THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Metabolic Modeling of the Gut Microbiome-Host Interactions and Meta'omics Integration

SAEED SHOAIE

CHALMERS

Department of Biology and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2015

Metabolic Modeling of the Gut Microbiome-Host Interactions and Meta'omics Integration
SAEED SHOAIE

ISBN 978-91-7597-256-5

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie Nr 3937
ISSN 0346-718X

Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2015

Metabolic Modeling of the Gut Microbiome-Host Interactions and Meta'omics Integration

SAEED SHOAIE
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

A large number of microbes with different strain types occupy the human gut. These gut inhabitant microbes have key roles in decomposition of indigestible dietary macronutrients before they are metabolised by the host. The gut microbiome has a clear effect on human health and perturbations in its diversity may lead to the development of disorders through changes in metabolic functions. To date, different studies have shown the association of the gut microbiome with metabolic disorders such as obesity, type2 diabetes and certain cancers. It has also been shown that there is a complex interaction between microbe-microbe, host-microbe and microbe-diet, and elucidation of the mechanisms behind these interactions and associations remains a challenge. Due to the extreme complexity of cellular metabolism, mathematical models may be employed for deciphering the role of its individual elements and may thereby assist in providing an increased understanding of these interactions. The emerging research field of systems biology can integrate different high-throughput data, in this case metagenomics and metatranscriptomics, through the use of mathematical models and thus provide a holistic interpretation for this complex system. In this context, genome-scale metabolic modeling has been applied to gain increased knowledge in important biotechnology applications.

This thesis presents approaches to facilitate understanding of the causalities and go beyond the association analysis by considering the interactions between microbiome, host and diet. Using genome-scale metabolic models (GEMs), we investigated the contribution of key species in the overall metabolism of the gut microbiome. We developed methods and generated stand-alone software to apply for different case studies on modeling of gut microbiome and finally addressed relevant biological questions. First, GEMs for three bacteria being representatives of dominant phyla in the human gut microbiome were reconstructed. This modeling approach allowed us to establish effective resources for understanding the microbe interactions in the gut. Increasing the number of relevant GEMs representing all key microbes in the human gut resulted in more complexity and therefore we developed the CASINO toolbox, a comprehensive software platform for the analysis of microbial communities. CASINO was validated based on *in-vitro* studies and thereafter applied to human studies that showed its capability to predict the phenotype of individuals based on their dietary pattern and gut microbes' abundances. Finally, the application of CASINO was extended and used for modeling of the interactions between gut microbiota and host metabolism. The overall metabolic differences between germ-free and conventionally raised mice were revealed through the use of CASINO. In conclusion, this thesis provides a new approach to human gut analysis by using valuable resources (GEMs) and novel methods (CASINO). As such it contributes to advancing the role of metabolic modeling in human health and designing new clinical interventions.

Keywords: Gut microbiome; genome-scale metabolic model; obesity; diabetes; gene richness; CASINO; flux balance analysis, network topology; complexity; systems biology; meta'omics

List of publications

This thesis is based on the work in the following publications:

I. **S. Shoaie**, F. Karlsson, A. Mardinoglu, S. Bordel, I. Nookaew and J. Nielsen, Understanding the interactions between bacteria in the human gut through metabolic modeling, S*cientific Reports 3, 2532 (2013).*

II. **S. Shoaie**, P. Ghaffari, P. Kovatcheva-Datchary, A. Mardinoglu, P. Sen, E. Pujos-Guillot, T. de Wouters, C. Juste, S. Rizkalla, J. Chilloux, L. Hoyles, J. K. Nicholson, MICRO-Obes Consortium, J. Dore, M. E. Dumas, K. Clement, F. Bäckhed, and J. Nielsen, Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome, *Cell Metabolism 22, 320-331 (2015).*

III. P. Kovatcheva-Datchary, **S. Shoaie**, J. Nielsen and F. Bäckhed, SIM mice – a tool to obtain basic and mechanistic understanding of the microbe-microbe interactions in the mammalian gut, *For Submission (2015).*

IV. **S. Shoaie** and J. Nielsen, Elucidating the interactions between human gut microbiota and its host through metabolic modelling, Frontiers in Genetics, 5:86, 2014.

V. A. Mardinoglu**#**, **S. Shoaie#,** M. Bergental, P. Ghaffari1, C. Zhang, E. Larsson, F. Bäckhed and J. Nielsen, The gut microbiota modulates host amino acid and glutathione metabolism in mice, *Submitted for publication (2015).*
#Authors contributed equally to this work

Additional publications not included in this thesis:

VI. L. Caspeta[#], S. **Shoaie**[#], R. Agren, I. Nookaew, and J. Nielsen, Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. *BMC Systems Biology 6, 24 (2012).*

VII. K.K. Hong, J. Hou, **S. Shoaie**, J. Nielsen, and S. Bordel, Dynamic 13C-labeling experiments prove important differences in protein turnover rate between two Saccharomyces cerevisiae strains, *FEMS yeast research, 12, 741-747 (2012).*

VIII. R. Ågren, L. Liu, **S. Shoaie**, W. Vongsangnak, I. Nookaew and J. Nielsen , The RAVEN Toolbox and its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLOS Computational Biology, D-12-01566R1 (2013).*

IX. IE. El-Semman, F. Karlsson, **S. Shoaie**, I. Nookaew, TH. Soliman and J. Nielsen, Genome-scale metabolic reconstructions of Bifidobacterium adolescentis L2-32 and Faecalibacterium prausnitzii A2-165 and their interaction, *BMC systems biology 8 (1), 41 (2014).*

X. P. Ghaffari, A. Mardinoglu, A. Asplund, **S. Shoaie**, C. Kampf, M. Uhlen and J. Nielsen, Identifying anti-growth factors for human cancer cell lines through genome-scale metabolic modeling, *5, 8183, Scientific Reports (2015).*

XI. **S. Shoaie**[#], B. Ji[#], P. Kovatcheva-Datchary, F. Bäckhed, L. Engstrand and J. Nielsen. Systematic analysis of spatial heterogeneity along human colon. *Manuscript in preparation (2015).*

XII. P. Ghaffari, A. Mardinoglu, A. Asplund, **S. Shoaie**, C. Kampf, M. Uhlen and J. Nielsen, Comparative analysis of metabolic reprograming across human cancer cell lines by genome-wide integration of transcriptomics and proteomics. *Manuscript in preparation (2015).*

XIII. A. Asplund, **S. Shoaie,** P. Ghaffari, A. Mardinoglu, C. Kampf, J. Nielsen and M. Uhlen. Characterization of metabolic differences between 2D and 3D culture of human colorectal adenocarcinoma-derived cell line. *Manuscript in preparation (2015).*

[#]*Authors equally contributed.*

Contribution Summary

I. Reconstructed the models, performed the modeling, performed the analysis, prepared and submitted the paper.

II. Formulated, implemented and participated in the development of software. Participated in analysis of data and prepared and submitted the paper.

III. Analyzed the data and assisted in preparation of the paper.

IV. Prepared and submitted the paper.

V. Performed modeling, assisted in analysis and preparation of the paper.

VI. Reconstructed the model, performed modeling, assisted in analysis of data and preparation and submission of the paper.

VII. Assisted in formulation and implementation of mathematical model. Assisted in preparation of the paper.

VIII. Participated in development of an algorithm in the software.

IX. Supervised the work, assisted in the reconstruction. Modeling, analysis and preparation of the paper.

X. Assisted in modeling and analysis of data.

XI. Performed data analysis. Prepared the paper.

XII. Assisted in modeling and analysis of data.

XIII. Assisted in modeling and analysis of data.

“He deals the cards as a meditation
And those he plays never suspect
He doesn't play for the money he wins
He don't play for respect

He deals the cards to find the answer
The sacred geometry of chance
The hidden law of a probable outcome
The numbers lead a dance”

By Sting & Dominic Miller

Preface

This dissertation is submitted for the partial fulfillment of the degree of doctor of philosophy. It is based on work carried out between 2011 and 2015 in the Systems and Synthetic Biology group, Department of Biology and Biological Engineering, Chalmers University of Technology under the supervision of Professor Jens Nielsen. The research was funded by the Knut and Alice Wallenberg Foundation, the Chalmers Foundation, European Commission FP7 project METACARDIS and the Bill & Melinda Gates Foundation.

Saeed Shoaie
September 2015

Table of Contents

List of figures and tables

Abbreviations and symbols

AA	Amino acid
BMI	Body mass index
bp	Base pairs
CASINO	Community And Systems-level INteractive Optimization
CBM	Constraint-based model/modeling
COG	Cluster of orthologous groups
DNA	Deoxyribonucleic acid
EFM	Elementary flux mode
FBA	Flux balance analysis
GEM	Genome-scale metabolic model
HDL	High-density lipoprotein
HGC	High gene count
HMP	Human microbiome project
HMR	Human metabolic reaction
HOMA-IR	Homeostatic model assessment-insulin resistance
INIT	Integrative network inference for tissues
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG orthology
LDL	Low-density lipoprotein
LGC	Low gene count
LP	Linear programming
MCA	Metabolic control analysis
MIP1b	Macrophage inflammatory protein 1b
MILP	Mixed-integer linear programming
MOMA	Minimization of metabolic adjustment
NAFLD	Non-alcoholic fatty liver disease
PCA	Principal component analysis
PCR	Polymerase chain reaction
QP	Quadratic programming
RNA	Ribonucleic acid
SCFA	Short chain fatty acid
SD	Standard deviation
T2D	Type 2 diabetes
WMS	Whole metagenome shotgun
NEFA	Non-esterified fatty acids

1. Introduction

The first evidence for the existence of microbes in the mouth and faeces of human was observed by the end of the seventeen century. From that time till now, it has been realised that the human body contains a vast number of microbes, colonised in different parts of the body such as the skin, oral cavity, gastrointestinal tract and urogenital tract. Extensive computational methods and experimental instrument developments also took place, particularly for sequence technology. Since then, several software programmes have been designed to overcome challenges to optimum genome assembly and annotation. For microbiome studies, the major goal to achieve is a comprehensive understanding of the microbial consortia, their interaction within the community, and their cross-talk with their environments. Top-down approaches such as metagenomics were to some extent made successful due to having meaningful insight into the microbial community, but were unable to assess the individual members. This is where the mathematical modeling and particularly genome-scale metabolic model as the common denominator of systems biology can be applied. The history of systems biology of single organisms shows us the effectiveness of this approach in metabolic engineering and studying human health and disease. After I worked on metabolic modeling of different organisms at the beginning of my PhD, I focused mainly on integrating the experience of single organism modeling into community modeling. Within this, I tried to address three questions in the field of the gut microbiome as an important research area for human health and diseases:

-Can metabolic modeling be applied on human gut microbiota to describe the interactions between individuals?

-Can metabolic modeling be applied to test perturbations and pinpoint the changes of individual bacteria and their metabolism?

-Can GEMs be used to better understand the interactions of microbiota with their host and changes in their metabolism?

GEMs allowed us to simulate the gut microbiome ecosystem. We obtained information about the functional role of different species in this complex ecosystem and evaluated the role of different species *in silico*. We mathematically formalised two different scenarios to understand the interactions of the bacteria in the gut ecosystem. In the first case, the composition of the diet and the species abundances in the microbiota is known and constituted the input for the model whereas in the second case, we predicted the abundances of the different species in the microbiota as a function of the diet. We further demonstrated that the integration of the topological information of GEMs with diet compositions, genomic and transcriptomic data, as well as community metabolic modeling provided a mechanistic interpretation of the statistical findings revealed by metagenomics.

The use of the GEM approach would facilitate the testing of a different hypothesis. The effects of different diets were simulated, the interactions between the microbes and host in response to the diet were studied and the contribution of each bacteria to the faecal metabolite profiling was quantified. The knowledge harvested from our simulations can be used for rational designing of prebiotics. Our approach can also be used for identifying beneficial bacteria for human health and can be used for the treatment of metabolic disorders that are associated with the gut microbiota. For instance, malnourished individuals who lack digestive enzymes due to the absence of certain microbes in their gut can be determined and appropriate probiotics can be designed specific to these

subjects. Metabolic modeling can be employed to understand the effect of the deletion or over-expression of individual genes or gene sets on human health in different clinical conditions by evaluating the abundance of species and the level of metabolites. Animal models can also be applied to examine the link between the gut microbiome and human host. These models can be a great platform for investigating different hypotheses through changing different parameters, whereas this is not practical in human studies.

To better understand the metabolic changes in the host due to microbiome activity, germ-free (GF) and conventionally-raised (CONV-R) mice are a great tool. Mice can be raised with different diets and then euthanised to get samples from the host that can be used for high-throughput analysis. Then different analyses can be performed on the data and GEMs can be applied to provide network topology and a study of network-dependent analyses, or used for predicting the diet consumption of gut microbes and the host, and their interactions.

This thesis describes the concept of metabolic modeling for simplified community. Then it defines two tools for unravelling the interactions of microbe-microbe in a more complex community and the possibility of testing different diet patterns on them. At the final step, through using GEMs and high-throughput data, the metabolic changes between CONV-R and GF mice are revealed as a platform for understanding the microbe-host interactions and their impact on different metabolic disorders.

2. Background

2.1. The human gut microbiome

Large numbers of microorganisms inhabit various sites in the human body. This collective set of microorganisms in and on the human body is generally referred to as a microbiota and their collective genome is called the microbiome. Microorganisms inhabiting our body include bacteria, eukaryotes and viruses, and can be located on the skin, in the vagina, oral cavity and intestine (Huttenhower et al., 2012). The number of somatic cells in the human body is 10^{13}, while the number of microorganisms that live in and on the human body is 10 times larger. Moreover, the gut metagenome is two orders of magnitudes greater than the human genome.

The gastrointestinal tract accommodates a large part of the human microbiota and 50% of the faecal matter consists of bacteria with an overall weight of 1.5 kg (Zhao, 2013), similar to the weight of the adult human brain. The density of cells changes in different locations of the gut, such as 10^{11}-10^{12} cells per gram of luminal content which is aggregated in the distal colon whereas this is 10^{3} to 10^{8} cells in the small intestine (Backhed et al., 2005). The bacterial taxonomic diversity varies in accordance with human gastrointestinal sites. Two major bacterial phyla, *Firmicutes* and *Bacteroides* inhabit the gut (Qin et al., 2010) . The Gram positive *Firmicutes* phylum contains different classes. Clostridia with clusters of IV (*Clostridium leptum* group), IX and XIVa (*Clostridium coccoides* group) are the most abundant classes of *Firmicutes* (Collins et al., 1994). Clusters IV, IX and XIVa include *Clostridium*, *Eubacterium*, *Roseburia* and *Ruminococcus* as the main genera, so that two butyrate-producing bacteria *Eubacterium rectale* and the *Roseburia* species belong to cluster *Clostridium* XIVa, comprising 5% to 10% of the total microbiota (Aminov et al., 2006; Scott et al., 2006). The gram negative *Bacteroides* phylum has been studied systematically together with the *Prevotella*. There are also less abundant phyla dominating the gut such as *Actinobacteria*, *Proteobacteria* and *Verrucomicobia*. The *Actinobacteria* includes two important genera, *Bifidobacterium* and *Collinsella*. *Akkermansia muciniphila* (Derrien et al., 2004), the known bacteria for mucus degradation belongs to recently found *Verrucomicobia* phyla. There are also important species belong to archaea in the gut microbiota like *Methanobrevibacter smithii*, which belongs to the *Euryachaeota* phyla (Gill et al., 2006).

The diversity of microbiota is highly variable between the subjects. Shaping of the gut microbiota composition can be driven through complex interactions between age, geography, diet, lifestyle, diseases and drug usage (Biagi et al., 2010; Claesson et al., 2012; Dethlefsen and Relman, 2011). The diversity and composition of the gut microbiota is mainly measured from faecal samples and the structures of these microbial communities change along the gastrointestinal tract. The spatial heterogeneity and co-occurrence patterns of microbiota phylum along the intestinal tract have been studied (Zhang et al., 2014). The faecal samples are still the most common sources for gaining knowledge about the gut microbiota due to difficulties in obtaining samples from other parts of the intestinal tract. Despite the extensive research on the inter-subject variation of gut microbiota, the stability of this community over time has been disregarded. However, a few studies have shown that the microbiota composition between two time points for the same subject is very similar compared to other subjects (Faith et al., 2013). These colonised microbes in the human gut have dynamic and beneficial functions for the human body. Through their symbiotic relationships, the indigestible part of diet by human cells is assimilated (Backhed et al., 2005). Moreover, the gut microbiota affects the immune system by regulating immune homeostasis and autoimmunity and maintains the stability of the immune system by providing resistance against pathogens (Wu and Wu, 2012).

2.1.1. Meta'omics of the gut microbiota

In the last two decades, these complex microbial communities have been studied through culture independent experiments. These approaches proved the diversity and functionality of these communities mainly through low-throughput sequencing, 16S rRNA gene. Exponential cost reduction of high-throughput sequencing resulted in more genome-wide approaches such as metagenomics and metatranscriptomics to explore the microbial communities. These meta'omics tools extract the molecular details of thousands of microbes and expose the host and microbe metabolism and their interactions at a systems level. Meta'omics approaches mainly aim to depict the uncultured microbial community in the form of genes, transcripts, proteins, metabolites and their variations. The metagenomics refers to DNA sequencing of the whole community, while the metatranscriptomics is their cDNA sequencing. Metagenomics and metatranscriptomics are used to gain access to genome composition and diversity in the microbial community through targeted rRNA sequencing and whole metagenome shotgun (WMS) sequencing. WMS sequencing is followed by isolation of DNA/RNA from the community. After library construction, the short-read sequencing on the mixture of genomes or transcripts is performed. These millions of short reads are either assembled for gene prediction or used for recognising the presence of organisms and their possible metabolic functions. The collection, analysis and mining of this massive data remains computationally challenging. Bioinformatics tools have also advanced following the evolution of sequencing technology. All generated pipelines start with filtering a low quality sequence and the removal of contaminating genomes, and this is then followed by sequence assembly, gene prediction and species diversity. A catalogue of reference genes plays a key role in the analysis of the human gut microbiome. However, most gut metagenome studies have been done based on single cohorts or reference genomes, which causes limitations for functional metagenomics analyses (Huttenhower et al., 2012; Qin et al., 2010). Recently, an integrated catalogue of reference genes in the human gut microbiome has been constructed to overcome this issue (Li et al., 2014). This catalogue is established by using published studies from MetHIT (760 European samples), HMP (139 American samples), a Chinese cohort (368 samples) as well as the genome of gut-related or already sequenced bacteria and archaea. This has resulted in a non-redundant reference catalogue, comprising 9,879,896 genes. Besides the metagenomics analysis, the integrated catalogue also enables the quantitative characterisation other meta'omics data. Although metagenomics is becoming the common tool and study for the human gut microbiome, metatranscriptomics has become more common for profiling the expressed functions and regulations in the community (Ridaura et al., 2013). Since isolating the microbial community with high quality and quantity is challenging, there are few gene expression analyses of microbial communities in association with the host (Byrd and Segre, 2015). These difficulties in isolation are a result of few microbial mRNA in total microbial RNA as well as the presence of host nucleotides. In the later step, the metagenomics data are used for functional annotation and metabolic model reconstruction. The easy way is to align the reads with gene catalogues and directly determine the functions. The functions can also be inferred through aligning reads to assembled contigs or genes. This later step leads to the *de novo* assembly of metagenomics reads. The function of metagenomic genes are linked to databases such as KEGG (Kanehisa et al., 2004), NCBI (www.ncbi.nlm.nih.gov) and COG (Tatusov et al., 2003). The KEGG database is composed of many microbial genomes with links between their genes and functions. Since the majority focus of KEGG is on metabolism, it makes it a great tool for metabolic model reconstruction.

2.1.2. Metabolism by the gut microbiota

The colon was mainly assumed as the loading space for undigested food, but now different microbial activities based on diet consumption and metabolic formation have been studied in this organ. Amino acids and xenobiotic metabolism, vitamin biosynthesis, microbial regulation of bile-acid metabolism and microbial metabolism of choline are some of the examples of microbial metabolic activities (Gill et al., 2006).

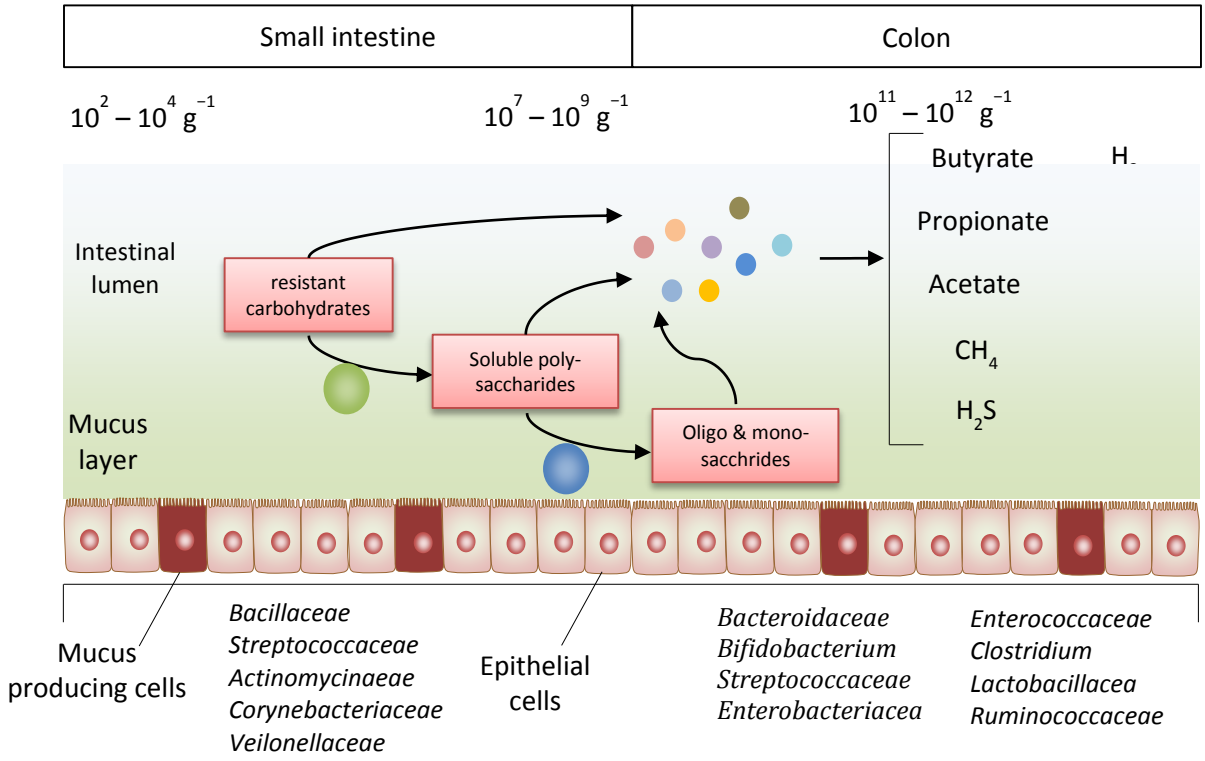


Figure 1 Schematic representation of variations of gut microbiota composition and concentration along the intestinal tract. As well the conversion of dietary carbohydrates through microbes' anaerobic fermentation and syntrophy between the carbohydrate degraders and other gut microbial members. Adapted from (Kovatcheva-Datchary et al., 2013).

