan_hclust_heatmap.py –tax_lev k -c bbcry –top 25 –minv 0.1 -s log –in output/merged_adundances_reads.txt –out output_images/adundances_heatmap.png

FragGene Scan 1.16

FragGeneScan is an application for finding (fragmented) genes in short reads.

$ ./run_FragGeneScan.pl –genome=read12.fasta –out=output12 –complete=1 –train=illumina_5

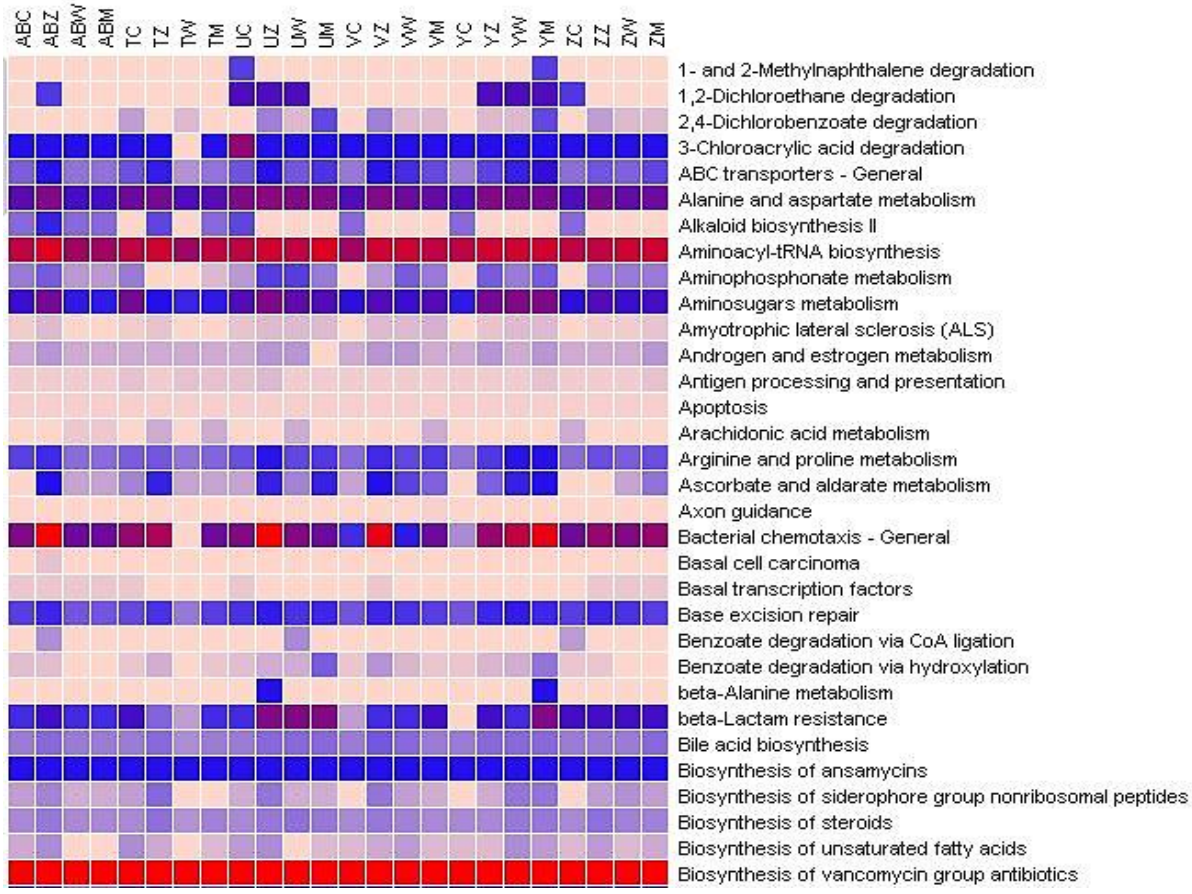
Upon completion, FragGeneScanR generates three files. The first file is [output_file_name].out, which lists the coordinates of putative genes. This file consists of five columns: start position, end position, strand, frame, and score. The second file is [output_file_name].ffn, which lists nucleotide sequences corresponding to the putative genes in "[output_file_name].out". The third file is [output_file_name].faa, which lists amino acid sequences corresponding to the putative genes in [output_file_name].out.