Various complex carbohydrates including resistance starch, fibres with plant origin, non-starch polysaccharides and host driven glycans are not digestible by host enzymes (Backhed et al., 2005). These non-digestible foods are not digested in the small intestine and moved over the large intestine with a complex microbial community (Ouwehand et al., 2005). In the presence of microbes, these foods are anaerobically fermented. This degradation of various complex compounds requires the collection of hydrolytic enzymes that are provided by the microbial community. The carbohydrate digestion occurs through metabolic cross-feeding and different levels of conversion in microbiota. Capturing this information about which microbe is involved in the degradation of specific carbohydrates is missing and difficult to extract from metagenomics data. It is known that the indigestible dietary carbohydrates are degraded into many poly-saccharides and provide a substrate for the colonic microbiota (Rossi et al., 2005; Samuel and Gordon, 2006). The conversion of poly-saccharides to oligo-saccharide occurs via different classes of microbial activities which many bacteria utilised and depend on. There are recent studies, which show that the first level of

breakdown of complex indigestible carbohydrates is established with a few microbes, such as species in the *Bacteroides* and *Ruminococcus* genus (Flint et al., 2008; Jindou et al., 2008).

The main products from the microbial fermentation of these substrates are different gases including hydrogen, carbon dioxide and methane. Moreover, short chain fatty acids (SCFA) (butyrate, propionate and acetate) are the end products of this fermentation and they have significant impacts on host physiology. The concentration of SCFAs varies between different sites in the colon and SCFA measurements showed that their concentration drops from proximal to distal since a significant portion of SCFAs are absorbed by the proximal and 5% to 10% is transferred to the end part of colon (Cummings and Englyst, 1987; Macfarlane et al., 1992; Wong et al., 2006). SCFAs have different benefits for the human body by decreasing the luminal PH, increasing the microbial biomass, the beneficial bacteria population as well as mucus production and eventually shaping the peripheral metabolism.

Among the SCFAs, butyrate has a key role as an energy source for colonic epithelium which is the 5% to 10% of total energy needed by a healthy human body (Donohoe et al., 2011). Studies have shown that absorbed butyrate inhibits histone deacetylase, which stimulates the proliferation of colorectal cancer cells (Davie, 2003; Hamer et al., 2008). Different metabolic pathways can synthesise butyrate. Major butyrate producing bacteria (e.g. Clostridium cluster XIVa) ferment SCFAs through a CoA-transferase pathway whereas *Eubacterium Hallii* utilise lactate (Duncan et al., 2004; Louis et al., 2004). Propionate is a preferred substrate for liver gluconeogenesis and regulation of cholesterol synthesis and it has a protective role against hepatic cancer and relevant cancers that metastasise in the liver (Chambers et al., 2002; Comalada et al., 2006). The fermentation of propionate occurs in three routes; succinate decarboxylase, acrylate and propanediol pathways. The relevant examples are species from *Bacteroides*, Clostridium cluster IX and *Roseburia inulinivorans.* Acetate is a required substrate for liver lipogenesis and cholesterol synthesis while it stimulates the colonic blood flow and also has a protective role against hepatic cancer (Chambers et al., 2002; Scheppach, 1994). In addition, metabolites like succinate and lactate are produced by microbial fermentation. The list of important metabolites and contributed gut microbes within their functions were summarised in table 1.

Moving from the proximal colon towards the distal colon, the level of dietary carbohydrates is decreased and protein becomes the available source for microbes. Hence, the human colon has been reported as the place for protein turnover (Macfarlane et al., 1986). Possibly due to the existence of branched chain fatty acids, e.g. isobutyrate, isovalerate and 2-methylbutyrate, and other toxic compounds as one of the end products of protein fermentation in the distal section, the risk for the occurrence of colon cancer is increased (Bingham et al., 1996; Cummings et al., 1979). It has been shown than colon cancer development has been positively correlated to protein fermentation and the distal part of the colon has a severe impact on cancer progression (Muir et al., 2004). The hydrogen gas as one of the microbial fermentation is diminished through utilisation of colonic methanogens. By this process, the hydrogen gas is oxidised and methane gas is released. *Methanbrevibacter smithii* is the well-known archaeon for such conversion in the human microbiota and it has a key role in overall gut metabolism despite its low abundance (Gill et al., 2006; Salonen et al., 2010). The left over hydrogen gas is absorbed into the blood and finally excreted from the body through the lungs. Moreover, in those cases with availability of sulphate and mucins in the lumen, the hydrogen gas can be oxidised to hydrogen sulphite through sulphate-reducing bacteria (Christl et al., 1992; Gibson et al., 1988).

2.1.3. Factors affecting the gut microbiota composition

The composition of gut microbiota may alter due to different factors, such as dietary pattern, environmental factors, age, drugs and antibiotics as well as host genetics. The host is one of the influencing factors that can have a strong impact on the distribution of the microbiota along the gastrointestinal (GI) tract. This occurs through changes in PH, host-secreted metabolites and other physiochemical parameters. Hence, the densities of bacterial cells in the stomach, small intestine and distal part of the colon can vary from 10^4 (per ml of digesta) to 10^8 to 10^{12} (per gram of stool), respectively (Figure 1) . The relative low abundances of microbes in the upper part of the GI can affect the host metabolism via acid secretion in the stomach, pancreatic enzymes and bile acid secretion in the small intestine. Moreover, different distributions of microbiotas have been correlated with different compartments along the GI tract. This observation can be associated with the epithelial cells covered by mucus and nutrient availability (Eckburg et al., 2005; Frank et al., 2007). While in the stomach the most abundant classes of bacteria are *Lactobacillacea, Veilonellaceae* and *Helicobacterceae,* the PH varies between 1 and 4.4. In the small intestine, the composition changes to *Bacillaceae* and *Streptococcaceae.* In the large intestinal lumen, the abundant classes are *Bacteroidaceae, Clostridium* and *Bifidobacterium,* while in the epithelial surface of the large intestine, these are *Clostridium, Lactobacillaceae, Enterococcaceae.* The PH in the intestines varies from 5.5 to 7 (Figure 1).

The other factor that may play a key role in changing the composition of gut microbiotas is the diet. The substrates available for microbial fermentation in the human colon can be categorised as dietary protein, resistance starch, non-starch polysaccharides, unabsorbed sugars and oligosaccharides. Beside these, bile, urea, peptides and mucus are the other available substrates for the microbial fermentation (Hughes et al., 2000; Scott et al., 2013). Around 10 to 60 grams of dietary carbohydrates reach the colon on a daily basis, which are mainly fermented in the proximal location. The amount of dietary proteins reaching the colon varies between 12 to 18 g/day. Figure 1 shows the different substrate availability for human colonic fermentation. This availability of nutrients in different regions influences the distribution of the microbiota along the colon. This has been studied in colonoscopy of colonic mucosa for 11 healthy subjects (Zhang et al., 2014). The samples have been taken from seven regions of the colon where 16S rRNA gene sequencing showed the composition of the bacterial phyla changes along the colon. In another case, the metagenomics data have shown the positive correlation between long-term diet and microbiota composition. The metagenomics data generated for faecal samples of American adults have been clearly clustered together with the metagenomics data of European faecal samples (De Filippo et al., 2010). The dietary pattern for both groups is enriched in protein. In a similar case, metagenomics data of Malawians and Africans have been clustered together, while it is known that their diet is rich in plant polysaccharides (Yatsunenko et al., 2012). The seeded microbiota in infants is significantly dependent on feeding. The predominant bacteria in breast-fed infants are *Bifidobacteria*, *Bacteroides*, *Clostridia* and facultative anaerobes, while the formula-fed infants have a higher abundance of pathogenic species such as *Clostridium difficile(Wall et al., 2009).* Before the age of 2 years, the composition of the gut microbiota is simple. At the first few months, *Bifidobacteria*, *E.coli*, *Lactobacillus* species and *Bacteroidetes* are colonised in the gut. After one year of age, *Bacteroidetes* and *Lactobacillus* remain abundant ones together with clostridia, while the abundance of *Bifidobacteria* and *E.coli* drops. After 2 years of life, the composition of the gut microbiota is quite similar to adults (Koenig et al., 2011).

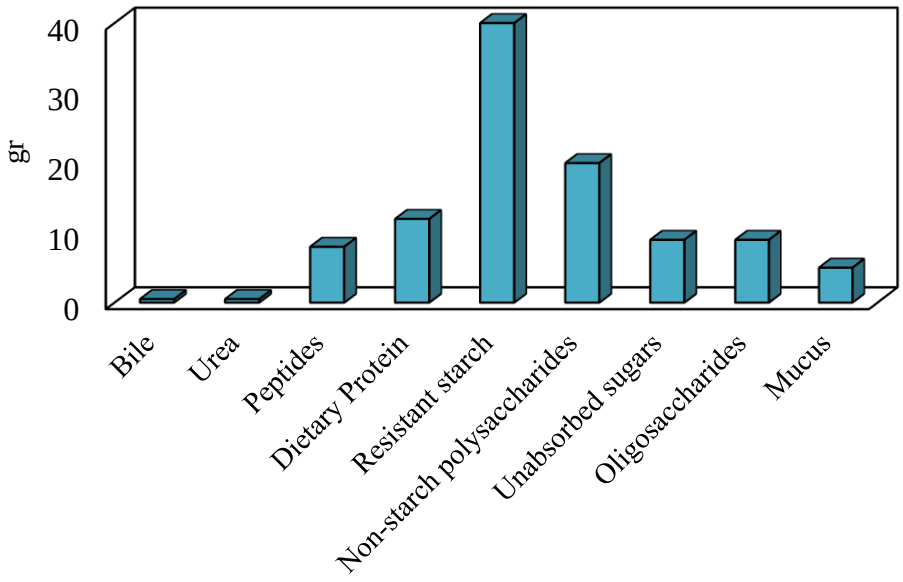


Figure 2 The maximum availability of different substrates for human colonic fermentation. The amount for peptides, dietary protein resistant starch, non-starch polysaccharides, unabsorbed sugars and oligosaccharides can vary between 4-8, 1-12, 5-40, 10-20, 1-9 and 2-9 gr, respectively. Adapted from (Kovatcheva-Datchary et al., 2013).

2.1.4. The human host-gut microbiota interactions and association with diseases

Increased numbers of studies on host-microbe interactions indicate the significant impact of microbiota composition on host metabolism. Particularly, the host and gut microbiota symbiosis results in the exchange of small bioactive molecules between each other which have influence on health and disease states of the host. Along the GI tract, different molecules can be secreted by the microbes or host cells/tissues. There are different metabolites that regulate the host-microbiota interactions, which SCFAs is the well-studied ones. Butyrate, besides being the energy source, has been shown as energy homeostasis triggering leptin secretion in adipose cells (Xiong et al., 2004). Moreover, SCFAs decrease the colonic PH, prevent the growth of pathogens and have an effect on the occurrence of insulin resistance and obesity (Samuel et al., 2008; Scheppach, 1994; Wong et al., 2006). The production of SCFAs in the human gut microbiota has been attributed to Firmicutes phyla, including *Clostridia clusters IV* and *XIVa* as well as the *Eubacterium, Roseburia, Faecalibacterium and Coprococcus* species. Table 1 summarise the metabolites, contributed microbes and their role in health and disease.

The other metabolite that is utilised by both gut microbes and the liver is Choline, which should be absorbed from the diet. Choline is metabolised to trimethylamine by gut microbes and further to trimethylamine-N-oxide, and it may have a potential pathological role in the progression of atherosclerosis (Dumas et al., 2006). Moreover, choline stimulates the lipid metabolism and glucose homeostasis and may contribute to the progression of non-alcoholic fatty liver disease and diabetes (Dumas et al., 2006; Wang et al., 2011). The related bacteria in the metabolism of choline are *Bifidobacterium Spp.* and *Faecalibacterium prausnitzii*. The human liver produces bile acids, a type of steroid acid produced from cholesterol-derived precursor molecules. Bile acids assist in dietary lipid metabolism and the absorption of fat-soluble vitamins and cholesterol. More than 90%

to 95% of released bile acids in the small intestine are absorbed into the distal ileum and transported back to the liver. Humans can have 4 to 12 of this circulation which is called enterohepatic (Dawson et al., 2009). The rest of the bile acids that escape from intestinal absorption are transferred to the colon and metabolised by the microbiota. The main bacterial metabolism of bile acids is deconjugation and conversion of primary to secondary bile acids such as deoxycholate and lithocholate (Groh et al., 1993; Ridlon et al., 2006). Then the secondary bile acids are reabsorbed into the liver while a high level of secondary bile acids is one of the colon cancer markers (Ridlon et al., 2006). The gut bacteria involved in bile acid metabolism are mainly from the *Lactobacillus*, *Bifidobacteria*, *Enterobacter*, *Bacteroides* and *Clostridium* species.

Table 1 Relevant contribution of gut microbiota, metabolites and their potential biological functions

Metabolites	Species	Microbial group	Metabolic functions	Disease association	Ref.
Acetate	*Bacteroidetes spp.*	*Bacteroidetes*	Polysaccharide breakdown	Cardiovascular disease,	(Samuel et al., 2008; Scheppach, 1994; Wong et al., 2006)
	Prevotella spp.		Polysaccharide breakdown	Obesity,	
	B. thetaiotaomicron		Polysaccharide breakdown, mucin degradation	type 2 diabetes	
	Bifidobacterium adolescentis	*Actinobacteria*	Carbohydrate metabolism		
	B. longum		Carbohydrate metabolism		
	Collinsella spp.		Carbohydrate metabolism		
	Ruminococcus bromii	*Clostridium cluster IV*	Carbohydrate metabolism		
	R. flavefaciens		Plant fibre breakdown		
	Victivallis vadenis	*Verrucomicrobia*	Cellobiose degradation		
	Akkermansia muciniphila	*Verrucomicrobia*	Mucin degradation		
Butyrate	*F. prausnitzii*	*Clostridium cluster IV*	Carbohydrate metabolism	Colorectal cancer	
	R. inulinovarans	*Clostridium cluster*	Carbohydrate metabolism	type 2 diabetes	
	E. hallii		Carbohydrate metabolism, Acetate utiliser	Obesity	
	A. caccae		Carbohydrate metabolism, Acetate utiliser		
	E. rectale		Carbohydrate metabolism, Acetate utiliser		
Propionate	*B. thetaiotaomicron*	*Bacteroidetes*	Polysaccharides breakdown, mucin degradation	Cardiovascular disease,	
	R. inulinovarans	*Clostridium cluster*	Carbohydrate metabolism	Obesity,	
	Megasphaera elsdenii	*Clostridium cluster IX*	Gluconic acid metabolism, Acetate, lactate utiliser	type 2 diabetes	
	Akkermansia muciniphila	*Verrucomicrobia*	Mucin degradation		
Choline	*Faecalibacterium prausnitzii**	*Clostridium cluster IV*	Carbohydrate metabolism	non-alcoholic fatty liver, Obesity	

	*Bifidobacterium dentium**	*Actinobacteria*	Protein metabolism	diabetes, cardiovascular disease	(Martin et al., 2010; Wang et al., 2011)
Bile acids	*Bacteroides fragilis*	*Bacteroidetes*	Polysaccharides breakdown	Colon cancer marker	(Groh et al., 1993; Ridlon et al., 2006; Swann et al., 2011)
	Bacteroides vulgatus		Polysaccharides breakdown	Atherosclerotic coronary artery	
	Clostridium perfringens	*Clostridium cluster*		Diabetes	
	Listeria monocytogenes	*Bacilli*			
	Peptostreptococcus productus	*Clostridia*			
	Eggerthella lenta.	*Actinobacteria*			
Vitamins	*Bifidobacterium dentium*	*Actinobacteria*			(Koenig et al., 2011; Said, 2011)
	Lactobacillus rossiae	*Lactobacillus*			
Polyamines	*Campylobacter jejuni*	*Proteobacteria*	Carbohydrate metabolism	Potential tumour marker	(Hanfrey et al., 2011; Matsumoto and Benno, 2007)
	Clostridium saccharolyticum	*Clostridium cluster*			
H_2S	*Desulfovibrio piger*	*Proteobacteria*	sulphate reducing bacteria, Lactate utiliser	Colorectal cancer	(Lakhan and Kirchgessner, 2010)
	Desulfovibrio desulfuricans		sulphate reducing bacteria, Lactate utiliser	chronic inflammation	
CH_4	*Methanobrevibacter smithii*	*Archaea*	H2 utiliser	Bloating, abdominal pain	
H2	*E. rectale*	*Clostridium cluster*	Carbohydrate metabolism, Acetate utiliser	Bloating, flatulence, diarrhoea	

*The choline acts as a substrate for these species.

2.2 Systems biology and metabolic modeling

Part of the work described in this thesis is highly connected to the field of systems biology, where a holistic view of the gut microbiota and host metabolism was used to integrate high-throughput (HT) data. The rise and spread of systems biology is to a great extent due to the rapid development of HT technologies. Seen from one aspect of systems biology, this approach allows us to study the interaction of the biological components on a large-scale. The central dogma for this interpretation is that of *emergent properties.* This means the process of larger entities arising through the interaction of smaller entities where these don't have similar properties. A classic example can be the interactions between macromolecular biological catalysts and small biological molecules that make up the series of biochemical reactions, called metabolic pathways.
The other way to interpret the term 'systems biology' is by looking at this as a paradigm. This approach is normally in contradiction to classical reductionist paradigms. The reductionist view as Sauer et al. described has perceived limitations (Sauer et al., 2007). Although this approach can detect the components and their interactions, it is incapable of present definite methods for elucidating the emergent properties within the system.

Both interpretation of systems biology either as a field or a scientific paradigm require the measurement of multiple components simultaneously. Omics techniques like transcriptomics, metabolomics and proteomics fulfil this requirement. On the other hand, these data need to be integrated with mathematical models, which make the systems biology field reliant on bioinformatics and computational biology. As Kitano *et al*. (Kitano, 2002a, b) described, systems biology can be considered as a cycle. It is initiated with the biological information and different data sets, followed with computational platforms to integrate these data where ended up to hypothesis-driven modeling. These hypotheses are analysed within the system and formed as a prediction to set up the right experimental design. After all these experiments are performed in the lab for predictions, validation is performed and knowledge is extracted from the new experimental data, which itself is then fed into this process loop (Kitano, 2002a, b). In this section, the elements of success for the systems biology of individual microorganisms were described following the current efforts towards outlining the systems biology of the community, in this case gut microbiota.

The current status of our understanding of systems biology of individual microorganisms goes back to the first whole genome sequencing of bacteria (Fleischmann et al., 1995). This facilitates the development of a computational framework to determine the phenotypes of single organisms from genotypes which rely on the molecular mechanisms and biochemical information of targeted organisms on the genome scale. This mathematical framework allows testing of different hypotheses about the phenotypic status of the biological system. One of the *in-silico* models that have this capability to link the genotype to phenotype is genome-scale metabolic model (GEM). In the next section, GEM, as one of the common denominators of systems biology, and its application in constrained-based modeling is described.

2.2.1. Constrained-based modeling

In 1913, the famous Michaelis and Menten equation for describing enzyme kinetics led to the application of mathematical modeling in the field of metabolism. This made it possible to set up small models to explain basic metabolism of a cell through determining kinetic parameters of a sufficient number of cells (Othmer, 1976). Later, another platform called metabolic control analysis was developed to measure the control of enzymes in a certain model (Heinrich et al., 1977). However, due to limitations on the availability of kinetic parameters, mathematical methods that

rely on steady state metabolic fluxes got attention. A method called metabolic flux analysis has been developed that uses the exchange fluxes of metabolites and determines the internal fluxes by linear regression (Aiba and Matsuoka, 1979). This approach had the limitation that sufficient exchange fluxes had to be measured to have a determined model. Through the years, additional constraints were added such as reaction reversibility to decrease the degree of freedom, but the breakthrough was the optimality constraint of the model (Fell and Small, 1986). Optimising some sort of cellular objective became the starting point of constraint-based modeling (CBM). CBM on the metabolic modeling imposes detailed constraints on the solution space to direct the feasibility towards relevant phenotypes. For metabolic modeling the CBM is highly identical to flux balance analysis (FBA), which is a less comprehensive term.
To understand better the general constraints in the metabolic network, the summary of these is explained. Over a metabolite (x_i) a mass balance is given by:

$$(\frac{dx_i}{dt}) = \nu_{out,i} + \nu_{in,i} - \nu_{generation,i} - \nu_{consumption,i} - \mu x_i$$

The μx_i is related to concentration reduction in relation to cell expansion and the value is neglected over the fluxes. By considering all the rates around the metabolites in a vector for each reaction and a matrix containing stoichiometric coefficients which describe the metabolic model, the end equation can be written as:

$$\frac{dx}{dt} = S \cdot \nu$$

Considering that the change of internal metabolites over time is quicker than growth, it can be assumed the change in internal metabolite concentration ($\frac{dx}{dt}$) is zero, meaning they are in a steady state (Varma and Palsson, 1994a). Converting the network to a matrix makes it possible to use in computational platforms. Thus up to now, the metabolic network is constrained by the metabolite connectivity and steady state assumption. CBM rather than optimum solution gives back a solution space. The challenge behind CBM is to introduce the right constraints in order to reduce the solution space to a more biologically relevant space. Thermodynamics can also impose another constraint by identifying the reversibility of chemical reactions through calculation around Gibbs free energy (Bianucci et al., 1995). By measuring some fluxes, another constraint can be imposed by the upper bond and lower bond of individual reaction fluxes. Meanwhile, many efforts are being made to impose different constraints to shrink the solution space even more towards physiologically relevant phenotypes, such as determining diffusion rates, enzyme capability and activity (Covert et al., 2001) and thermodynamic feasibility (Beard et al., 2002). After parametrisation of the model, the flux distribution can be calculated. However, the essential component in the FBA is maximising or minimising an objective, where it is a linear combination of reactions as the product of flux and objective coefficient. This objective, even in very large problems, can be solved through linear programming where only one optimal solution is found (Karp, 2008). There are many attempts to define the good objective function, but maximising growth for microbial cells where their biomass compositions are determinable is still by far the standard one (Varma and Palsson, 1994b). To date, a number of different objective functions have been implemented in FBA, such as minimisation of substrate consumption (Oliveira et al., 2005; Ramakrishna et al., 2001), maximisation of ATP yield (Vangulik and Heijnen, 1995), minimisation of ATP production or redox potential (Knorr et al., 2007). Combination of multiple objective

functions has also been implemented for optimality of microbial metabolism (Schuetz et al., 2012). Moreover, different unbiased approaches such as sampling of the solution space and HT data integration are used to go beyond the assumption of an objective function.

2.2.2. Reconstruction of GEMs

The availability of whole genome sequencing and physiological experiments enabled the illustration of metabolic genes in the assessment of phenotype behaviour of target organism through genome-scale metabolic modeling. GEMs reconstruction can be considered as an iterative process. To begin with, genomic and bibliomic data for the organism of interest is extracted. These data are integrated to determine the enzyme substrates and reaction stoichiometry, metabolite chemical formulae, reaction directionality and other necessary information, involve the spontaneous reactions and define biomass composition in the case of microbes. Then, this set up is tested in a controlling pipeline to check the quality of the model. After passing the quality control, a functional GEM is reconstructed, otherwise the model goes through the data integration and evaluation. There have been different reviews and protocols published for the detailed description of GEM reconstruction (Thiele and Palsson, 2010), where the process has been summarised in 96 steps with practical information. This process can be very time consuming and a laborious task, which can take anywhere from months to a year. However, there have been a number of methods and toolboxes to automate some parts of the reconstruction process, such as Model Seed (Henry et al., 2010), AUTOGRAPH (Notebaart et al., 2006), IdentiCS (Sun and Zeng, 2004), GEM system (Arakawa et al., 2006) and the RAVEN toolbox (Agren et al., 2013).

2.2.3. Application of GEMs in CBM

GEMs are a great platform for understanding and assessing phenotypic functions of a target organism and through this, to identify new targets for different fields of interest like metabolic engineering or drug discovery. Different methods and algorithms have been published based on CBM using GEMs and they have been extensively reviewed by *Lewis et al.* and *Zomorrodi et al.* (Lewis et al., 2012; Zomorrodi et al., 2012). Most of these methods use optimisation either for predictions and explore metabolic capabilities or model improvement and reconcile inconsistencies. These methods rely on one of the following optimisation frameworks: **Linear programming** (LP) requires a linear objective function and finds one optimum solution. LP can find an optimum solution even for large models efficiently. **Quadratic programming** (QP) makes it possible to have a quadratic term in an objective function although the optimisation formulation is similar to LP. **Mixed-integer linear programming** (MILP) is founded on LP and when some or all the variables are constrained to integer values**. Bi-level optimisation** refers to when one problem is nested in another one and so two level optimisation is needed. This programming optimises the objective in respect that another cellular objective should be optimised. **Heuristic optimisation** finds the approximate solution in a short time frame. This method trades away the exact solution, but is useful when classical methods fail to find one or they are computationally expensive. Although heuristic methods can produce a result alone, sometimes they are also used together with other optimisation methods for greater efficiency. Examples of the algorithm for CBM using GEMs together with their functionality and the method in use were summarised as examples in Table 2.

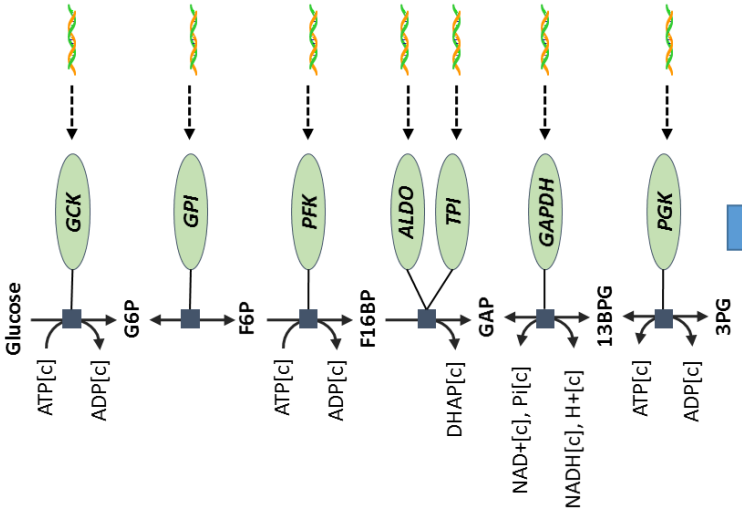
Table 2 Examples of optimisation algorithms to explore and redesign metabolic models. Adapted from (Zomorrodi et al., 2012)

Name	Description	Optimisation	Ref.
MOMA	Minimisation of metabolic adjustment in response to genetic modifications	QP, MILP	(Segre et al., 2002)
ROOM	Regulatory on/off optimisation for strain design	MILP	(Shlomi et al., 2005)
ObjFind	Determining the consistency of obj.fun. With experimental data	LP	(Burgard and Maranas, 2003)
OptKnock	Gene deletion strategies	Bi-level	
OptGene	Knockouts for *in-silico* metabolic engineering	QP, LP	(Patil et al., 2005)
OptStrain	Addition/deletion of reactions to redesign microbial production system	Bi-level, MILP	(Pharkya et al., 2004)
OptReg	identifying reaction activation/inhibition/elimination for overproduction	Bi-level, MILP	(Pharkya and Maranas, 2006)
OptCom	A Multi-Level Optimisation for modeling microbial communities	Bi-level	(Zomorrodi and Maranas, 2012)
INIT	GEM reconstruction for cell/cancer type based on transcriptome & proteome data	MILP	(Agren et al., 2012)
GDLS	Knockouts, upregulations/downregulations	Heuristic	(Lun et al., 2009)

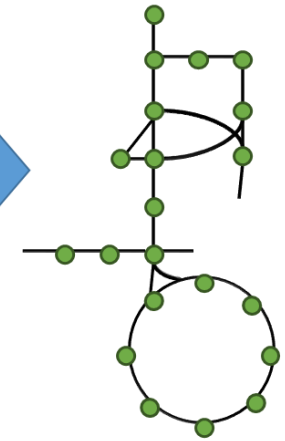
2.2.4. Application of GEMs in data integration

GEMs are also widely used for integration, interpretation and contextualising omics data. Different algorithms have been developed and extensively reviewed (Durot et al., 2009). Since describing all the methods and details is outside the scope of this thesis, a few examples of omics data using GEMs will be briefly illustrated. Fluxomics are one of the omics that are directly applicable for integration with GEMs. Methods like ^{13}C labelling substrates measure intracellular fluxes through atom mapping model, but they can build a small model and in most cases, the central carbon metabolism. It is therefore seen that some important metabolisms are not covered, but GEMs help to expand the atomic model and elucidate the metabolic flux for large-scale models (Suthers et al., 2007). There is also an optimisation framework, OptMeas, that can identify the optimal measurement sets to complete the flux elucidation metabolic flux analysis experiments through integer linear programming (Chang et al., 2008). Another algorithm, ObjFind, uses fluxomics data to determine the objective function that brings about the right phenotype (Burgard and Maranas, 2003). Metabolomics refers to large-scale quantification of internal metabolites that can provide an instantaneous snapshot of the cell physiology. Concentration of metabolites is not directly applicable to GEMs, but is very much in use for model improvement through finding missing/misleading functions (Kummel et al., 2006; Oh et al., 2007). An algorithm, (Integrative omics analysis) IOMA, uses the quantitative proteomics and metabolomics together with GEMs to more accurately predict the flux distribution (Yizhak et al., 2010). Transcriptomics are generated widely in different studies. Due to the multiple layers of regulation between transcripts level and fluxes, it is therefore challenging to change the model constraints. Nevertheless, the expression levels have been used in a binary mode to decide which gene is highly or lowly expressed. A method called Gene Inactivity Moderated by Metabolism

Reconstruction of GEMs

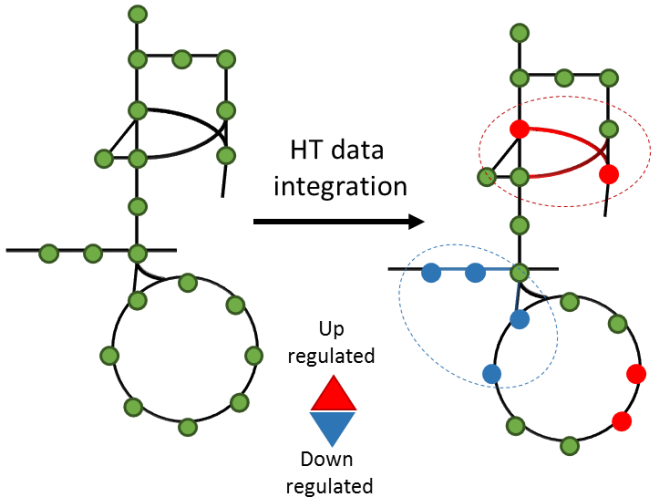


	R_1	R_2	R_3	R_4	R_5	R_6
Glucose	-1	0	0	0	0	0
G6P	1	-1	0	0	0	0
F6P	0	1	-1	0	0	0
F16BP	0	0	1	-1	0	0
GAP	0	0	0	1	-1	0
13BPG	0	0	0	0	0	-1
3PG	0	0	0	0	0	1
DHAP	0	0	0	1	0	0
ATP	-1	0	-1	0	0	-1
ADP	1	0	1	0	0	1
NAD+	0	0	0	0	-1	0
NADH	0	0	0	0	1	0
Pi	0	0	0	0	-1	0
H+	0	0	0	0	1	0

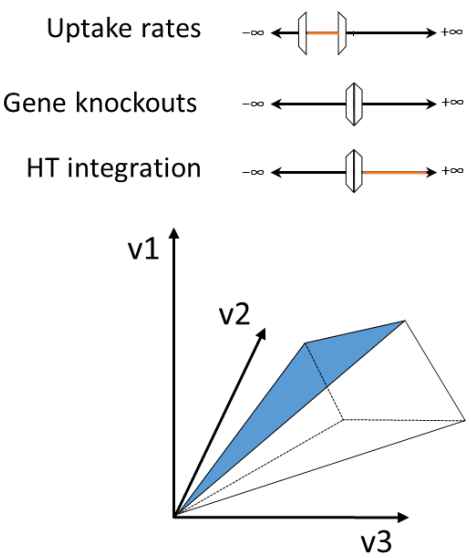


Applications of GEMs

Network dependent Analysis



Constrained-based modeling



Cell type specific modeling

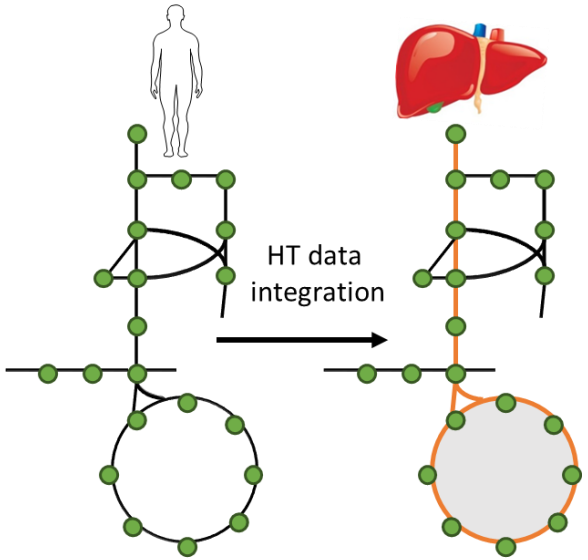


Figure 3 The layout of GEM reconstruction and examples of GEM applications. These are metabolic reactions and metabolites associated with enzymes. These enzymes are linked to their transcripts, which are linked to corresponding genes. Through the stoichiometric coefficients the S matrix is constructed, which defines the metabolic network. GEMs can have different applications. One simple application uses the network dependent analysis through HT data integration. CBM is another method that uses GEMs to predict the cellular phenotype. The generic GEMs can also be used to build cell-specific models through multi-omics integration.

and Expression (GIMME) uses gene expression together with a presumed biological objective to generate a specific model through minimising the utilisation of inactive reactions associated with low expressed genes (Becker and Palsson, 2008). Another algorithm called Integrative Network Inference for Tissues (INIT) uses transcriptomics or proteomics data to maximise the utilisation of reactions associated with an abundance of proteins/genes. In this algorithm, all the reactions should carry a flux and all experimentally observed metabolites should be produced (Agren et al., 2012). Another way of using expression data is to analyse it based on the network topology and significance of expressed genes between two conditions (Patil and Nielsen, 2005). Based on links between metabolites and genes through their associated reactions, a bipartite graph can be generated from the metabolic model. Using this graph and expression, data a meta-analysis can be performed to see if the genes that were associated with a metabolite are deferentially expressed or not. If yes, this metabolite is called a reporter metabolite. Almost exactly the same concept has used to introduce the reporter subnetworks. This time a unipartite enzyme interaction graph is generated from the model and the heuristic method is used to find the set of connected enzymes that are associated with overall significant change in expression. A very useful algorithm has been generated to reveal transcriptional regulation in enzymes by using flux distribution and gene expression data (Bordel et al., 2010). This method uses some external fluxes as the constraints and it finds a set of flux distributions by sampling the solution space with average and standard deviation of each flux. These values between two conditions show the significant changes for each reaction and it makes it comparable with significant changes in gene transcription of associated enzymes.

GEMs together with omics data have been integrated to reconstruct cell/tissue specific GEMs. Different global metabolic models have been generated for human (Duarte et al., 2007; Ma et al., 2007; Mardinoglu et al., 2014). Since metabolism changes from each cell/tissue to anotherand therefore their physiological states, it is necessary to look at them in specific GEMs. Algorithms have therefore been developed to reconstruct such models. The tINIT (Task-driven Integrative Network Inference for Tissues) algorithm is a method that uses protein evidence and set of task-driven to reconstruct the cell/tissue GEMs (Agren et al., 2014). These specific GEMs are a great scaffold for identifying metabolic changes and molecular mechanisms in response to disease progression and therefore facilitate the development of efficient treatments and discovery of drug targets.

2.3. Community systems biology

In most microbiome studies, top-down methods are applied such as 16S rRNA sequencing or metagenomics, which I have previously described in this thesis. These methods provide valuable details about the community but are incapable of identifying some key information. For example, ribosomal RNA genes are used to assess the community phylogenetically but this method is unable to capture the metabolic capacities of the community. Even a metagenomics analysis with all the results from an intact community often misses the information regarding single cell/species. Recently, a computational pipeline has been generated to elucidate the strain-level copy-number variation across human gut microbiome species from metagenomics data (Greenblum et al., 2015).

Extensive understanding of community structures and dynamics is captured through the use of the computational methods, different correlation and association analysis which can be performed between the community members (Chaffron et al., 2010). Nevertheless, even close correlation cannot delineate the causality. It is therefore necessary to apply other approaches to describe the inner workings of the microbial community. In this context, bottom-up methods may aid in elucidating the metabolic interactions between the members of the community in simple synthetic communities or animal models. Another powerful alternative is the use of metabolic modeling to understand the mechanisms of interactions. These methods are briefly described below.

2.3.1. *In-vivo* based methods to study the community

In-vivo models are great scaffolds for interpreting the link between human gut microbiome and human phenotypes by studying the host-microbiome interplay and manipulating different parameters while it is not practicable in human studies. Different model organisms have been considered to study the microbiome such as drosophila, zebrafish and mice (Chandler et al., 2011; Mahowald et al., 2009; Mouse Genome Sequencing et al., 2002; Roeselers et al., 2011) . Among all of these, mice have shown similarity to the human gut system in phylum through family levels (Mouse Genome Sequencing et al., 2002; Spor et al., 2011). Practically, when the mice are sterilised from exposure to any microorganisms, these mice are called germ-free (GF) mice and when they are colonised with microorganisms, they are called gnotobiotic mice. These types of animals provide great opportunities to govern variables like host genotype, diet and microbial ecosystem. Different combinations of microbial community have been colonised into GF mice to identify the gut microbe-microbe and microbe-host phenotype relationships (Faith et al., 2014). In one study, the response of diet on a colonised mouse with ten sequence human gut microbiota was evaluated (Faith et al., 2011). A statistical model has been developed accordingly to predict the variation of species abundance in response to different diets.

Even though the distal gut microbiota of humans and mice have the same phyla distribution, they are dissimilar in genera and species levels. The human faecal microbiota have thus also been transplanted to the GF mice to have more applicable models (Turnbaugh et al., 2009b). In one study, this humanised mouse gut microbiota from discordant twins showed the obesity modulate metabolism through gut bacteria and the diet dependency on the bacteria interactions (Ridaura et al., 2013).

Despite all the advantages of mouse models in gut microbiota studies, there are certain limitations that should be considered. Some of these limitations are: 1) Variety of factors that shape the human gut microbiota throughout life is absent in mouse models. 2) Different strains of mice can develop divergent gut microbiota composition. 3) Incapability of mice to recapitulate the human inherent genetic variations.

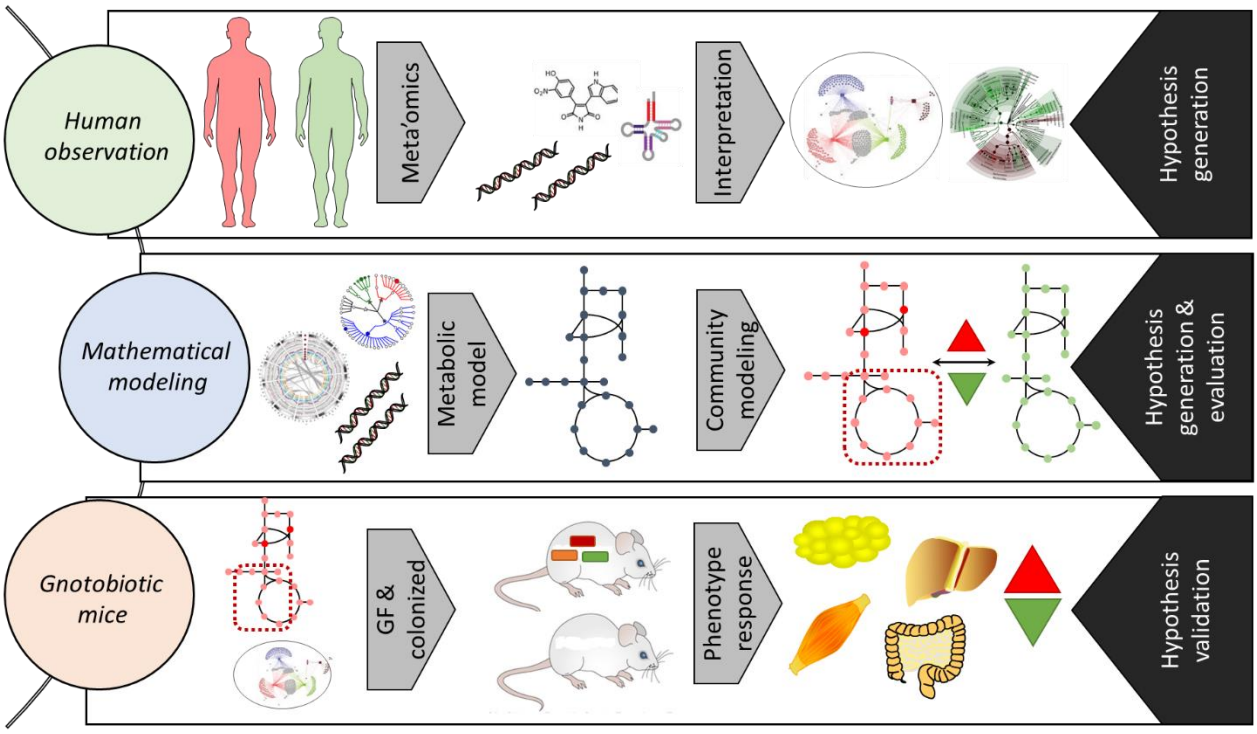


Figure 4 Proposed pipeline for understanding the causality of microbial communities and the interactions between the members. The microbial imbalance and the microbes associated with disease can be defined by applying metanomics approaches to different human microbial ecosystems in healthy and diseased conditions. These findings can be harvested to establish a workflow for genome-scale modeling. Through different applications of GEMs, the hidden causalities of this community can be elucidated and different hypothesis can be generated or evaluated. In order to validate the hypothesis and directly link the microbes with the disease state, the GF mice can be colonised with the target microbes and the responses of different host tissues can be analysed.