KEGGKAAS

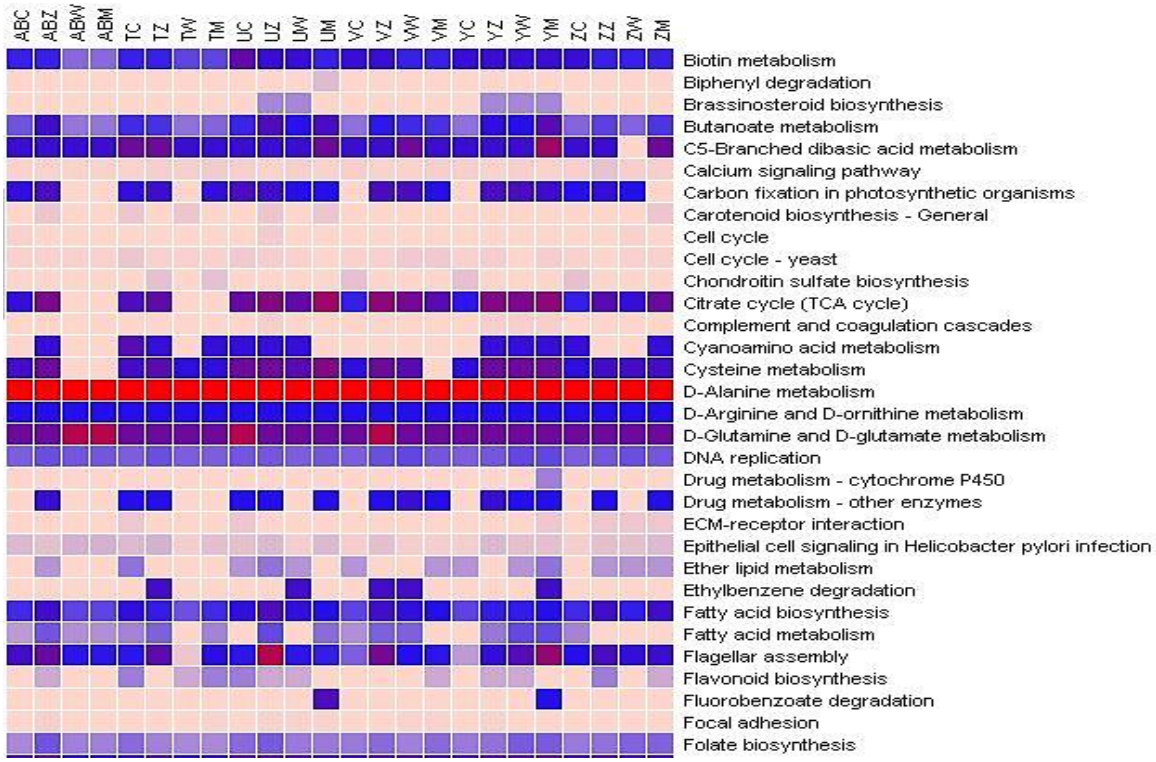
KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST or GHOST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways. The FASTA sequence file is uploaded on the web tool and e-mail ID is mentioned alongside to obtain results when done.