2.3.2. Modeling based methods for studying the community

The animal models can be limited and sometimes fail to discover the mechanisms of different interactions between the community members. As we understood from the modeling approaches for a single organism, they are a great scaffold for elucidating the underlying mechanisms behind them. It is therefore necessary to use these modeling approaches in the community to provide a better understanding of mechanisms within the complex community. As described before, GEMs are a great platform for elucidating underlying genotype-phenotype relationships at the single species level. With developing different frameworks, such modeling can be performed for the communities. Generating different GEMs of species within a community can facilitate interpretation of their phenotypic behaviour and elucidate the metabolic machinery of their interactions with other members. However, to describe the overall metabolism of community, the metagenome-scale models can be reconstructed (Greenblum et al., 2012). To address the questions concerning the community through modeling, two alternative approaches; network-based and constraint-based modeling, can be proposed. These two are briefly described in the next two sections.

Network-based models for studying the community

One way of using GEMs is the topology information that can be extracted from these models. Although network-based models employ a simple approach, they generates valuable insights into the metabolism of single species and multi-species, regardless of flux and stoichiometric parameters (Borenstein et al., 2008; Jeong et al., 2000). These models can be set up by using minimum information and so can be applied to complex microbiome studies. Through reconstruction of multiple relevant GEMs in different communities, the metabolically cross-species interactions have been investigated. In one study using this approach, the bacterial ecological co-occurrence in large-scale was explored (Freilich et al., 2010). This approach has been applied to human gut ecosystems through reconstruction of more than 150 GEMs for gut-inhabited species (Levy and Borenstein, 2013). Looking at each pair of species and using the network-based method, the metabolic competition and cross-feeding over them was predicted. Using the predictions and taking to account the co-occurrence of species through metagenomics data, the rules behind the gut community-level assembly have been studied. Cottret et al have linked the network-based information from interacting species and host to exchanges in metabolites (Cottret et al., 2010). Moving away from an individual species approach, GEMs can be reconstructed in the microbiome scale through pooling all relevant biochemical reactions that occur in the community. Using these metagenome-scale models may facilitate the integration of meta'omics data and lead to better exploration of the community behaviour. Greenblum et al. has integrated the metagenomics data into the community level metabolic model and linked the microbiome enzyme abundances to obesity as well as inflammatory bowel disease (Greenblum et al., 2012). This group of enzymes had close interplay between microbiome and host, and affects the physiology of the host. In different studies, metatranscriptomics and metametabolomics data have also been integrated into these large scale models and key components have been identified in the community from different ecological changes (Hartman et al., 2009; Jorth et al., 2014).

Constrained-based modeling for studying the community

The predictive and descriptive power of CBM has shown for identifying the characteristics of single species. Taking account the stoichiometry information and introduced constraints for the metabolic fluxes, the overall distribution of fluxes can be determined through optimising for predefined objective functions. As I mentioned before, the most applied objective function for microbe GEMs is biomass which is optimisation of the uptake and secretion of metabolites to predict the growth. This successfully established workflow for individual species persuades researchers to propose the analogous road map for using CBM in communities. The multispecies CBM can be practical through revision of the current problem formulation and tuning more parameters. The metabolic modeling between methanogenic archaea and sulphate-reducing bacteria was the very first effort towards multispecies CBM (Stolyar et al., 2007). The formulation for this modeling was scaling up the stoichiometric matrix and considering each model as a discrete compartment and one additional compartment for metabolites exchanged between the two species. This modeling approach predicted the overall metabolic fluxes and the phenotype of each species in the community as well as the growth parameters and the composition of the community. This expansion of stoichiometric matrix with different compartments started to be popular for modeling communities. Freilich et al explored the interactions between the bacterial pairs in the forms of competition and cooperation using metabolic models for 118 species and FBA (Freilich et al., 2011). They observed that the cooperative interactions have a tendency to be unidirectional although these species make a cooperative loop so all the species can benefit. Most of these studies

that have taken the advantage of the compartmentalisation approach have integrated the measured abundances ratio into the community objective function. The weighted summation of biomass for microbial community is therefore considered as an objective. However, determining this objective has a limitation due to indefinite growth rate ratios, so a more applicable method has been proposed. A powerful approach called OptCom by *Zomorrodi et al. 2012* has suggested using the multi-level objective function to use in community modeling (Zomorrodi and Maranas, 2012). An inner objective for species-level and one outer objective for overall community growth, the overall flux distribution and growth phenotype of species and community can be predicted through bi-level optimisation. This has been followed by a dynamic approach for multi-level and multi-objective metabolic modeling of the communities, called d-OptCom. So this framework makes it possible to simultaneously optimise for the community-level and species-level, while capable of extracting the dynamic biomass concentration of the community and individuals as well as the exchanged metabolites. In another study, community modeling has advanced using CBM for the gut microbe and its host interactions, wherein the microbe growth has been constrained and the host phenotype has been optimised (Heinken et al., 2013).

3. Results and discussion

In the following sections summarize the publications underlying the thesis. The result section can be divided to understanding the microbe-microbe interactions in gut using GEMs (Section 3.1), comprehensive computational and *in-vivo* tools for analysis of microbe-microbe interactions in gut (section 3.2) and quantifying the metabolic changes of host-microbes (section 3.3).

3.1. Understanding the microbe-microbe interactions in gut using GEMs

The human gastrointestinal system contains a variety of microbes outnumbering our own cells by a factor of 10, and previous studies have shown that the composition of the gut microbiota is influenced by diet, environment and age. The gut microbiota is involved in the conversion of nutrients, the stimulation of the immune system and in providing resistance to pathogens. The human gut microbiome can therefore play an influential role in maintaining human health, and it is a potential target for prevention and treatment of disease. This will, however, require understanding of the complex interactions between diet, microbiota and the host phenotype. Due to the extreme complexity of this system, mathematical models will be needed for deciphering the role of its individual elements and hereby assist in providing an understanding of the mechanisms behind the effects of diet, the genotype-phenotype relationships and microbial robustness. Recently genome-scale metabolic models (GEMs) have been used to explore the interactions between microbes in community and microbes and human host, but these models are not directly applicable to the gut microbiome and its interactions with the host metabolism.

3.1.2. Paper I: Understanding the interactions between bacteria in the human gut through metabolic modeling

The gut microbiota functions as a metabolically active organ and digests dietary components that are indigestible to human cells which can then be absorbed and metabolized by the human body (Backhed et al., 2005). The microorganism which inhabits the human gastrointestinal system is referred to as the gut microbiota and the genome sequence of this ecosystem is called the gut microbiome. The human gut microbiota carries out different metabolic functions relevant to the host, and perturbation or diversion of the microbiota can lead to disease development (Kinross et al., 2011; Turnbaugh et al., 2009a). Improvements in DNA sequencing technology and cost reductions open new possibilities to study the human microbiome in health and disease. As expected from previous studies, the 16 S rRNA sequencing has indicated the gut microbiota to be mainly dominated by the phyla Bacteroidetes and Firmicutes with 17-60% being Bacteroides and 35-80% being Firmicutes (Costello et al., 2009; Eckburg et al., 2005; Karlsson et al., 2011; Tap et al., 2009). Other key phyla in the human gut microbiota are Actinobacteria, Proteobacteria and Euryarchaeota (Arumugam et al., 2011). In this work we aim make use of two recently available technologies: metagenomics and genome scale metabolic model (GEM) simulation in order to generate an integrated mathematical model that describe metabolism in the gut ecosystem and its interactions with the host. Two well characterized bacteria, *Bacteroides thetaiotamicron* and *Eubacterium rectale* as representatives of the two abundant phyla, Bacteroides and Firmicutes were chosen for GEM reconstruction. *Methanobrevibacter smithii*, as a methanogenic and dominant archaeon in the human gut microbiome was selected as a third species because it plays a key role in gut microbial metabolism of hydrogen (Samuel et al., 2007). Despite of low abundance of *M. smithii* in metagenomics studies, it has a significant role in the human gut by removal of hydrogen gas and production of methane. Removal of hydrogen gas is important to consider as it affects

bacterial fermentation and energy harvesting (Armougom et al., 2009). The interactions between these three species are mediated by the production of short chain fatty acids (SCFAs) (acetate, propionate and butyrate), hydrogen and methane. The SCFAs absorbed through the gut epithelial cells, have strong effects on the energy regulation and the immune system of the host (Comalada et al., 2006). The relative absorption of SCFAs by colon varies between 60-90% (McNeil et al., 1978; Ruppin et al., 1980), and oxidization of SCFAs can provide energy for colonic mucosa and may contributes as much as 5-10% of the total energy for a healthy body (McNeil, 1984) (Figure 5).

To model the function of bacteria in the human gut ecosystem, we reconstructed GEMs of three key species; *iBth1201* (*B. thetaiotaomicron*), *iEre400* (*E. rectale*) and *iMsi385* (*M. smithii*), which are relevant representatives of three phyla in the human gut (Bacteroidetes, Firmicutes and Euryarchaeota). We formalized mathematically two different scenarios to which the modeling of microbial communities can be applied. In the first case the composition of the diet and the species abundances in the microbiota is known and constitute the input of the model. In this scenario we aim to predict the profile of compounds produced by the microbiota and hence represent metabolites that can be taken up by the host. We refer to this simulation problem as the α-problem and a solution is found by minimizing the substrate uptake rate. Alternatively we might be interested in predicting the abundances of the different species in the microbiota as a function of the diet. We refer to this problem as the β-problem. We further demonstrate that by integrating the topological information provided by GEMs with diet compositions, genomic and transcriptomic data, community metabolic modeling provide a mechanistic interpretation to statistical findings provided by metagenomics, and this leads to improved understanding of the relationships between diet, microbiota and disease and hereby enables a rational design of prebiotic and probiotic treatments.

After in-silico evaluations of single species as a simple model for gut microbiome, gut ecosystems involving two species were simulated. When *E. rectale* is together with *B. thetaiotaomicron*, two bacteria cooperate through exchanging acetate. In presence of *B. thetaiotaomicron*, *E. rectale* utilizes some portion of acetate produced by *B. thetaiotaomicron* and produce butyrate. Minor proportions of carbohydrates are utlized by *E. rectale* and most are taken up by *B. thetaiotaomicron*. The community stoichiometric matrix was established based on single stoichiometric matrixes within another compartment for the exchange of metabolites between the two species. Through solving α-problem, the biomass of each individual in the community was constrained to its experimental value and FBA was employed to minimize the input utilization. Then the SCFAs production was compared to available mice data (Figure 6A). Co-colonization of *B. thetaiotaomicron* and *M. smithii* was applied for modeling as well, where the key interactions between them are acetate and formate. These two metabolites are taken up by *M. smithii* and methane is produced through methanogenesis pathway (Figure 5). The α-problem was solved as well for this set up.

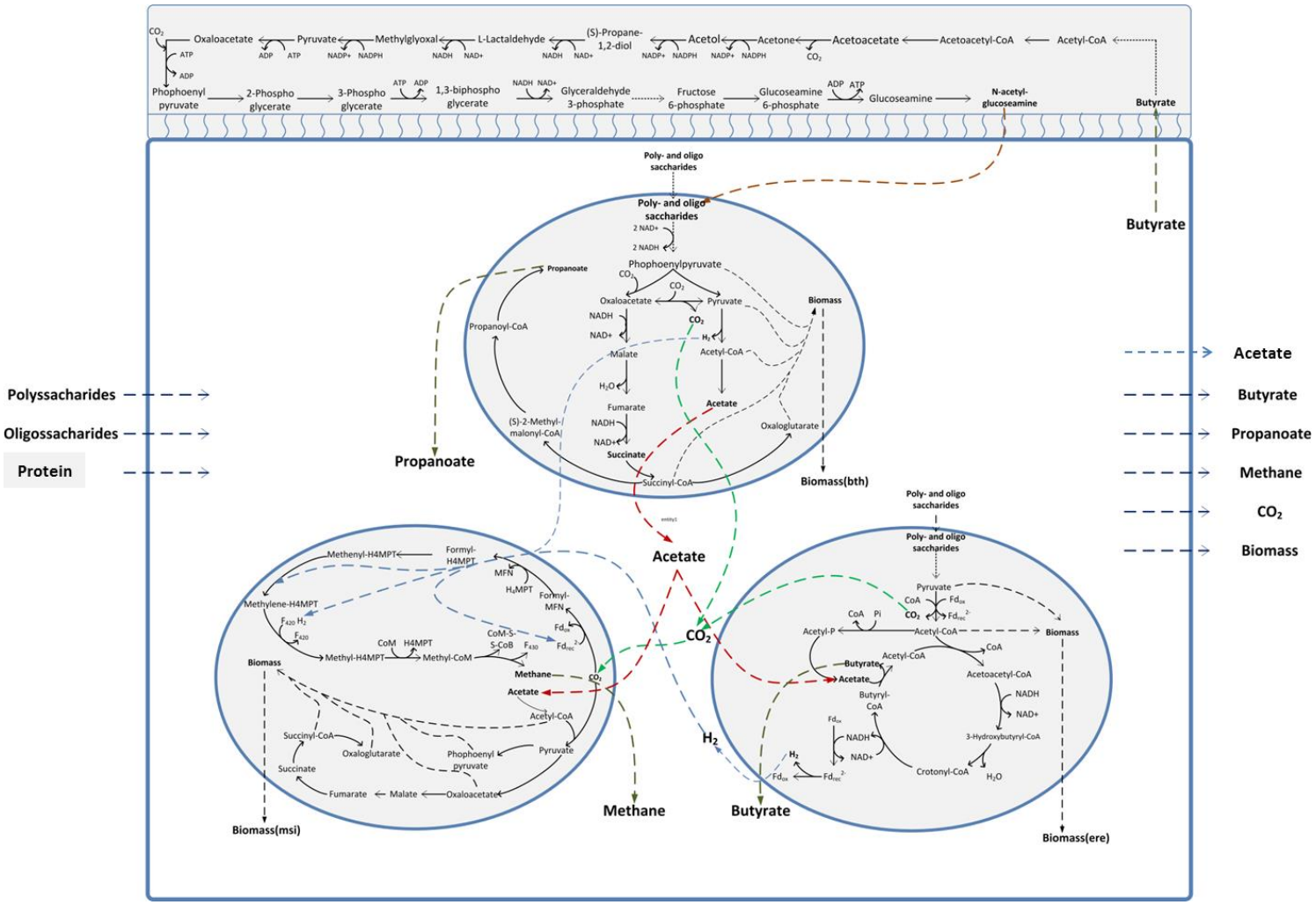


Figure 5 Overview of metabolism by the three bacteria in the gut. This figure show simplified pathways in bacteria for production of SCFAs and other products.

The β-problem is described as the biomass of individuals as objective function and preditc the other products through given only the composition of the diet as an input. Through solving the β-problem for a community composed of *B. thetaiotaomicron*, *E. rectale*, *M. smithii* as representatives of predominant phyla in the human gut microbiota, leads to determine the optimum SCFAs production based on different abundances of *B. thetaiotaomicron*, *E. rectale*, *M. smithii* and substrate. The interplay between the members of this community were acetate through secretion by B. thetaiotaomicron and utilization by *E. rectale*, *M. smithii*. As well CO_2 produced by *E. rectale* and consumed by *M. smithii*. The substrate was divided between *E. rectale* and *B. thetaiotaomicron*, based on the ratios obtained from simulations with the α-problem with *E. rectale* and *B. thetaiotaomicron*. By solving this problem SCFAs production, abundances of each individual and other byproducts such as methane and succinate were predicted. (Figure 6B).

GEMs link genotype and phenotype and allow for making new hypothesis for cellular metabolism and generating novel biological targets based on omics data. The transcriptome dataset (Mahowald et al., 2009) that profile the transcriptome of *E. rectale* and *B. thetaiotaomicron* in monocolonized mice with co-colonized mice were used to integrate with the models. Reporter metabolites

algorithm that identifies a set of metabolites around which show a strong transcriptional response and identify the hot spot of the metabolism (Patil and Nielsen, 2005). Metabolites involved in amino acid metabolism, aminoacyl-tRNA biosynthesis, TCA cycle, NAD and CoA and nucleotide biosynthesis were shown as reporter metabolites for the case of *E. rectale* response to co-colonization of *B. thetaiotaomicron*. The neighbor genes for these metabolites were upregulated. Moreover down regulation genes involved in carbohydrate metabolism, identified metabolites like melibiose, fructose, galactose and raffinose as reporter metabolites. In different case when *B. thetaiotaomicron* was co-colonized with *E. ractale*, genes involved in poly and mono-saccharides, carbohydrates and glycans were upregulated and accordingly the reporter metabolites mannan, fucose and glucose were identified.

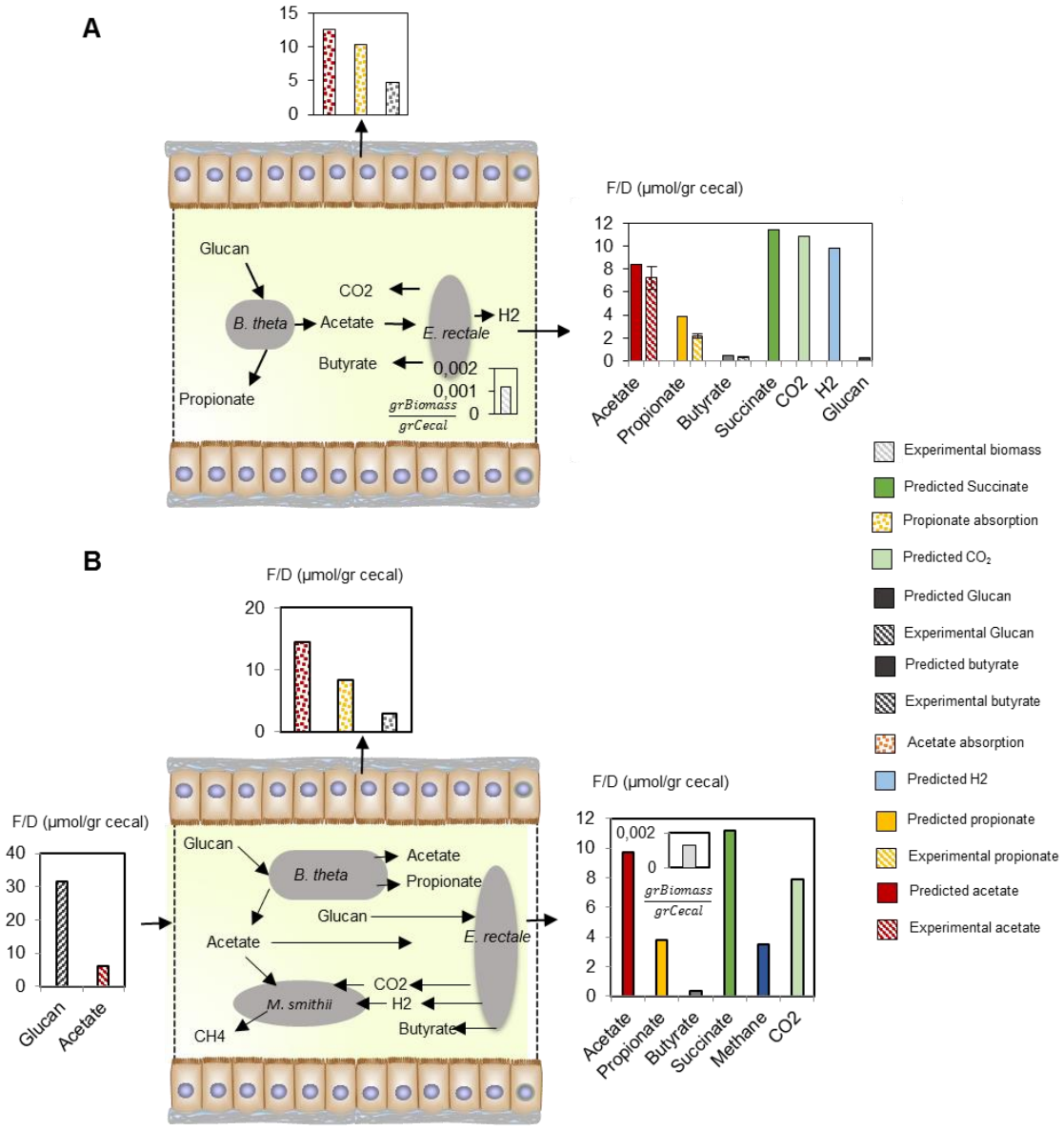


Figure 6 Predictions and comparison of modeling two bacteria through solving, α-problem (A), and modeling the three bacteria through solving β-problem (B).

The reporter subnetworks algorithm was used as well to determine the collection of metabolic reactions in association with transcriptional data. Utilization of amino acids were increased and degradation of carbohydrates decreased through adaptation of *E. rectale* to *B. thetaiotaomicron* (Figure7A) whereas utilization of polysaccharides were boosted up when *B. thetaiotaomicron* was adapted to *E. rectale* (Figure 7B).

Through integration of GEMs with transcription data, the transcriptional metabolic responses for each member in response to symbiotic conditions were identified. *E. rectale* when co-colonized with *B. thetaiotaomicron*, expression of genes involved in amino acid metabolism, TCA cycle and purine belong to *E. rectale*, were up regulated and genes involved in the degradation of carbohydrates were downregulated. This findings were supported by Reporter Metabolite and subnetworks. It appears *E. rectale* increase its abundance dependent on the amino acid consumption in particular glutamine which is the most abundant amino acid in the blood. Glutamine is a source of nitrogen and is utilized as the precursor for the biosynthesis major biomass macromolecules. Glutamine consumed by *E. rectale* is transformed to glutamate and to other intermediates, required for the synthesis of alanine, aspartate, arginine and proline. It has been shown the decreased level of glutamine in the PPAR-α null mouse, indicates number of discrepancies linked to diabetes and the metabolic syndrome (Atherton et al., 2006) and the high abundance of Firmicutes species were reported in obese mice compare to lean mice. Here we observed *E. rectale* may be involved in reducing glutamine level in the blood through consuming that to increase the biomass.

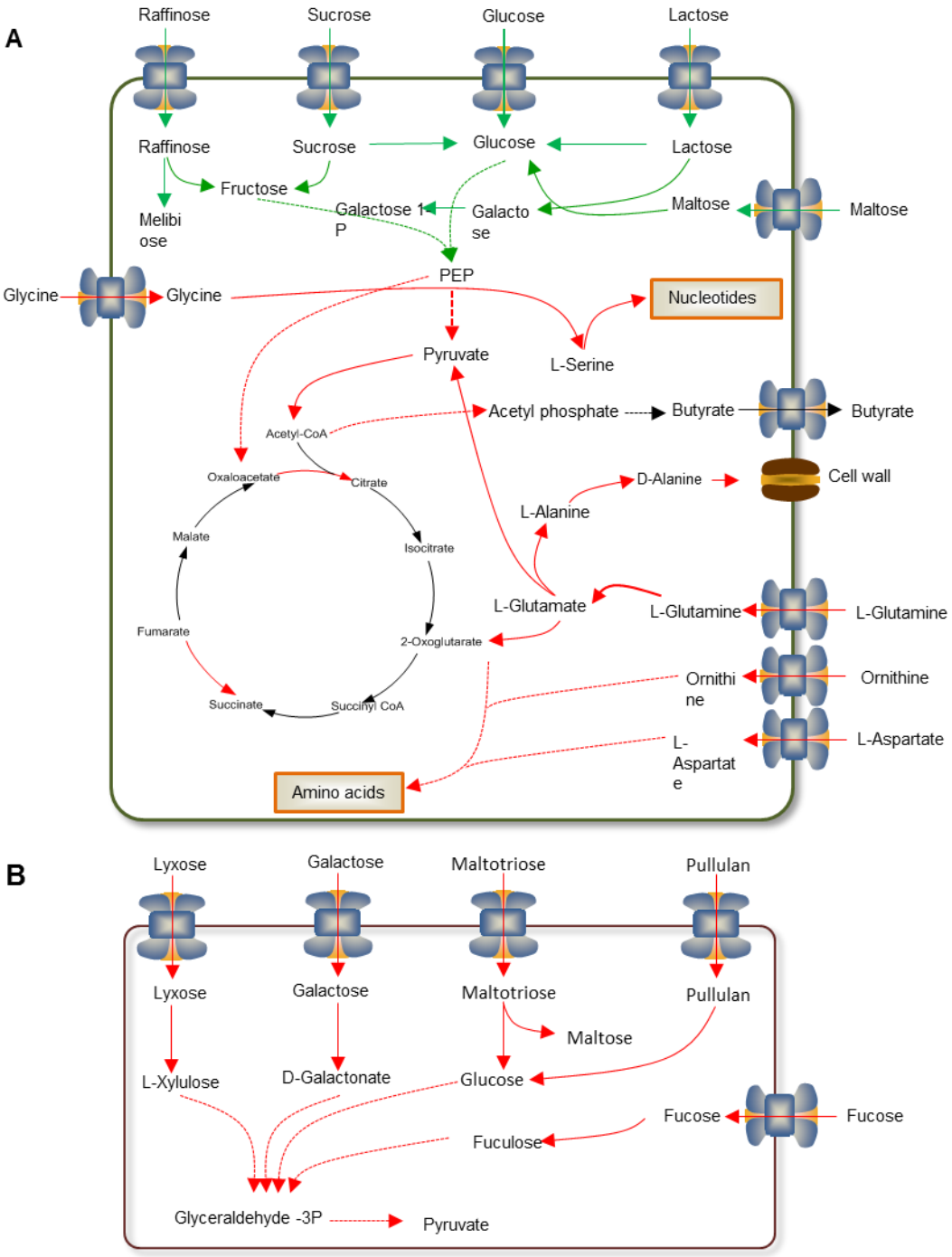


Figure 7 Result of reporter subnetworks for transcriptional response of co-colonization, when *E. rectale* responds to *B. thetaiotaomicron* colonization (A) and of *B. thetaiotaomicron* responds *to E. rectale* colonization.

3.2. Comprehensive computational and *in-vivo* tools for analysis of microbe-microbe interactions in gut

In this section we present two platforms for analysis of microbe interactions in a community. These platform can be used as well to look at the interactions between microbes-host. Here, we describe the CASINO (Community and Systems-level Interactive Optimization) Toolbox for analysis of microbial communities through metabolic modeling. We tested CASINO on a gut ecosystem to understand the mechanistic insight into the contribution of individual species in the gut microbiome to the overall metabolism of the ecosystem and further how the ecosystem and the individual species contribute to host metabolism. Moreover as in-vivo model, we developed SIM (Simplified Intestinal Microbiota) consists of ten representative strains isolated from the human gut colonized in to germ-free mice. This great tool empowers the understanding the microbe-microbe and host interactions and facilitates the generation and evaluation of new hypothesis.

3.2.1. Paper II: Quantifying diet-induced changes in metabolism of the human gut microbiome through metabolic modeling

We have previously reconstructed GEMs for *Bacteroides thetaiotamicron*, *Eubacterium rectale* and *Methanobrevibacter smithii* which are the relevant representatives of three dominant phyla in the human gut and studied the interactions between these bacteria in different combinations of gut ecosystems (Shoaie et al., 2013). We also reconstructed GEMs for *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii* and simulated the interactions between these bacteria (El-Semman et al., 2014). In both studies, the interactions between the bacteria were identified manually and the consumption and production rates of the defined interacting metabolites by each bacteria were quantified. However, approach cannot be used for simulation of the interactions for a large number of species representing the complete gut ecosystem. We therefore developed the CASINO (Community And Systems-level INteractive Optimization) Toolbox that comprises an optimization algorithm along with diet analysis and allocation algorithm for prediction of the phenotypes within the human gut microbiota. The optimization algorithm was designed based on collaborative and a multi-dimensional distribution approach. In CASINO, the model of the community system is linearized by separating the lower level describing the individual species (organism-level) and the higher level describing the community (systems-level). CASINO as well takes advantage of community topology through identification of species as effector (producing metabolites) and receptor (consuming the metabolites produced by effector). Using this information in the first stage the metabolites secreted by each model determined through organism-level optimization. Then the iterative multi-level optimization is performed until the total community biomass is optimized. We used CASINO to simulate the interactions between the microbes in two *in-silico* microbial communities including EBBR (*E. rectale, B. adolescentis, B. thetaiotaomicron* and *R. bromii*) and FBBR (*F. prausnitzii, B. adolescentis, B. thetaiotaomicron* and *R. bromii*). We predicted the net production and consumption of the metabolites produced/consumed by each community as well as the contribution of individual bacteria to the phenotype of the each community. Figure 8 summarizes different stages of the study. After validation of the individual GEMs and confirming the predictive power of the CASINO toolbox, we simulated the effect of the diet on two different subject groups that have been classified based on their gut microbial gene richness.

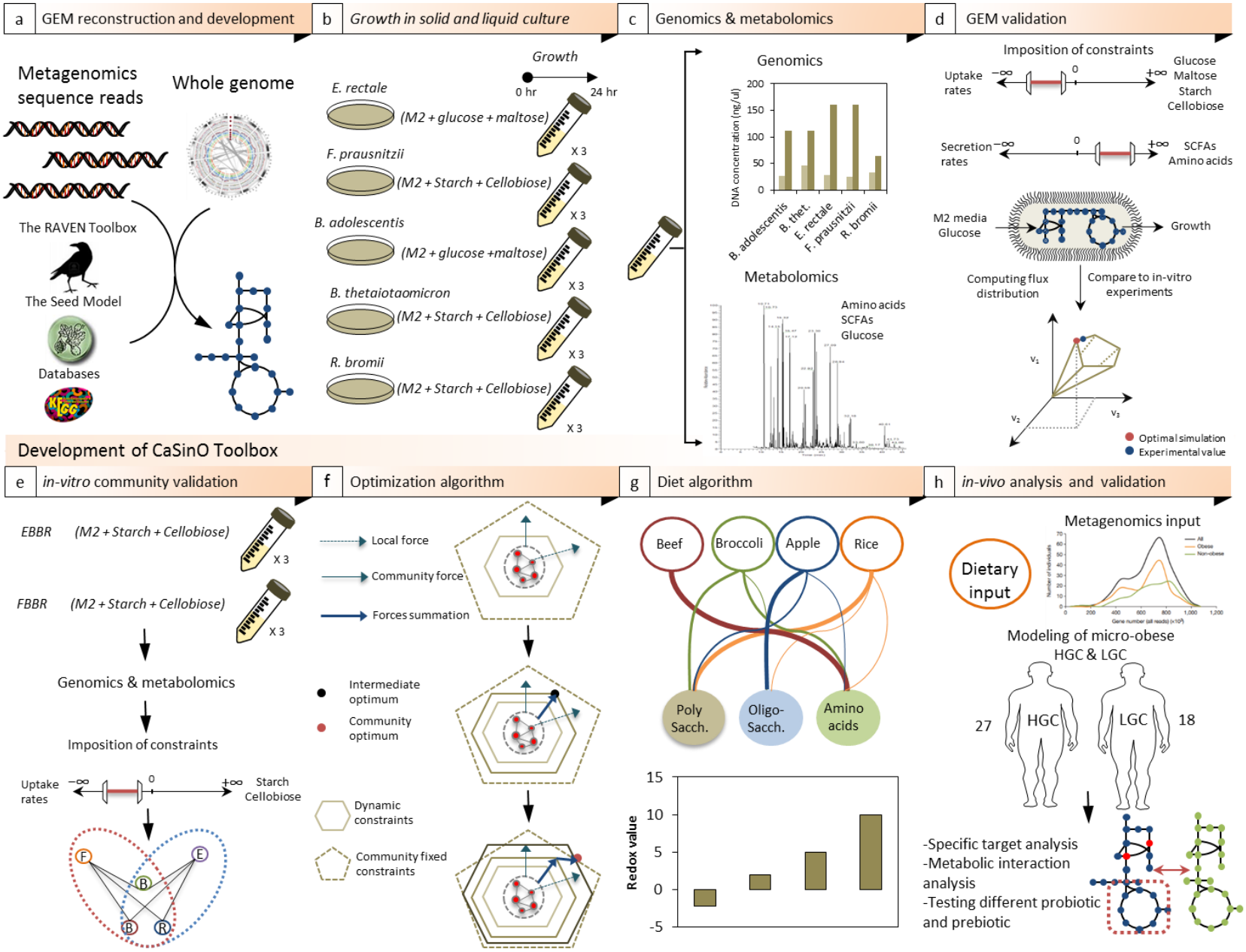

a GEM reconstruction and development
b Growth in solid and liquid culture
c Genomics & metabolomics
d GEM validation
Metagenomics sequence reads
Whole genome
The RAVEN Toolbox
The Seed Model
Databases
Growth
0 hr
24 hr
E. rectale
(M2 + glucose + maltose)
F. prausnitzii
(M2 + Starch + Cellobiose)
B. adolescentis
(M2 + glucose +maltose)
B. thetaiotaomicron
(M2 + Starch + Cellobiose)
R. bromii
(M2 + Starch + Cellobiose)
X 3
Genomics
DNA concentration (ng/ul)
B. adolescentis
B. thet.
E. rectale
F. prausnitzii
R. bromii
Metabolomics
Amino acids
SCFAs
Glucose
Imposition of constraints
Uptake rates
Glucose
Maltose
Starch
Cellobiose
Secretion rates
SCFAs
Amino acids
M2 media
Glucose
Growth
Computing flux distribution
Compare to in-vitro experiments
Optimal simulation
Experimental value
Development of CaSinO Toolbox
e in-vitro community validation
f Optimization algorithm
g Diet algorithm
h in-vivo analysis and validation
EBBR (M2 + Starch + Cellobiose)
FBBR (M2 + Starch + Cellobiose)
Genomics & metabolomics
Imposition of constraints
Uptake rates
Starch
Cellobiose
Local force
Community force
Forces summation
Intermediate optimum
Community optimum
Dynamic constraints
Community fixed constraints
Beef
Broccoli
Apple
Rice
Poly Sacch.
Oligo-Sacch.
Amino acids
Redox value
Metagenomics input
Dietary input
Modeling of micro-obese
HGC & LGC
27
HGC
LGC
18
-Specific target analysis
-Metabolic interaction analysis
-Testing different probiotic and prebiotic

Figure 8 The outline of study (a) Genome-scale metabolic models (GEMs) for *L. reuteri and R. bromii* were generated by the RAVEN Toolbox using whole genome sequencing and metagenomics data, and these models were manually curated based on literature information for the reconstruction of functional models. Previously reconstructed GEMs for *B. thetaiotaomicron, E. rectale, B. adolescentis* and *F. prausnitzii* were also expanded using the same pipeline. (b) Each model was validated by comparison with experimental data obtained from culturing bacteria for 24h in M2 media supplemented with different carbon sources for optimum growth of each bacterium. Growth was measured after 24 hours and used for the validation of the model. (c) Samples were collected at baseline and after 24h for performing 16S rRNA and metabolomics analysis. Bacteria were identified using 16S rRNA and DNA concentration was measured using nano drops. The concentration of SCFAs, amino acids and carbohydrates were also quantified. The carbohydrate measurements were done by hydrolyzing all the polysaccharides to glucose. (d) Each GEM was validated based on the genomics and metabolomics data generated by *in vitro* experiments. The byproducts and the substrate usage were constrained in the models and the growth rate was compared with the experimental data. (e) Two *in-silico* microbial communities *EBBR (E. rectale + B. adolescentis + B. thetaiotaomicron + R. bromii)* and *FBBR (F. prausnitzii + B. adolescentis + B. thetaiotaomicron + R. bromii)* were designed and simulated using CASINO. The results were compared with *in-vitro* experiments for *EBBR* and *FBBR* communities. (f) In CASINO the interactions of the bacteria as well as the phenotype of the community were identified using an optimization algorithm. Growth of each bacterium had local optimum whereas the community had global optimum. The community optimum was detected by the intersection point of the community fixed constraints and calculated dynamic constraints which is obtained by local and community forces summation. (g) A diet algorithm was developed and implemented for prediction of the macromolecules of diet that is used as an input for simulations. The diet is converted to three main categories of macromolecules polysaccharide, oligosaccharide and amino acids. (h) CASINO was applied to study the effect of the diet on the gut microbiota composition of subjects that were classified based on their microbial gene richness. Simulations were conducted and the predictions were confirmed by metabolomics data performed for fecal and serum samples obtained from these subjects and new hypothesis were generated.

The gene richness of the human gut microbiome has been analyzed for 123 non-obese and 169 obese individuals based on the threshold of 480,000 genes (Le Chatelier et al., 2013). 45 obese and overweighed individuals from this study were subjected to energy restricted high-protein diet for 6 weeks and to weight maintenance diet for another 6 weeks. The diet uptake was recorded before and after diet interventions for each individual. Among the 45 individuals 18 were grouped as low gene count (LGC) and 27 as high gene count (HGC). It was earlier shown that the dietary interventions resulted in different phenotypic responses based on the microbial gene richness of the individuals (Cotillard et al., 2013). Here we simulated the effect of the diet on the human gut microbiota composition at baseline and after the diet-induced weight-loss (6 weeks). We wanted to use CASINO to predict the contribution of the whole gut microbiota to the overall host metabolism as well as the contribution from each bacteria. Before metagenomics analysis identified six species dominated in all individuals (Cotillard et al., 2013). Following the metagenomics we performed 16s rRNA qPCR on fecal samples obtained from the 45 individuals at baseline and after 6 weeks to capture the quantitative data for the six species (Figure 9).

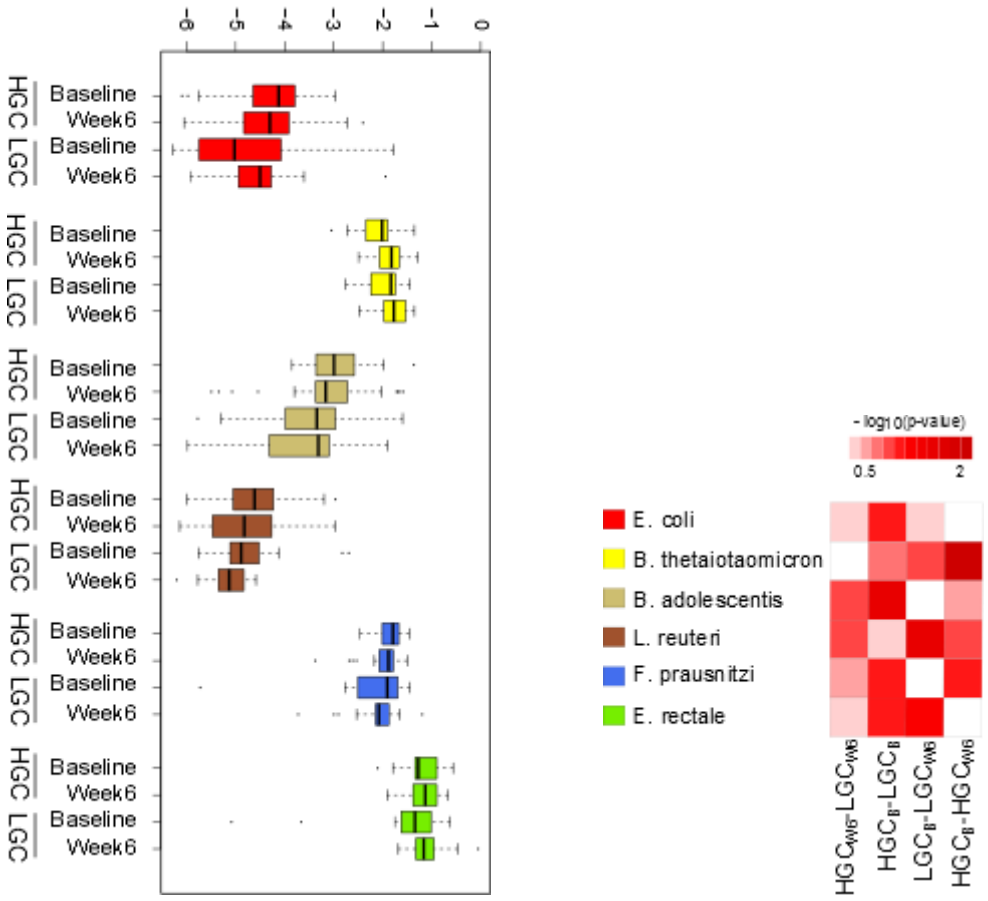


Figure 9 The effect of the 6 weeks diet interventions in HGC and LGC individuals. The abundance of the bacteria before and after diet interventions in HGC and LGC subjects are shown.

In order to simulate the effect of the diet based on the abundant bacteria in LGC and HGC individuals, we used GEMs for *B. thetaiotaomicron*, *B. adolescentis*, *F. prausnitzi*, *E. rectale* and *Lactobacillus reuteri*, a representative of the Lactobacillus genus and obesity-associated species (Million et al., 2012). Using the GEMs for the abundant bacteria we used CASINO to simulate the effect of dietary intake on the human gut microbiota composition at baseline and after the diet-induced weight loss (6 weeks) for each of the 45 subjects. Dietary information for each individual was recorded by a registered dietician and used to analyze the diets of the 45 individuals. The dietary macronutrients were computed and changes between the macronutrients at baseline and after 6 weeks of dietary interventions were calculated and used as inputs to CASINO for simulations. The model simulations predicted the profile of three SCFAs and 14 amino acids produced by the gut ecosystem as well as the contribution of each microbial species to the overall metabolite production by the ecosystem at baseline and after 6 weeks for each individual. Looking at prediction, we observed the level of SCFAs and amino acids were significantly reduced in 6 weeks diet interventions and this reduction was greater for LGC individuals (Figure 10).

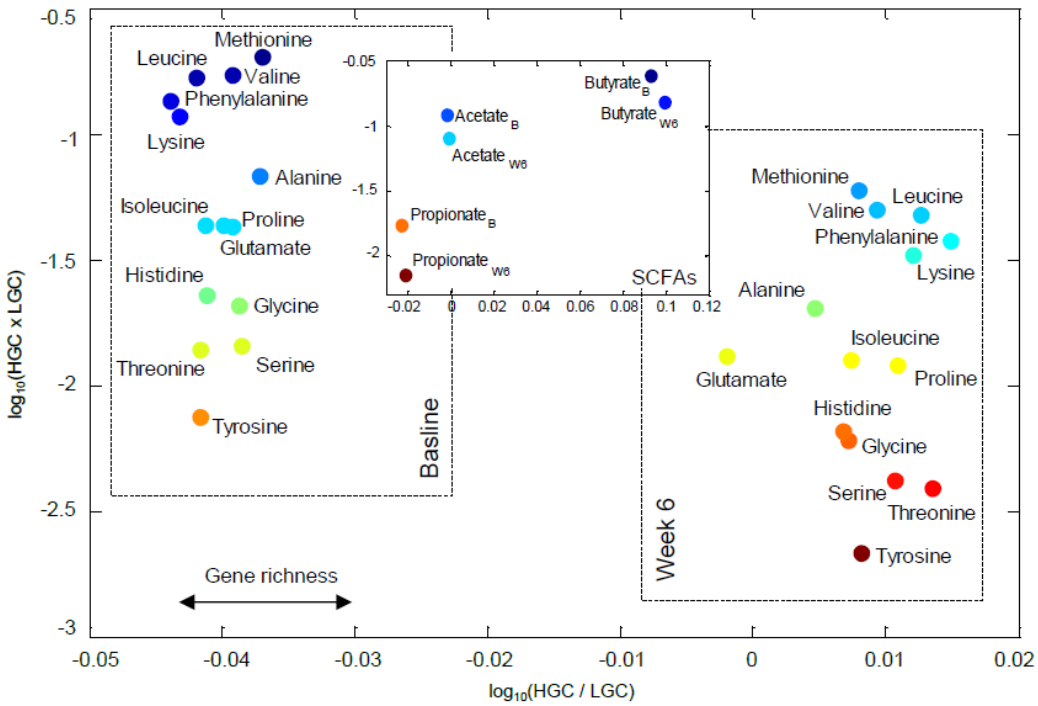


Figure 10 Predictions of SCFAs and amino acids for HGC and LGC subjects at baseline and after 6 weeks of diet interventions. The group of metabolites at the rights show the baseline predictions and at the left the week 6. The middle is the SCFAs predictions for baseline and week 6.