MinPATH

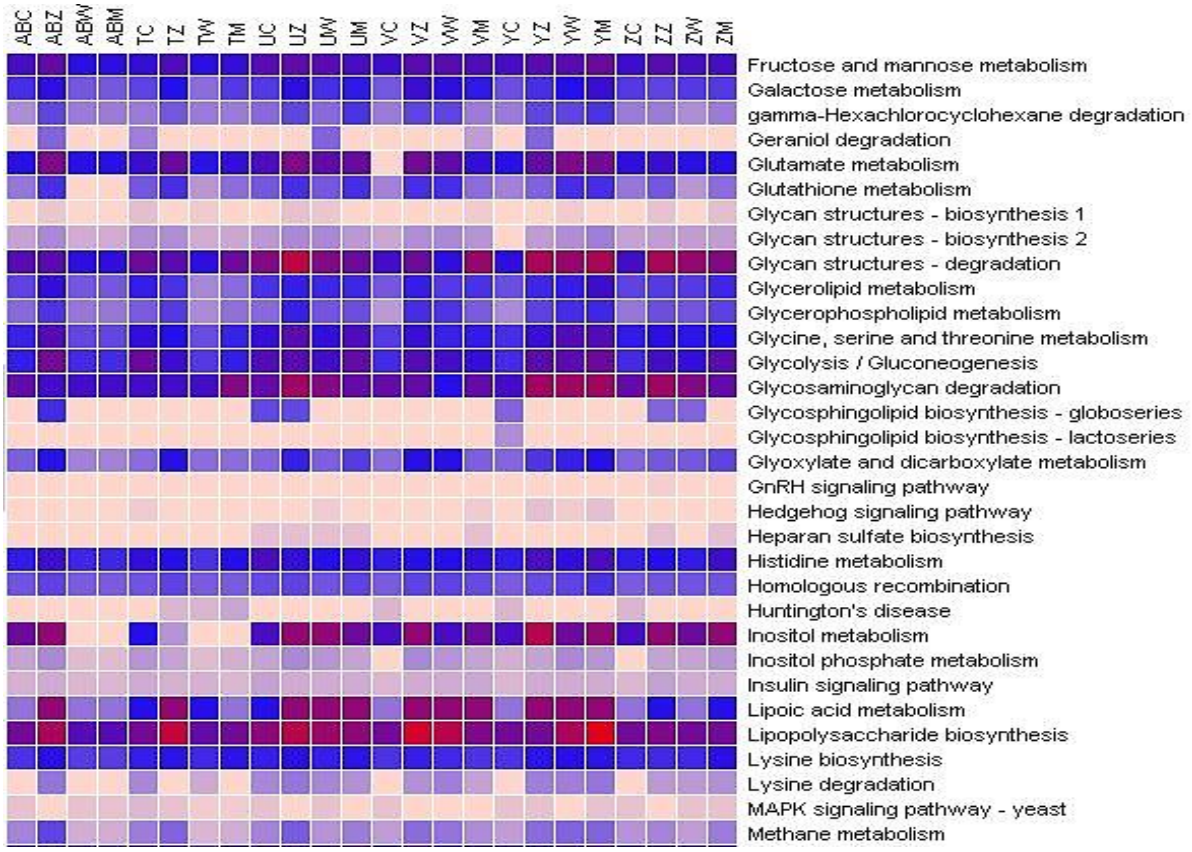
MinPath (Minimal set of Pathways) is a parsimony approach for biological pathway reconstructions using protein family predictions, achieving a more conservative, yet more faithful, estimation of the biological pathways for a query dataset. The ko files obtained using KEGGKAAS were uploaded directly onto the web tool to obtain the pathways involved in the given sample.


ABC
ABZ
ABW
ABM
TC
TZ
TW
TM
UC
UZ
UW
UM
VC
VZ
VW
VM
YC
YZ
YW
YM
ZC
ZZ
ZW
ZM
1- and 2-Methylnaphthalene degradation
1,2-Dichloroethane degradation
2,4-Dichlorobenzoate degradation
3-Chloroacrylic acid degradation
ABC transporters - General
Alanine and aspartate metabolism
Alkaloid biosynthesis II
Aminoacyl-tRNA biosynthesis
Aminophosphonate metabolism
Aminosugars metabolism
Amyotrophic lateral sclerosis (ALS)
Androgen and estrogen metabolism
Antigen processing and presentation
Apoptosis
Arachidonic acid metabolism
Arginine and proline metabolism
Ascorbate and aldarate metabolism
Axon guidance
Bacterial chemotaxis - General
Basal cell carcinoma
Basal transcription factors
Base excision repair
Benzoate degradation via CoA ligation
Benzoate degradation via hydroxylation
beta-Alanine metabolism
beta-Lactam resistance
Bile acid biosynthesis
Biosynthesis of ansamycins
Biosynthesis of siderophore group nonribosomal peptides
Biosynthesis of steroids
Biosynthesis of unsaturated fatty acids
Biosynthesis of vancomycin group antibiotics