In order to confirm our predictions on altered metabolite production by the gut ecosystem, we performed metabolome analysis of the 45 fecal samples obtained from the same HGC and LGC individuals at baseline and after 6 weeks of dietary intervention. The metabolomics data confirmed several of the predictions. It showed the decrease level of proline, glycine, serine, phenylalanine and tyrosine for all the subjects from baseline to week 6 while the reduction was higher for LGC subjects (Figure 11). Looking at phenylalanine predictions, the level of that for LGC subjects is higher at baseline compare to HGC and after 6 weeks of dietary intervention phenylalanine level was predicted higher for HGC subjects. This findings from predictions were in agreement with the fecal metabolomics. We performed as well serum metabolomics to evaluate the predictions that were not in agreement with fecal metabolomics, which may be is a result of absorption by the host. In addition to many metabolites that we found the correspondence between the predictions and the measured changes, we observed phenylalanine level was higher in LGC subjects at baseline and the level decreased after 6 weeks of diet intervention. Besides, the correlation between amino acids level in serum and bioclinical parameters revealed phenylalanine was positively correlated with variables insulin resistance and BMI (Figure 12).

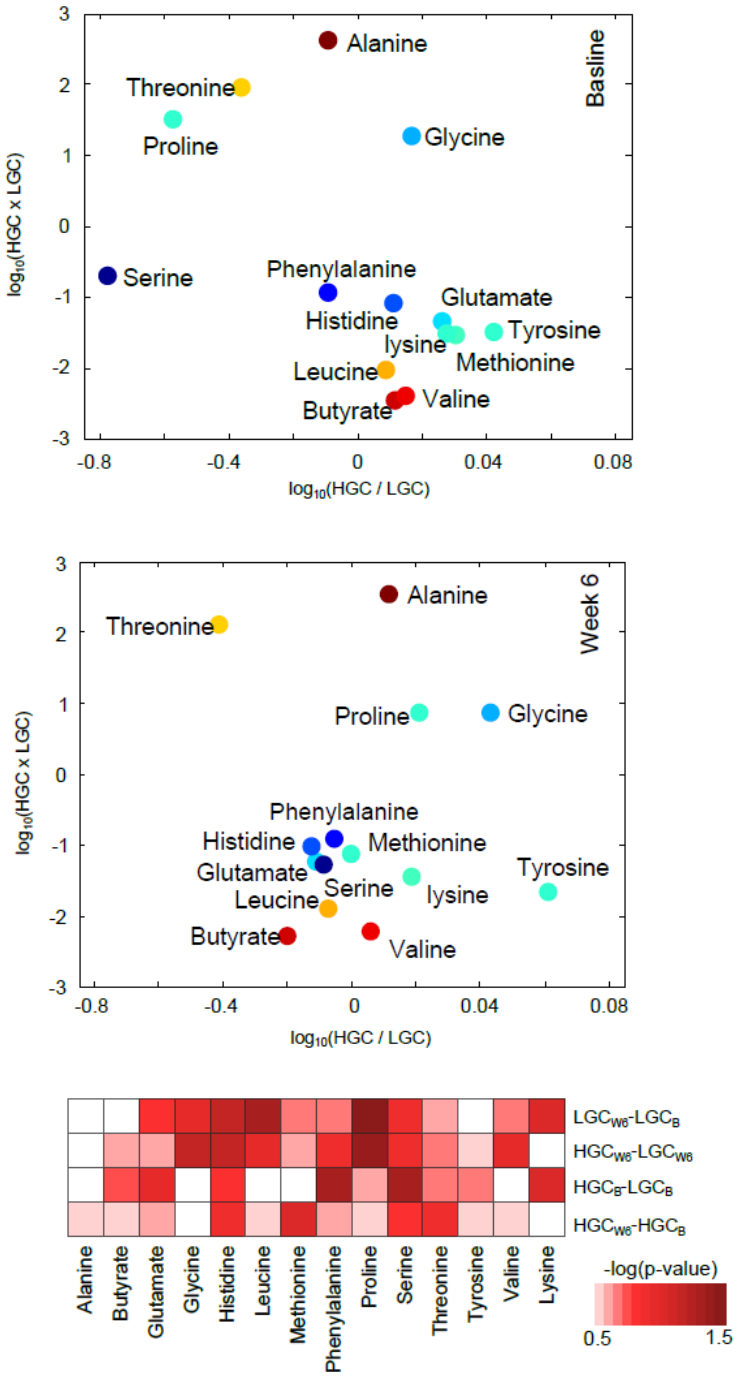


Figure 11 Metabolomics analysis of fecal samples obtained from HGC and LGC subjects. The differences for 14 detected AAs as well as butyrate are shown. Students- t-test was applied to detect significantly changed metabolites

In the final step we tried to use the modeling approach to design a diet to improve metabolism of LGC individuals, which are assumed have a non-optimal gut microbiome metabolism. We performed the simulation through assumption that LGC subjects would enable to reach the optimal gut microbiome metabolism of HGC after 6 weeks of dietary intervention that are associated with an improved metabolic phenotype. Using the abundance of five different species *B. thetaiotaomicron*, *B. adolescentis*, *F. prausnitzi*, *E. rectale*, and *L.reuteri* to predict the relative

utilization of eight essential amino acids by LGC subjects at baseline and HGC subjects at week 6 (Figure 13a). Through correlation of achieved pattern and amino acids in different food types, we identified LGC subjects should increase the consumption of e.g. dairy products, vegetables and fish while decrease e.g. bread and rice (Figure 13b).

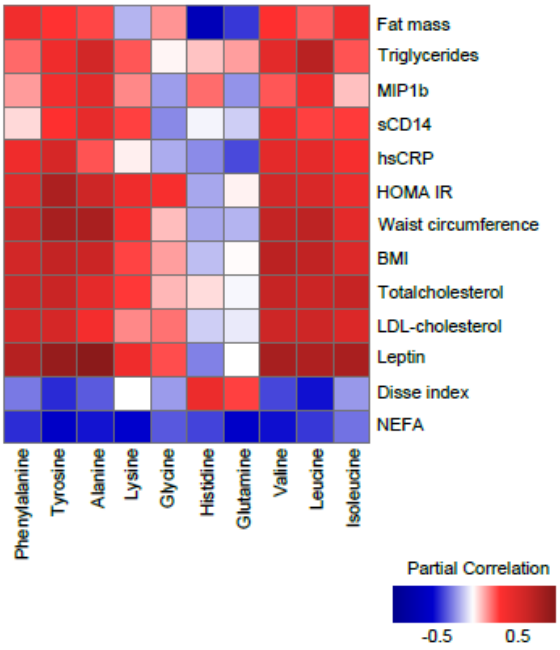


Figure 12 The serum metabolites were correlated with different parameters of HGC and LGC subjects. Fat mass was measured by biphotonic absorptiometry (DXA). MIP1b, macrophage inflammatory protein 1b; sCD14, soluble CD14; hsCRP, human sensitive CRP; HOMA-IR, homeostatic model assessment − insulin resistance = Math Eq; BMI, body mass index (kg/m2); Disse index = Math Eq; MIP1b: macrophage inflammatory protein 1b; hsCRP, human sensitive CRP; NEFA, non-esterified fatty acids. The color specifies the slope of correlation.

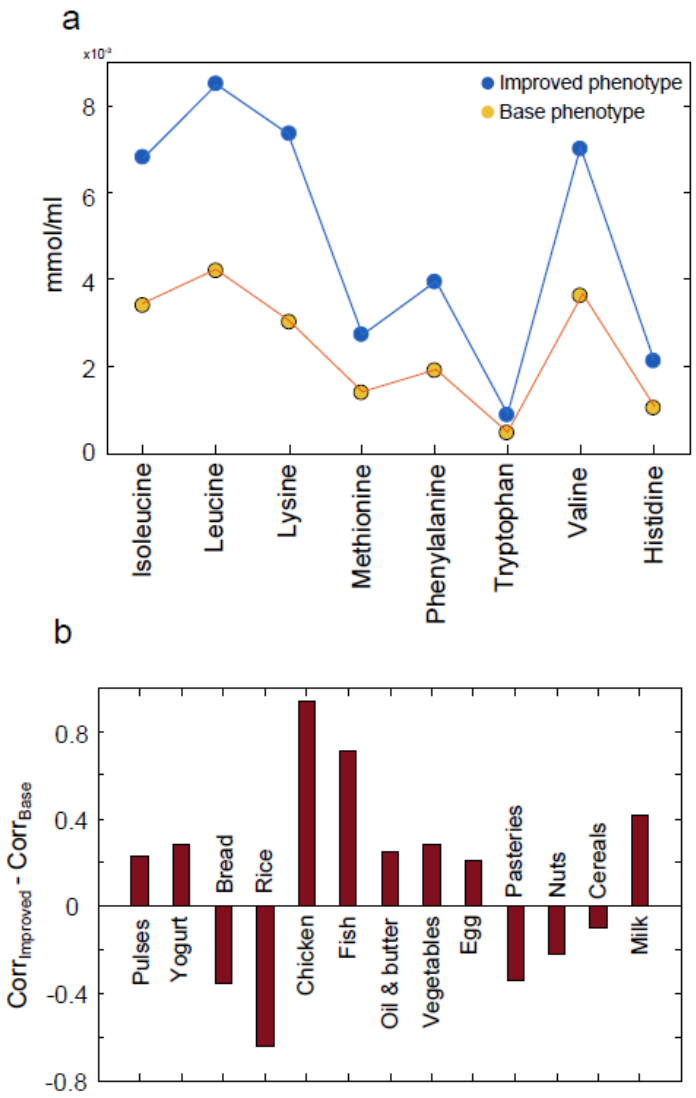


Figure 13 Model-based diet design to improve metabolism of LGC individuals. (a) The yellow circles depicts the consumption of eight amino acids by LGC at baseline and the blue ones are consumption of eight amino acids for HGC at week 6. (b) The figure shows the correlation between the required amino acids at baseline and week 6 with composition of amino acids in different food types. Subtract of improved correlation and baseline correlation the proposed pattern of increase or reduction of food for LGC at baseline to reach the optimal gut microbiota metabolism.

3.2.2. Paper III: SIM mice – a tool to obtain basic and mechanistic understanding of the microbe-microbe interactions in the mammalian gut

Here, we developed a mouse model colonized with ten representatives of the human gut microbiota, with known function relevant to human health microbiota. Here by the means of different omics analysis we study the respond of simplified intestinal microbiota (SIM) to changes in host diet. DNA from jejunum, ileum, caecum, colon and faeces of SIM mice were isolated and 16S rRNA quantitative PCR was performed to understand the distribution of colonized microbes along the intestinal tract and fecal. RNA was isolated from the SIM mice and sequenced to reveal the metatranscriptional of SIM mice fed with autoclaved chow diet. We generated as well metatranscriptome data for SIM mice fed with different diet and examine the respond of the SIM microbes to changes in host diet. We also generated metabolomics data on the plasma sample of SIM mice for different dietary pattern and compared them together and as well with GF mice to uncover the changes of metabolites in plasma of SIM and GF mice.

The selected microbes process metabolic functions identified in the anaerobic food web, considering also possible cooperation via metabolic cross-feeding that likely occurring in the human gut (Table 3). In particular we included strains that have ability to break down complex dietary polysaccharides not accessible to the host (*Bacteroides thetaiotaomicron, Prevotella copri* and *Ruminococcus bromii*), consume oligosaccharides and simple sugars (*Bifidobacterium adolescentis*, *Colinsella aerofaciens*, *Eubacterium hallii*, *Eubacterium rectale* and *Roseburia inulinivorans*), reduce sulfate (*Desulfovibrio piger*). We included also bacteria identified to reside in the mucus layer and use the mucus as a source of carbon and energy (*Akkermansia muciniphila* and *Bacteroides thetaiotaomicron*).

As the bacteria in the SIM microbiota are isolated from human feaces we further studied how the differences in the nutrient availability through the length of the GI tract and SIM microbiota establishes through the length of the mouse gastrointestinal tract. We quantified the levels of each bacteria by 16S rRNA qPCR in the jejunum, ileum, cecum, colon and faeces of the SIM mice. We found out that (i) all bacteria from the SIM microbiota were established in each of the regions of the mouse gut, (ii) the levels of SIM bacteria gradually increased from jejunum towards the distal part of the gut and reached the highest levels in the faeces. The two Bacteroidetes species in the SIM community, *Bacteroides thetaiotaomicron* and *Prevotella copri*, colonize in high levels in all regions of the gut, while from the Firmicutes species only *Roseburia inulinivorans* was highly abundant. The poor establishment of Firmicutes species in the mouse gut and the enrichment in the Bacteroidetes is in agreement with previous findings (Wos-Oxley et al., 2012).

Table 3 Phylogenetic genome size of the members of the simplified intestinal microbiota.

SIM- bacteria	Phyla	Genome size, MB	Reference
Akkermansia muciniphila	Verrucomicrobia	2.66	(Derrien et al., 2004)
Bacteroides thetaiotaomicron	Bacteroidetes	6.29	(Martens et al., 2008; Xu et al., 2003)
Bifidobacterium	Actinobacteria	2.09	(Marquet et al., 2009)
Colinsella aerofaciens	Actinobacteria	2.83	(Kageyama and Benno, 2000)
Desulfovibrio piger	Proteobacteria	3.34	(Marquet et al., 2009)
Eubacterium halli	Firmicutes	3.29	(Duncan et al., 2004)
Eubacterium rectale	Firmicutes	2.44	(Duncan and Flint, 2008)
Prevotella copri	Bacteroidetes	3.51	(Hayashi et al., 2007)
Roseburia inulinovaris	Firmicutes	4.05	(Duncan et al., 2006)
Ruminococcus bromii	Firmicutes	2.25	(Ze et al., 2012)

Next we used three different diets a standard chow diet (CHD - low in fat and high in plant polysaccharides), sucrose diet (ZFD - no fat, low in plant polysaccharides and high in sucrose) and high fat diet (HFD - low in plant polysaccharides, high in fat and sucrose). We found reduced number of bacteria in the SIM mice caecum after HFD feeding compare to CHD and ZFD fed mice. We quantified decreased caecum levels of bacteria such as *Prevotella copri* and *Ruminococcus bromii*, *Bifidobacterium adolescentis*, *Roseburia inulinivorans*, *Eubacterium rectale* and *Colinsella aerofaciens* when plant polysaccharides were omitted from the mouse diet, further supporting the extreme responsiveness of the gut microbiota to macronutrient intake. Other bacteria as *Akkermansia muciniphila* and *Bacteroides thetaiotaomicron* that have the ability to deplete complex glycan structures produced by the host (van Passel et al., 2011; Xu et al., 2007) were not affected by the absence or lower levels of plant polysaccharides in the mouse diet. *Desulfovibrio piger* was one of the few bacteria from the SIM microbiota, which levels in the caecum remain stable during the three dietary changes. (Rey et al., 2013). Unexpectedly, the 16S rRNA levels of the butyrate producer *Eubacetrium hallii* in the caecum of the SIM mice on HFD and ZFD were higher compare to SIM mice fed CHD. We further measured the levels of caecum metabolites, including SCFAs (acetate, propionate and butyrate) and lactate and succinate in the SIM mice fed CHD, ZFD and HFD. The molar ratio of SCFAs in the caecum of the SIM mice on CHD diet was 70:39:10 (acetate:propionate:butyrate) (Figure 14), which is higher than what have been reported recently from our group conventionally raised mice (CONV-R) (Wichmann et al., 2013).

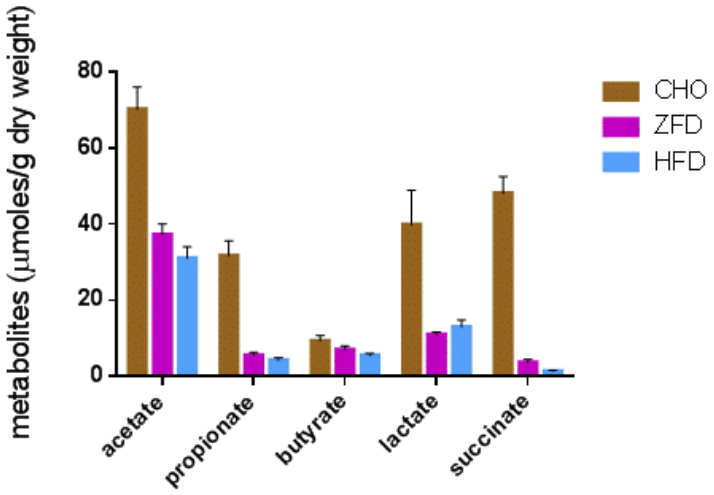


Figure 14 Shift in metabolites in the caecum of SIM mice on chow diet (CHD), sucrose diet (ZFD) and high-fat diet (HFD).

Meta-transcriptomic can be key supplement to metagenome analysis through identification of metabolic active profiles of gut microbiota members. Transcript abundance can uncover the response of microbial members to perturbation within the community. The application of RNA-seq have been applied widely to microbial communities such as marine (Gilbert et al., 2008), soil (Baldrian et al., 2012) and human microbiota (Turnbaugh et al., 2010).

Here we characterized the microbial gene expression of the SIM mice to unravel the activity of each ten member in response to diet changes. mRNA enriched samples of the caecal SIM microbiota at an average depth of 15 million paired-end reads per sample, and relative abundance in the form of counts per million (CPM) were calculated for each protein-coding gene. Then through the differential expression discovery task using edgeR (Robinson et al., 2010), three sets of two group differential comparison were performed. We compared the expression of SIM fed with HFD to CHD, ZFD to CHD and ZFD to HFD (Figure 15). P-values were corrected for multiple testing using adjustment method Benjamini & Hochberg and P-value adjusted were calculated. Afterward, we found there were 5765 genes significantly expressed genes in SIM microbiota fed with HFD compared to CHD (*adj. P* < 0.05).

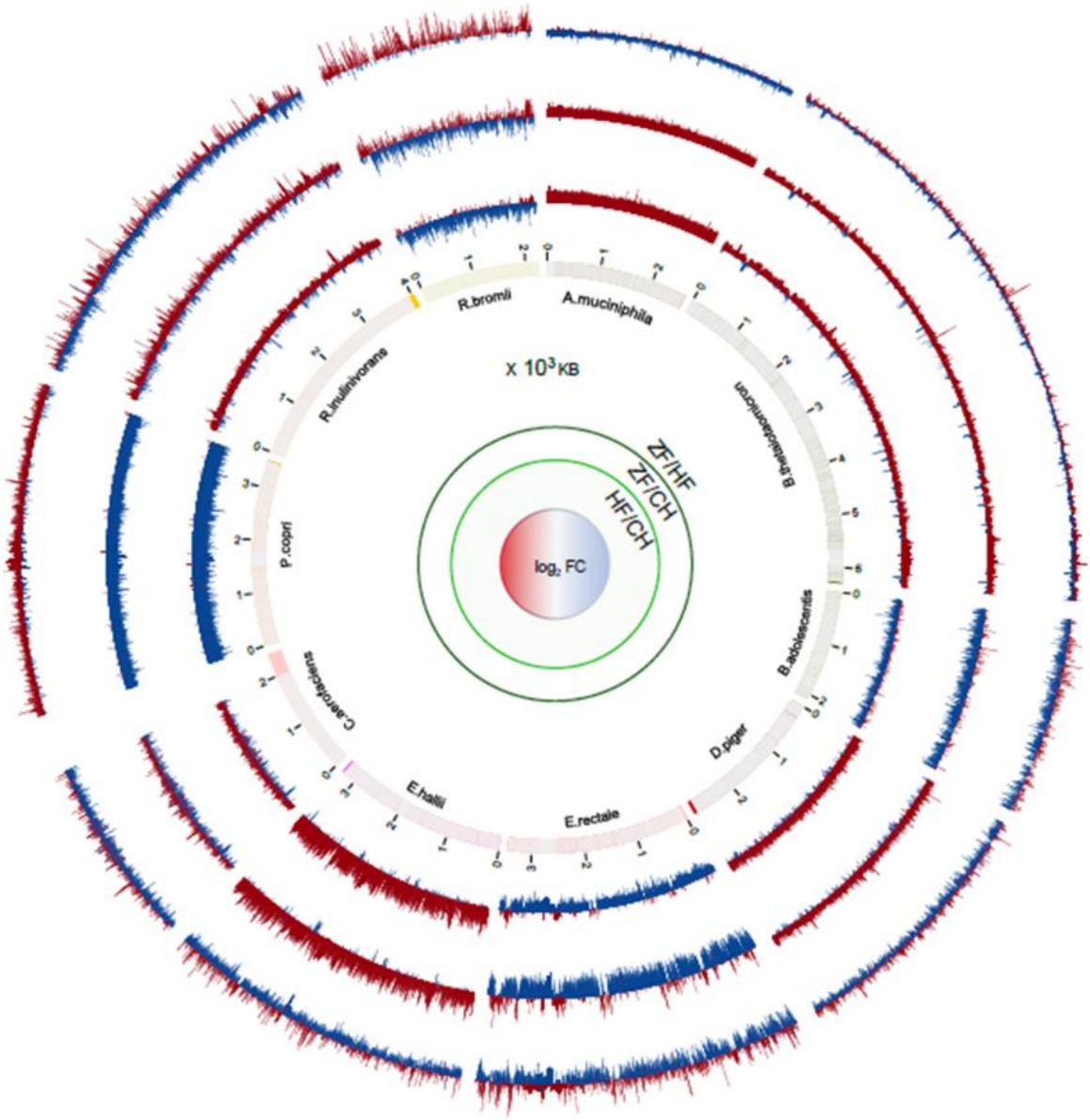


Figure 15 Comparison of differentially expressed genes for 10 members of SIM microbiota. The inside circle is the genome coverage of generated RNA-seq data. The second circle is the differentially expressed genes of SIM fed with HFD compared to CHD, the third circle for ZFD to CHD and the outer one is for ZFD to HFD.