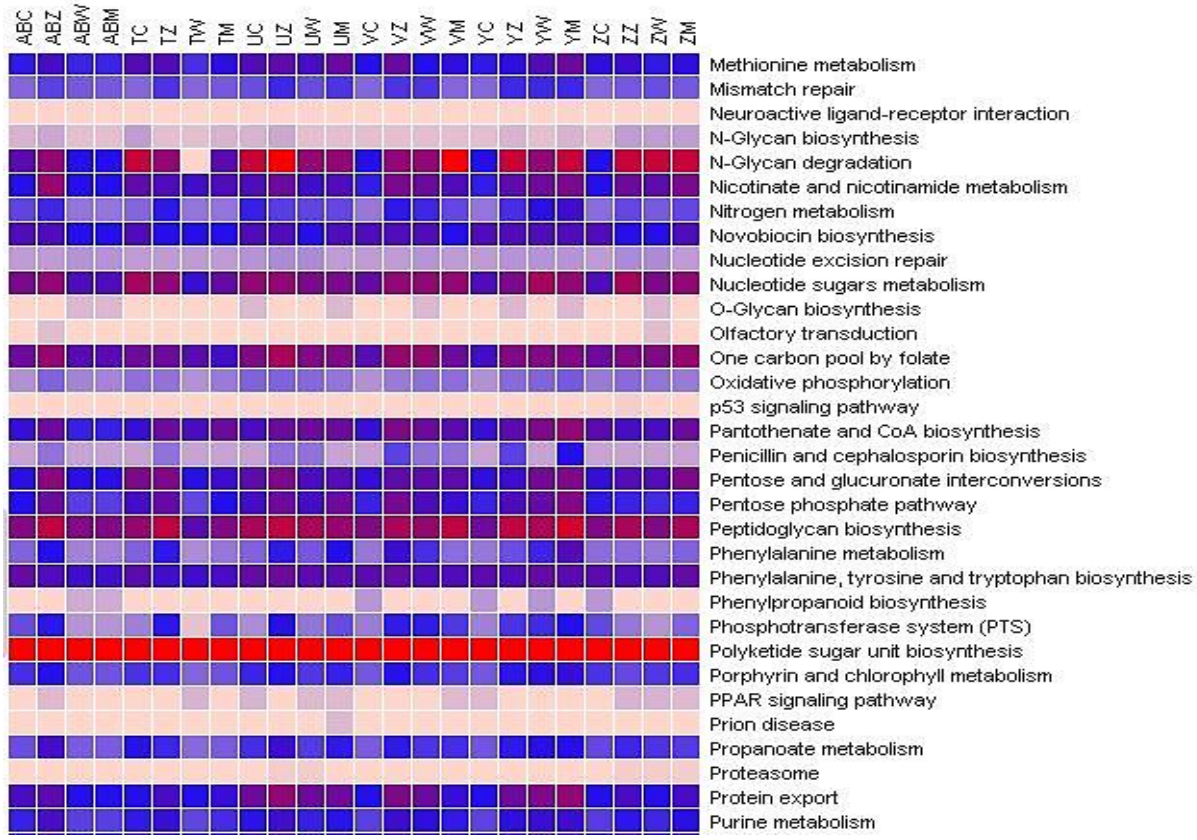
A


ABC
ABZ
ABW
ABM
TC
TZ
TW
TM
UC
UZ
UW
UM
VC
VZ
VW
VM
YC
YZ
YW
YM
ZC
ZZ
ZW
ZM
Biotin metabolism
Biphenyl degradation
Brassinosteroid biosynthesis
Butanoate metabolism
C5-Branched dibasic acid metabolism
Calcium signaling pathway
Carbon fixation in photosynthetic organisms
Carotenoid biosynthesis - General
Cell cycle
Cell cycle - yeast
Chondroitin sulfate biosynthesis
Citrate cycle (TCA cycle)
Complement and coagulation cascades
Cyanoamino acid metabolism
Cysteine metabolism
D-Alanine metabolism
D-Arginine and D-ornithine metabolism
D-Glutamine and D-glutamate metabolism
DNA replication
Drug metabolism - cytochrome P450
Drug metabolism - other enzymes
ECM-receptor interaction
Epithelial cell signaling in Helicobacter pylori infection
Ether lipid metabolism
Ethylbenzene degradation
Fatty acid biosynthesis
Fatty acid metabolism
Flagellar assembly
Flavonoid biosynthesis
Fluorobenzoate degradation
Focal adhesion
Folate biosynthesis

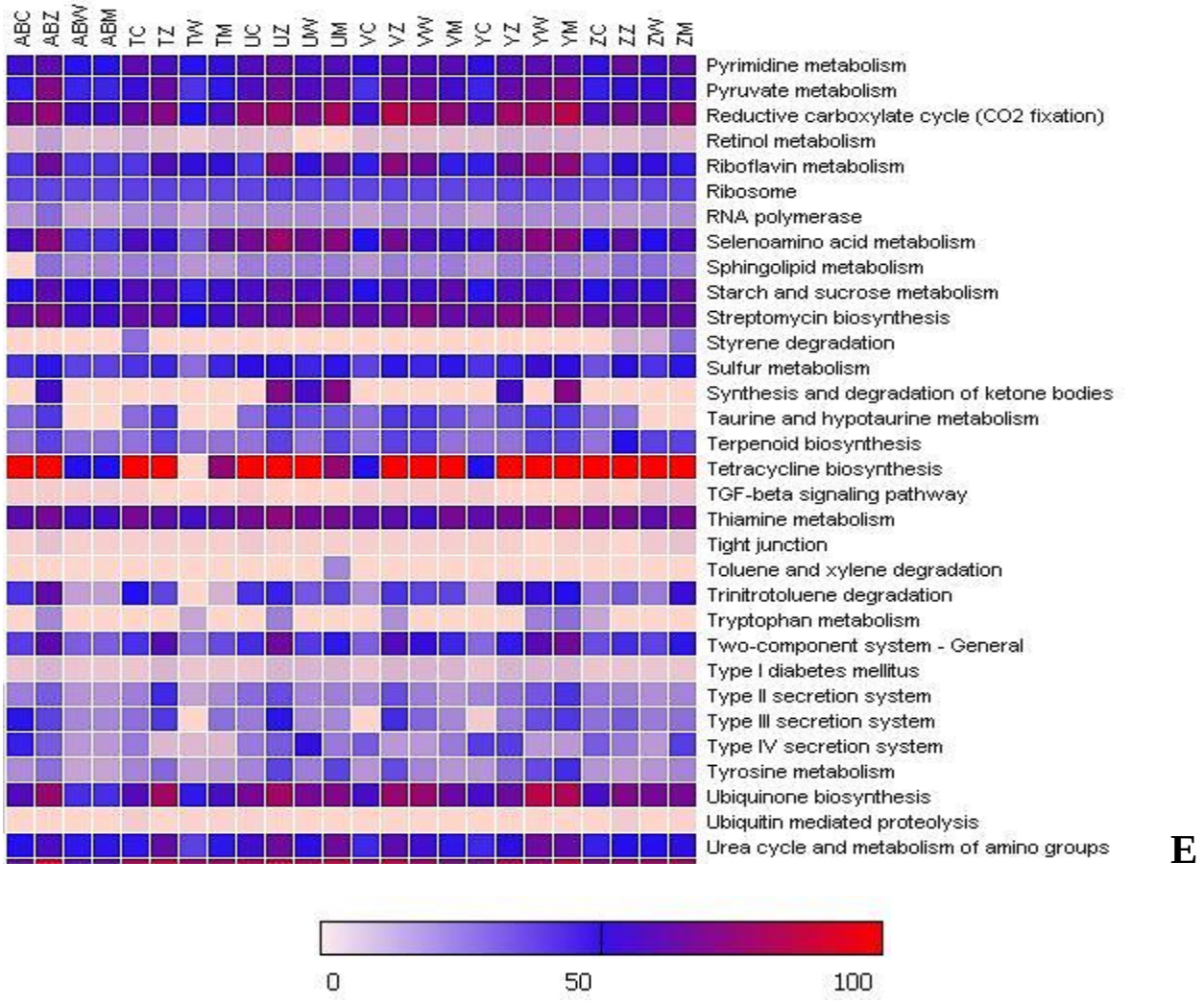
B


ABC
ABZ
ABW
ABM
TC
TZ
TW
TM
UC
UZ
UW
UM
VC
VZ
VW
VM
YC
YZ
YW
YM
ZC
ZZ
ZW
ZM
Fructose and mannose metabolism
Galactose metabolism
gamma-Hexachlorocyclohexane degradation
Geraniol degradation
Glutamate metabolism
Glutathione metabolism
Glycan structures - biosynthesis 1
Glycan structures - biosynthesis 2
Glycan structures - degradation
Glycerolipid metabolism
Glycerophospholipid metabolism
Glycine, serine and threonine metabolism
Glycolysis / Gluconeogenesis
Glycosaminoglycan degradation
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lactoseries
Glyoxylate and dicarboxylate metabolism
GnRH signaling pathway
Hedgehog signaling pathway
Heparan sulfate biosynthesis
Histidine metabolism
Homologous recombination
Huntington's disease
Inositol metabolism
Inositol phosphate metabolism
Insulin signaling pathway
Lipoic acid metabolism
Lipopolysaccharide biosynthesis
Lysine biosynthesis
Lysine degradation
MAPK signaling pathway - yeast
Methane metabolism

C


ABC
ABZ
ABW
ABM
TC
TZ
TW
TM
UC
UZ
UW
UM
VC
VZ
VW
VM
YC
YZ
YW
YM
ZC
ZZ
ZW
ZM
Methionine metabolism
Mismatch repair
Neuroactive ligand-receptor interaction
N-Glycan biosynthesis