Then we associated the genes to their KEGG orthologies (KO) to understand the active functions of them. Using the KEGG pathways associated with the Kos together with corrected P-values, we identified the differentially abundant Kos for each member of SIM microbiota. The glycolysis pathway was the one with major contribution from *A. muciniphila, E. hallii, P. copri and D. piger* when we compared the HFD to CHD. The responsible genes for 2,3-bisphosphoglycerate were highly expressed in *A. muciniphila* and *E. hallii*,while it was expressed significantly lower in *P. copri*. Fructose-bisphosphate aldolase classII as well has the same pattern. Another pathway that came out was galactose, which genes for α and β galactosidase were significantly expressed. The expression of genes were higher for *A. muciniphila, B. thetaiotaomicron* and lower for *B.*

adolescentis and *P. copri*. α- galactosidase is an enzyme that hydrolyzes ceramide trihexoside. We as well found peptidoglycan, fatty acid, glycerophospholipid, glycerolipid, amino and nucleotide sugar pathwyas contribute significantly (Figure 16).

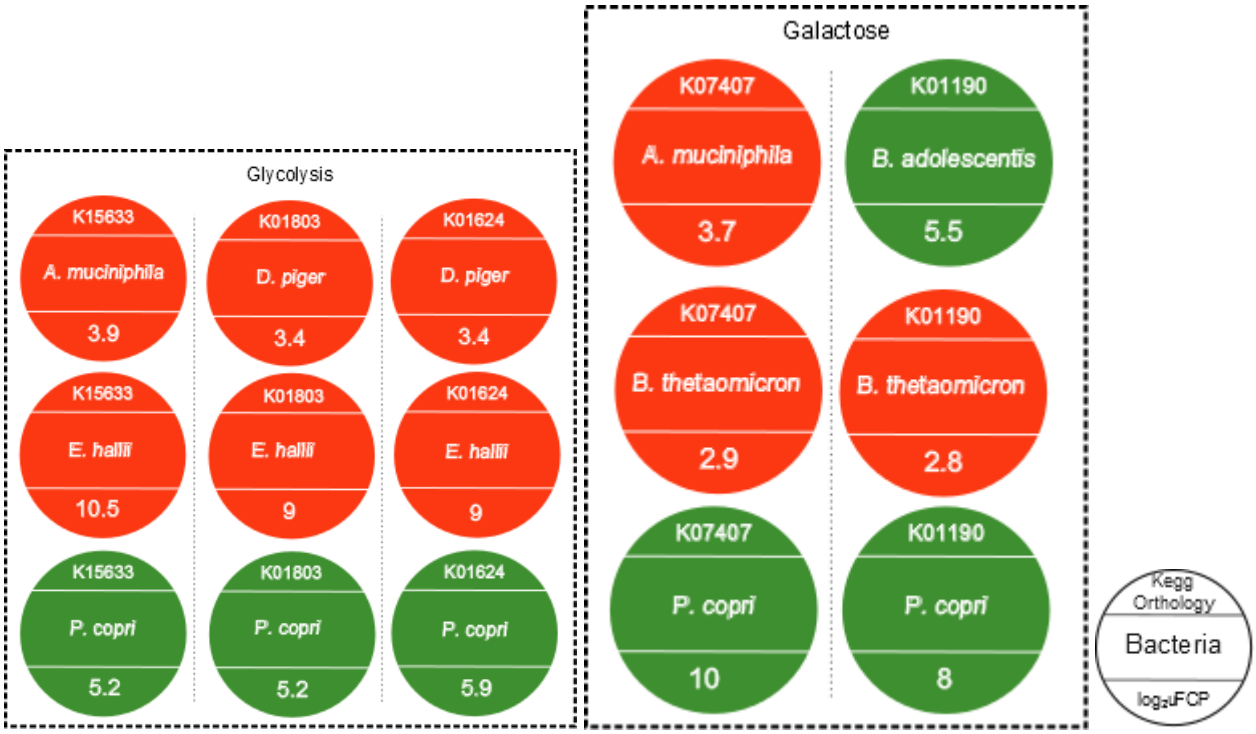


Figure 16 Two examples of contributed pathway with corresponding KOs and SIM members for comparison of HFD and CHD.

To evaluate the impact of the cross-talk between SIM microbiota and diet on microbial-host co-metabolism, we performed untargeted liquid chromatography - mass spectrometry (LC-MS) and gas chromatography - mass spectrometry (GC-MS) on plasma samples collected from the portal vein of the SIM mice fed CHD, ZFD and HFD. To discriminate between host and microbial derived plasma metabolites we included as control plasma samples from GF mice fed CHD, ZFD and HFD. We identified 17 metabolites associated with the SIM microbiota that significantly changed in the plasma metabolome of the SIM mice in respond to changes in diet. Interestingly, those metabolites have been linked mostly to lipid and amino acids metabolism, which suggest that introduction of SIM microbiota to the GF mice induced metabolic consequences to the host beyond the processing of polysaccharides.

3.3. Quantifying the metabolic changes of host-microbes

3.3.1. Paper V: Elucidating the interactions between human gut microbiota and host through metabolic modelling

The symbiosis of host and gut microbiota were described in the introduction section. Here, we will discuss the using metabolic modeling of cell/tissue types in the human body and integration of specific models with gut microbes. The generic human metabolic models were initiated by Edinburgh human metabolic model (EHMN) (Ma et al., 2007) and *Recon1* (Duarte et al., 2007) which mainly reconstructed based on bibliomic and includes the collection of reactions occur in human. After this more detailed and accurate models were reconstructed, like the human metabolic reaction (HMR) (Agren et al., 2012) and *Recon2* (Thiele et al., 2013).These generic models were applied to understand better the human metabolism through mainly the network dependent analysis and data integration. Taking to account that metabolism each tissue in the human body is different, looking to tissue specific models could make lot of sense for data integration and hypothesis generation. It is therefore different algorithm were developed to reconstruct the cell or tissue specific models from generic model through using transcriptome and proteome data from the Human Protein Atlas (Agren et al., 2012; Fagerberg et al., 2014; Uhlen et al., 2010). These algorithms have been extensively reviewed in (Mardinoglu and Nielsen, 2015)

Integration of relevant clinical data to the specific model made it possible for new hypothesis, therapies and design of new interventions. As well the metabolic modeling the interactions between tissues provide better understanding of different metabolic disorders and physiology of body. Towards this intercellular interactions between adipocytes, hepatocytes, and myocytes as the metabolically active human cell types has been studied (Bordbar et al., 2011). These interactive modeling were applied to study the metabolic variations and reaction activities.

To this end, metabolic modeling of gut microbe-host symbiosis can as well help for better understanding the mechanisms and effect of microbes on human host (Figure 18). The interactions between the microbes, small intestine and colon can reveal the effect of microbiota on the metabolism of epithelial and enterocyte cells (Sahoo and Thiele, 2013). A metabolic model for small intestinal enterocytes has been generated and observed that metabolism of carbohydrates, amino acid, dietary fibers, and lipids were most occurred ones. Using the model the effect of american and balanced diet was studied. Through interacting one abundant gut microbe with the mouse generic model the systems-level characterization of a host-microbe metabolic symbiosis has been studied (Heinken et al., 2013).

Considering the interaction between three bacteria in the gut from Paper I, the main products of the bacterial community are SCFAs: acetate, propionate, and butyrate. SCFAs are mainly absorbed by epithelial cells. The nutrient enters to the gastro0intestinal tract in categorizes of carbohydrates, proteins, and fats which some digested by host enzymes and indigestible ones by the microbiota. The available SCFAs in the portal vein can be taken up by liver. SCFAs can have impact on the mechanism of hepatocyte cells that regulate cholesterol levels by synthesizing primary bile acids and lipoproteins [chylomicrons, very low-density lipoprotein (VLDL), low-density lipoprotein (LDL), and high-density lipoprotein (HDL)] (Gabert et al., 2011). There are as well cross talks between myocyte, adipocyte and hepatocyte which each of them can get impact from the secreted products of microbes to portal vein. Investigation on interactions between organs is vital to overcome the complexity of metabolic modeling of host-microbe interactions.

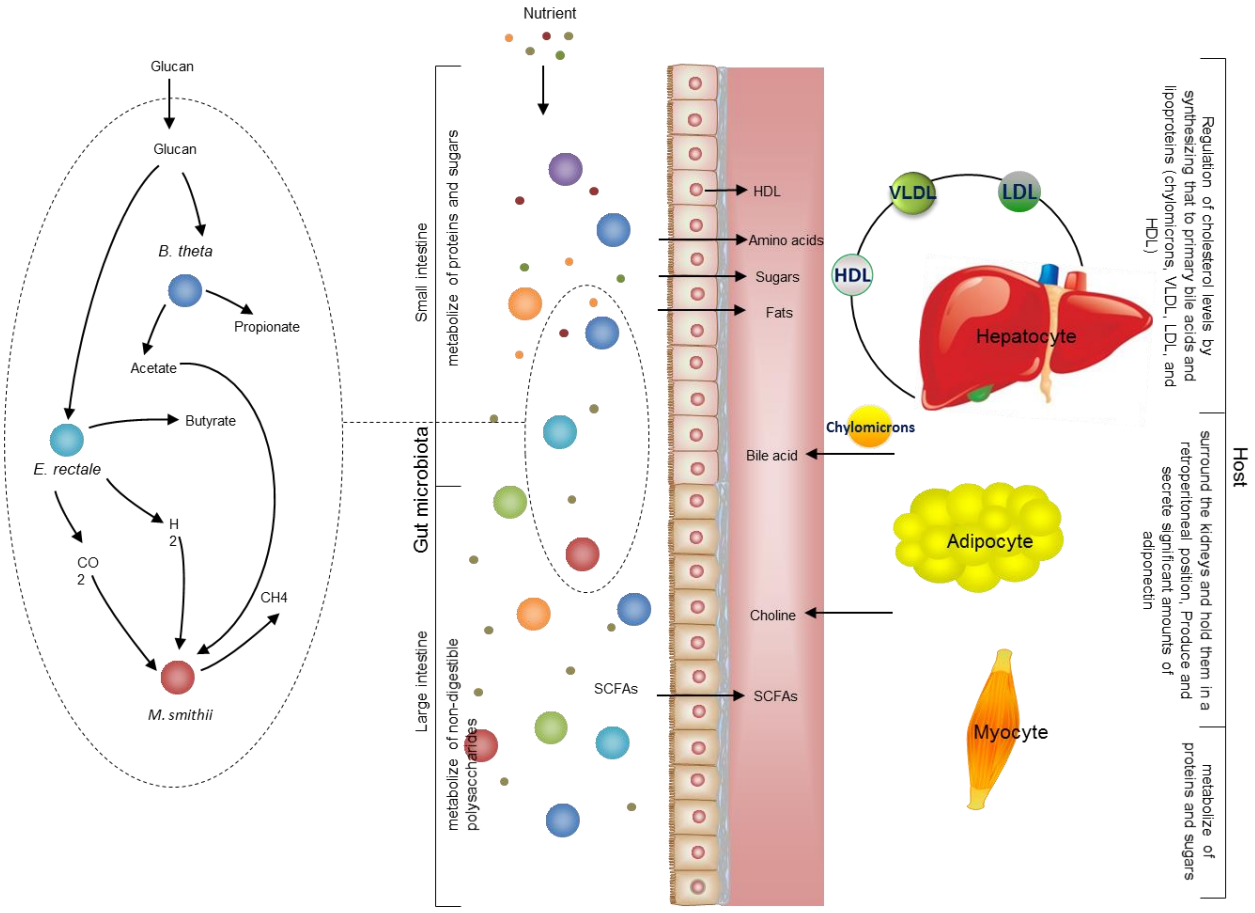


Figure 17 Simplified interactions between three bacteria in gut microbiota and their crosstalk with other tissues.

3.3.2. Paper IV: Revealing the metabolic differences between germ free and conventionally raised mice through metabolic modeling of gut microbiota and its host

Complex disorders such as obesity (Backhed et al., 2007; Turnbaugh et al., 2006), type 2 diabetes (T2D) (Karlsson et al., 2013), atherosclerosis (Karlsson et al., 2012), non-alcoholic fatty liver disease (NAFLD) (Henao-Mejia et al., 2012) and malnutrition, which appears at the opposite end of the spectrum (Smith et al., 2013; Subramanian et al., 2014), have been associated with the imbalances in the human gut microbiota. In order to gain a mechanistic insights about the contribution of the specific microbial populations in the development of such disorders, germ-free (GF) animals have been adopted for studying the association of the gut microbiota to disease pathogenesis (Ridaura et al., 2013). GF mice has been broadly used for investigating the effect of gut microbiota on host physiology. The differences between the GF and conventionally raised (CR) mice have been studied (Claus et al., 2008; El Aidy et al., 2013; Slack et al., 2009; Stappenbeck et al., 2002; Wostmann, 1981). Moreover, Larsson et al (Larsson et al., 2012) studied the response of the host induced by microbes along the gut in CR and GF C57Bl6/J mice and provided a detailed description for tissue-specific host transcriptional responses.

Here, in order to investigate the gut microbiota-induced transcriptional responses of the host metabolism, CR and GF C57Bl6/J mice were freely fed with autoclaved chow diet and mice were killed at 12 weeks of age. Small intestine and colon were removed and RNA was isolated from duodenum, jejunum, ileum segments of the small intestine, the proximal piece of the colon, liver, epididymal fat and subcutaneous fat, and microarray gene expression data was generated for these seven tissues. We identified significantly differentially expressed probe sets and metabolic genes by comparing gene expression profiles of tissues obtained from CR versus GF mice (Figure 19).

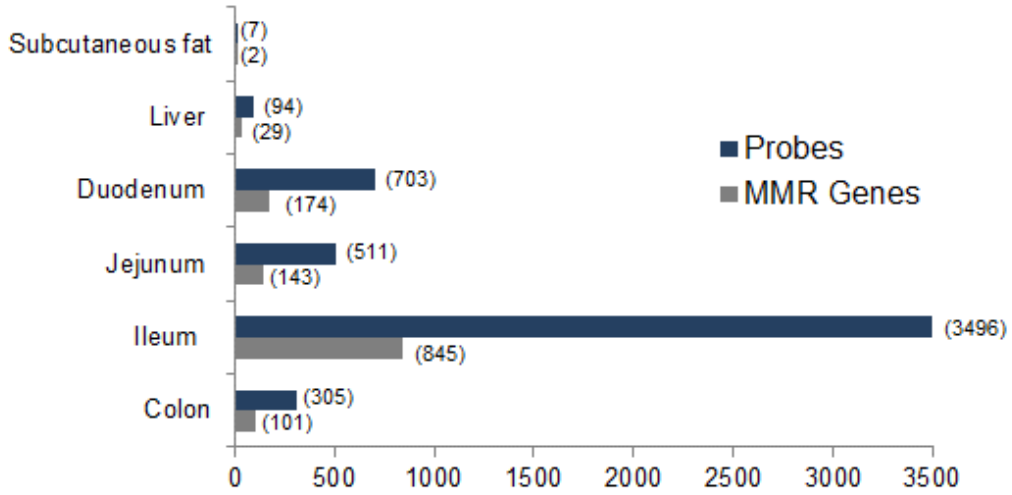


Figure 18 Gene expression profiling of different tissues including duodenum, jejunum, ileum, colon liver, and subcutaneous fat, tissues obtained from both CR and GF mice obtained from CR and GF mice. Gene expression data for each tissue has been normalized independent than other tissues and significantly (Q-value<0.05) differentially expressed probe sets and metabolic genes in Mouse Metabolic Reaction database was presented in each analyzed tissue.

During the identification of the significantly (Q-value<0.05) differentially expressed probe sets and metabolic genes, we corrected p-values for multiple testing by the Benjamini Hochberg method and calculated Q-values. Figure 20 shows the comparison between the differentially expressed metabolic genes of duodenum, jejunum, ileum, colon and liver tissues of CR and GF mice. Two genes were observed that differentially expressed in all tissues of CR compared to GF mice. Expression of Nicotinamide nucleotide transhydrogenase (Nnt) gene was higher and Ectonucleoside triphosphate diphosphohydrolase 4 (Entpd4) was lower in all five tissues of CR mice compared to GF mice.

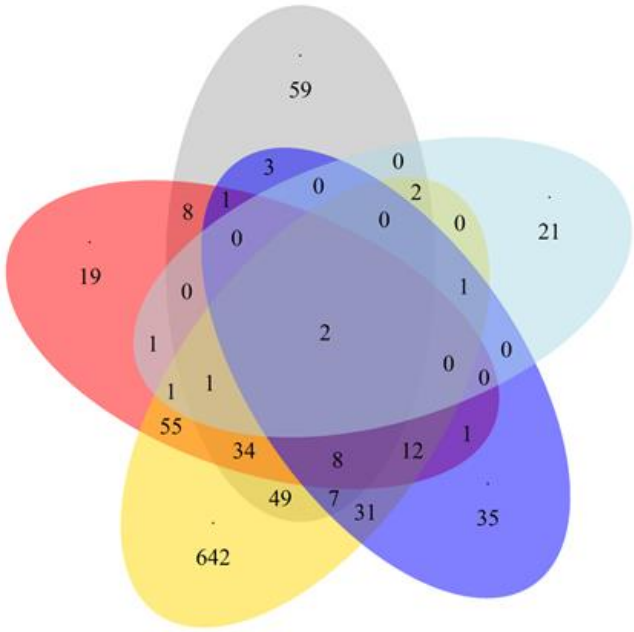


Figure 19 The overlap between the significantly (Q-value<0.05) differentially expressed metabolic genes in duodenum, jejunum, ileum, colon and liver are presented. The significantly (Q-value<0.05) differentially expressed metabolic genes, Nnt and Entpd4 as well as the associated reactions to Nnt is presented.

In the next step we developed MMR by using the mouse orthologs of human genes in HMR2.0. The resulting generic model includes 8,140 reactions, 3,579 associated metabolic genes to those reactions and 5,990 metabolites in eight different subcellular compartments. Previously, SILAC based proteomics data have been generated to analyze the expression of 7,349 proteins in 28 different major C57BL/6 mice tissues (Geiger et al., 2013) and it covers the 2,030 of the protein-coding genes in MMR. We reconstructed 28 tissue-specific GEMs, using proteomics data, MMR and tINIT algorithm (Agren et al., 2014). A total of 5,813 reactions, 4,574 metabolites and 1,838 genes were shared across the tissue-specific GEMs of which 2,750 (47.3%) reactions, 3,001 (65.6%) metabolites and 669 (36.4%) genes were common to all tissue-specific GEMs. GEMs for liver (*iMiceLiver*), adipose (*iMiceAdipose*), colon (*iMiceColon*) and small intestine (*iMiceSmallintestine*) by merging the GEMs for duodenum, jejunum and ileum tissues were incorporated with the differentially expressed genes between CR and GF mice tissues.

Using network structure provided by *iMiceSmallintestine* and gene expression profiling of the small intestine segments (duodenum, jejunum and ileum) between CR and GF mice, we investigated on the changes in the expression of the genes interacting with Nnt (Figure 21A). We found that the expression of Glutathione reductase (Gsr) which uses NADPH as an electron source to reduce the glutathione disulfide (GSSG) back to GSH is also significantly (Q-value<0.05) higher in all three small intestine segments of CR mice compared to GF mice (Figure 21). The decreased synthesis of the GSH in the small intestine segments of CR mice may be due to the limited availability of the substrates including glutamate, cysteine and glycine. Eventually, we identified significantly (Q-value<0.05) differentially expressed genes linked to glycine (Figures 21B) and glutamate (Figures 21C) and found that there are metabolic differences in the utilization of these AAs between CR and GF mice.

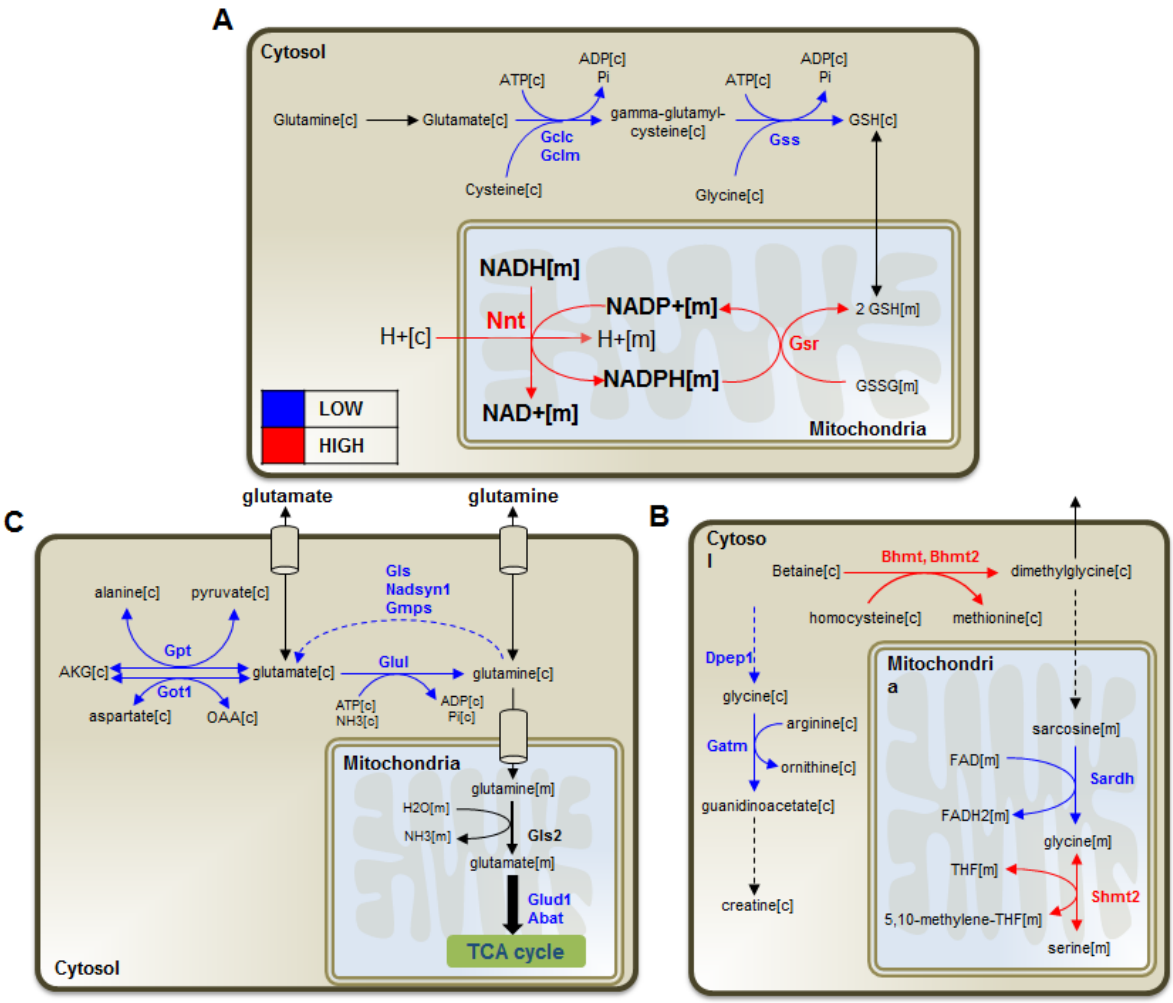


Figure 20 Metabolic differences in the small intestine tissues of CONV-R and GF mice. A) Metabolic genes together with the associated reactions involved in the formation of glutathione (GSH) are shown. Significant changes associated with the B) glycine and C) glutamine are represented.