N-Glycan degradation
Nicotinate and nicotinamide metabolism
Nitrogen metabolism
Novobiocin biosynthesis
Nucleotide excision repair
Nucleotide sugars metabolism
O-Glycan biosynthesis
Olfactory transduction
One carbon pool by folate
Oxidative phosphorylation
p53 signaling pathway
Pantothenate and CoA biosynthesis
Penicillin and cephalosporin biosynthesis
Pentose and glucuronate interconversions
Pentose phosphate pathway
Peptidoglycan biosynthesis
Phenylalanine metabolism
Phenylalanine, tyrosine and tryptophan biosynthesis
Phenylpropanoid biosynthesis
Phosphotransferase system (PTS)
Polyketide sugar unit biosynthesis
Porphyrin and chlorophyll metabolism
PPAR signaling pathway
Prion disease
Propanoate metabolism
Proteasome
Protein export
Purine metabolism

D



E

Supplementary Figure 1: Heatmap showing comparative abundances of pathways in the gut microbiome in all the 24 samples on a linear scale without clustering. “AB”, “T”, “U”, “V”, “Y” and “Z” are the sample names. “C” in the suffix represents control sample, “Z” in the suffix represents zero day samples, “W” in the suffix represent samples after 1 week of DOTS and “M” in the suffix represents samples after 1 month of DOTS.