Based on gene expression data, Nnt was significantly higher and Entpd4 was significantly lower expressed in the liver tissue of CR compared to GF mice based on microarray gene expression data. Expression levels of Nnt and Entpd4 in liver tissue obtained from CR and GF mice were validated by quantitative reverse transcription PCR. Higher Nnt expression in CR mice can be described with the response of liver tissue to the lower level of glycine required for the *de novo* synthesis of the GSH. HPV level of serine was measured, which can be utilized by the liver and converted to the glycine, and we observed that the level of serine was also significantly lower in CR mice compared to GF mice.

In the colon tissue of CR mice the expression of the Gsta4, Gstk1, Gstp1 and Gstt1 that reduce GSH to oxidized glutathione is significantly high (Q-value<0.05) (Figure 22). Metabolic differences between the colon tissue of CR and GF mice by integrating the significantly differentially expressed genes to *iMiceColon* were also investigated. Accordingly, it was found expression of Arg2 involved in arginine metabolism as well as Ces1g, Aldh1a2 and Rbp4 involved in vitamin A metabolism is higher in the colon tissue of CR mice compared to GF mice. On the other hand, we found that Aldob and Aldh9a1 involved in glycolysis, Hmgcs1, Hsd17b7, Nsdhl and Sc4mol involved in cholesterol synthesis, Mgam and Sis involved in starch and sucrose

metabolism, Slc2a5, Slc2a9, Sord and Khk involved in fructose metabolism as well as Ace2 transcription factor involved in th conversion of angiotensin were significantly lower in CR mice (Figure 22). Additionally, we found that the expression of the genes involved in the transport of AAs are significantly lower in the colon tissue of CR compared to GF mice.

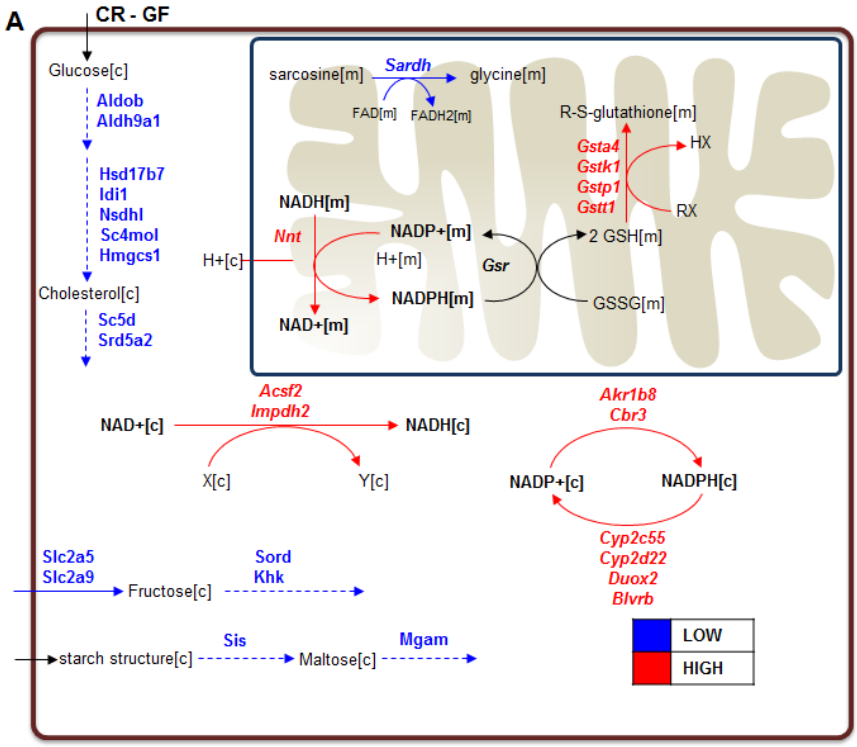


Figure 21 Metabolic differences in the colon tissue of CONV-R and GF mice.

To understand the metabolic interactions between the gut microbiota as well as their interactions with the small intestine in CONV-R mice, we simulated the crosstalk between the two key species *B. thetaiotamicron* and *E. rectale*. We used the content of the autoclaved chow diet and considered assumptions regarding the utilization of the diet; maximum 40% of the total protein may be consumed by the bacteria (set as upper bound), 5% of the total protein is transferred to the colon tissue and the remaining proteins are consumed by the small intestine (Figure 23). We also assumed that 40% of the digestible and 5% of the non-digestible carbohydrates in the diet is also consumed by the bacteria in small intestine and 5% of the digestible carbohydrates is transferred to the colon tissue (Gibson and Roberfroid, 1995).

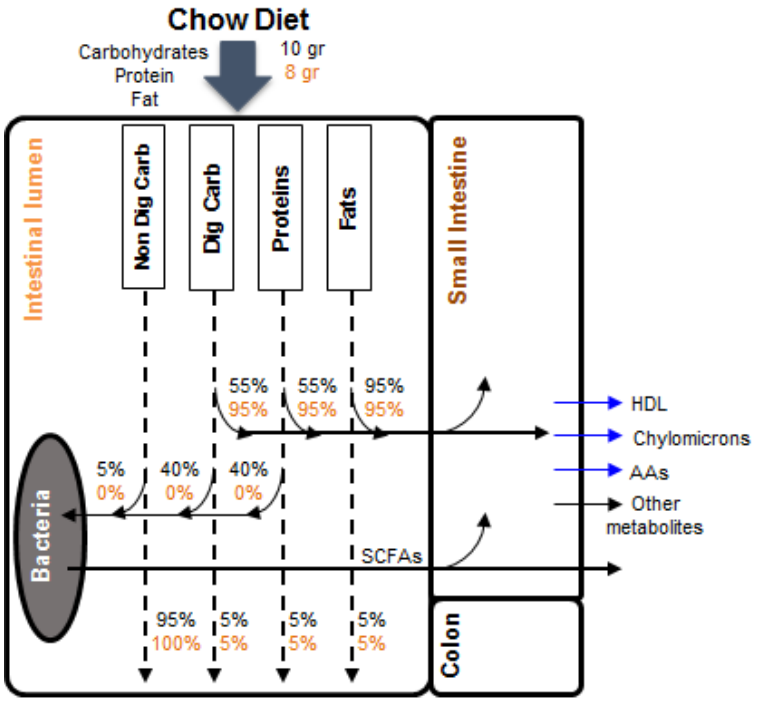


Figure 22 The differences in the utilization of proteins, carbohydrates, non-digestible carbohydrates and fates between the CR and GF mice is presented.

We simulated the interactions and predict the utilization of AAs and secretion of SCFAs by the bacteria. Simulations predicts isoleucine, proline and valine are only consumed by the *E. rectale*, glycine, serine, alanine, cystine, glutamate, histidine, leucine, lysine, methionine, phenylalanine, threonine and tyrosine are utilized by both *B. thetaiotamicron* and *E. rectale* whereas arginine, aspartate and tryptophan are not consumed by neither of these bacteria (Figure 24). Using the outcome of bacteria simulations, we studied the effect of SCFAs into the metabolic functions of small intestine in CR mice. We set up the production of chylomicrons and HDL in the small intestine of GF as objective function using the content of the diet. In order to produce chylomicrons and HDL together in the small intestine GEM, their lower bounds were constrained to 20% of the maximum production and performed flux variability analysis for all up and down regulated reactions based on the small intestine gene expression data. Next, we increased the upper and lower bounds of the reactions associated with significantly differentially expressed genes in GF mice and used these expanded bounds for the reactions in the small intestine model of CR mice.

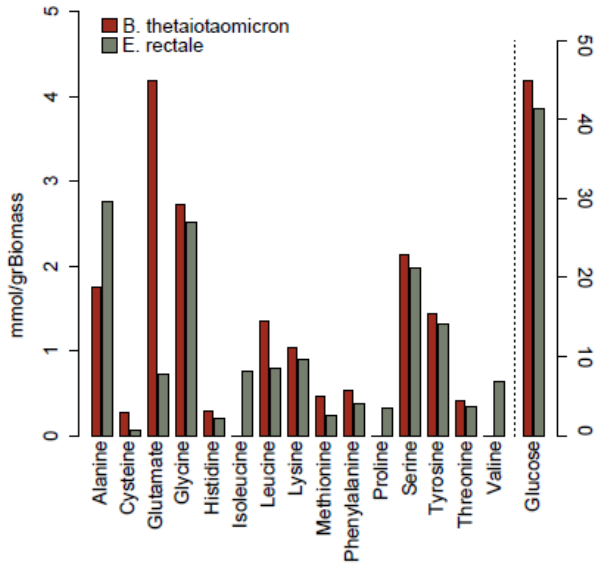


Figure 23 The amount of the glucose and AAs consumption for gut microbiota to optimize for their biomass and SCFAs production.

We optimized for the production the chylomicrons and eventually compared the amount of chylomicrons and HDL produced by CR and GF mice. SCFAs produced by the bacteria were used as an additional input to the *iMiceSmallintestine* in CR mice and we predicted that lower levels of chylomicrons and HDL are secreted by CR mice compared to GF mice. Through the use *iMiceSmallintestine,* calculated the set of flux distribution and ransom sampling algorithm (Bordel et al., 2010) we identified reactions involved in fatty acid biosynthesis and oxidations are transcriptionally downregulated in CONV-R compared to GF mice (Figure 25).

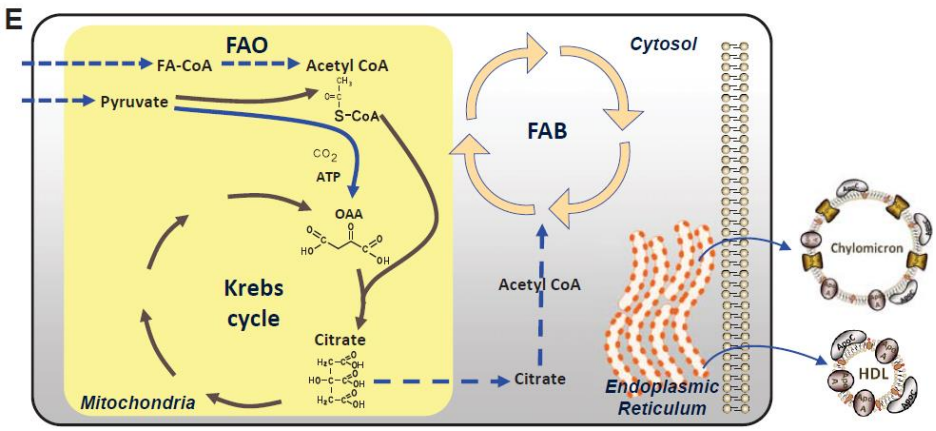


Figure 24 Transcriptionally up and downregulated reactions in CR and GF mice are identified.

4. Conclusion

In Paper I: We reconstructed relevant GEMs for the predominant phylum in the human gut, and we captured the main metabolic functions of the gut microbiome. We reconstructed GEMs of three key species; iBth1201 (*B. thetaiotaomicron*), iEre400 (E. rectale) and iMsi385 (M. smithii), which are relevant representatives of three phyla in the human gut (Bacteroidetes, Firmicutes and Euryarchaeota). We mathematically formalised two different scenarios to which the modeling of microbial communities can be applied. In the first case, the composition of the diet and the species abundances in the microbiota is known and constitute the input of the model. In this scenario, we aim to predict the profile of compounds produced by the microbiota and hence represent metabolites that can be taken up by the host. We refer to this simulation problem as the α-problem and a solution is found by minimising the substrate uptake rate. Alternatively we might be interested in predicting the abundances of the different species in the microbiota as a function of the diet. We refer to this problem as the β-problem. We further demonstrate that by integrating the topological information provided by GEMs with diet compositions, genomic and transcriptomic data, community metabolic modeling provides a mechanistic interpretation for statistical findings provided by metagenomics, and this leads to improved understanding of the relationships between diet, microbiota and disease and hereby enables a rational design of prebiotic and probiotic treatments.

In Paper II: We describe how the genome scale metabolic model (GEM) for the predominant human gut microbiota can elucidate the interactions between the human gut microbiota and its host. Focusing on metabolic interactions between the diet, gut microbiota and host metabolism, we hereby demonstrated that CASINO has excellent predictive power. As we demonstrated CASINO can also be used to predict dietary changes required in order to ensure a certain phenotype of the gut microbiome, here represented as a specific profile of consumption of 8 essential amino acids. CASINO could be used even further to simulate the impact of different diets on the production of different SCFAs and amino acids by the gut microbiome, and if a desired profile could be identified this would allow for computational evaluation of a suitable diet for subjects with a certain specified gut microbiome. As the gut microbiome composition is likely to change in response to changes in the diet, such predictions may have to be followed up with further metagenomics analysis and recalculations, but eventually, as more information is acquired about the impact of diet on gut microbiome composition, we would expect that CASINO can be used even to simulate how the gut microbiome may change in response to dietary changes, and hence a full predictive analysis can be performed. The modeling could quantitatively describe altered faecal amino acid levels in response to diet interventions. Our approach can also be used for identifying beneficial bacteria for human health and can be used for the treatment of metabolic disorders that are associated with gut microbiota. For instance, malnourished individuals who lack digestive enzymes due to the absence of certain microbes in their gut (Subramanian et al., 2014) can be determined and appropriate probiotics can be designed specifically for improving the overall metabolism in these subjects. We are therefore confident that CASINO will be a valuable tool for enriching the information content provided by gut metagenome analysis, and hereby advance our understanding of how this important metabolic organ contributes to disease development.

In Paper III: We developed a mouse model colonised with ten representatives of the human gut microbiota, with known function relevant to human health microbiota. Here, by means of different omics analyses, we study the response of simplified intestinal microbiota (SIM) to changes in the host diet. DNA from the jejunum, ileum, caecum, colon and faeces of SIM mice were isolated and

16S rRNA quantitative PCR was performed to understand the distribution of colonised microbes along the intestinal tract and faeces. RNA was isolated from the SIM mice and sequenced to reveal the metatranscription of SIM mice fed with an autoclaved chow diet. We also generated metatranscriptome data for SIM mice fed with a different diet and examine the response of the SIM microbes to changes in the host diet. We also generated metabolomics data on the plasma samples of SIM mice for different dietary patterns and compared them together with GF mice to uncover the changes in metabolites in the plasma of SIM and GF mice

In Paper IV & V: We created a generic Mouse Metabolic Reaction GEM (MMR) and generated tissue-specific GEMs for mice primarily based on proteomics data as well as the significantly differentially expressed genes between CONV-R and GF mice. We investigated the metabolic differences between CONV-R and GF mice using global gene expression profiling of the host tissues and the network topology provided by the tissue GEMs. Based on gene expression data, we found that the gut microbiota effect the host amino acid (AA) metabolism, which lead to modifications in glutathione metabolism. We validated our GEM-based predictions based on gene expression data by generating metabolomics data, and comparing the level of the metabolites in the HPV of the CONV-R and GF mice. Taken together, we found that the levels of the glycine and serine as well as the N-acetylated form of these two AAs that may be taken up by the liver tissue and used in GSH synthesis were significantly lower in the HPV of the CONV-R mice compared to GF mice. Hence, we observed that the expression of Nnt is increased in the liver tissue of CONV-R mice, potentially caused by the limited availability of the glycine used as a substrate in the GSH de novo synthesis. Moreover, we observed that the liver and colon tissue of CONV-R mice also responded to the lower level of glycine by higher expression of Nnt, and this indicated that the gut microbiota regulates AA metabolism not only in the small intestine but also in the liver and colon tissues. We investigated the global metabolic differences between the liver tissue of the CONV-R and GF mice, and found that the expression of the genes involved in the uptake of glucose, and biosynthesis of cholesterol and bile acids was significantly lower in CONV-R mice.

5. Future perspectives

Most current studies on human gut microbiome are focused on the metagenomics, metatranscriptomics and metaproteomics methods. Without prior information of the single microbes within the community, challenges in understanding the community still remain. This prior information plays a key role in bottom-up approaches. The concept of community systems biology is very much connected to progress in single cell sequencing studies, since the draft genome of single organism is a necessary input to construct detailed models, so they can accurately predict the single organism phenotype and hereby unravel their capabilities in the community. Another obstacle in the modeling approach that has to be addressed in the future is the spatial organization of microbial community that restricts the understanding of complex interaction between members of community. To overcome this, imaging mass spectrometry technique can be employed on microbial communities (Watrous and Dorrestein, 2011).

Despite all the varied applications of GEMs in biological studies, the field of reconstruction of GEM is still Imprecise. Most of the developed GEMs in different studies contain primary metabolism of the target organism and they hardly can be applied to different levels of systems biology studies. The proposed cellular objective is another limitation in CBM using GEMs. Most of the objectives are based on the simplification of the actual cellular process. For current studies the questions like “What kind of objective functions can give the best solution?” Which should be complemented by “what are the most relevant objectives and how can we evaluate them?” This question is more relevant when we want to perform CBM on tissue specific GEMs as well as the global objective function for the microbial communities and their interactions with their host. CASINO, OptCOm, COMMET and such methods are initial steps towards establishment of a workflow to predict deep capabilities of microbial communities and their interactions within the host.

Going back to the scope of this thesis and considering that one day we could construct comprehensive GEMs together with well-established methods and evaluate objective functions, it is clear that GEMs for additional species may have to be added in the future, in particular if one wants to study more specific bioconversions, e.g. related to bile acids and vitamins. It seems the future of the gut microbiome research field is pulled hardly towards identification of new possible drugs that can modulate the gut microbiome. Through applying drug discovery and different novel bioprocess engineering, our microbes and their mediated metabolites have become new targets for small molecule drugs. It will be important to develop workflows that can be employed to understand the effect on human health in different clinical conditions of the deletion or over expression of individual gene or gene sets in the gut microbiome, either though altering the abundance of species or the level of metabolites they produce. Here GEMs can assist and be used for design just as they have been recruited for metabolic engineering and the results presented in this thesis may therefore hopefully allow for new avenues in future modulation of the gut microbiome with the objective to develop new strategies for treating human diseases.

Acknowledgments

First, I would like to thank my supervisor Prof. Jens Nielsen for affording me the opportunity to be a young hope. Jens, I will never forget your continuous support of my Ph.D study research, your patience, motivation, and immense knowledge. Your guidance helped me in both professional and personal endeavours. I could not have imagined having a better advisor and mentor for my Ph.D study.

My sincere thanks also goes to my co-supervisor, Dr. Adil Mardinoglu. I am thankful for all the discussion, education and support, at times frustration could have easily derailed my efforts. You are great friend and mentor. This is only the beginning of a friendship that will last a life time. I would like to thank my dear friend Dr. Antonio Roldao. You have become the closest friend. Our trip to Portugal has become that of legends! Although you left two years ago, but during that time you had a great and positive impact on me. Thank you for your brotherhood.

I am forever indebted and thankful to Dr. Rasmus Ågren. You provided invaluable advice and guidance during supervising me as a master student and later as PhD student. It is often said your PhD research will be more defined by the people that surround you, it is therefore I am forever thankful to those that were fellow PhD students, Postdocs, Senior Researchers and Faculty at the time: Amir Feizi, Pouyan Ghaffari, Leif Väremo, Francesco Gatto, Klaas Buijs, Mingtao Huang, Stefan Tippmann, Dr. Boyang Ji, Dr. José L. Martínez, Dr. Clara Navarrete, Dr. Partho Sarathi Sen, Dr. Christoph Knuf, Dr. Siavash Partow, Dr. Fredrik Karlsson, Dr. Tobias Österlund and Dr. Kuk-ki Hong, Dr. Luis Caspeta-Guadarrama, Dr. Il-Kwon Kim, Dr. Verena Siewers, Dr. Yun Chen, Dr. Sakda Khoomrung and Dr. Ali Kazemi.

I would like as well to thank Dr. Dina Petranovic, Dr. Fredrik Karlsson, Dr. Sergio Bordel, and Dr. Intawat Nookaew for all the support and discussion during my research. I had the pleasure to collaborate with several scientist around the world. I would like to thank Prof. Fredrik Bäckhed, Dr. Petia Kovatcheva-Datchary (Wallenberg lab, GU) and Prof. Karine Clément (ICAN). I had a great two weeks in France as a visiting PhD student. Thanks to Dr. Bridget Holmes (Danone) and Prof. Karine Clément for hosting me.

I want to give a warm thank you to Erica Dahlin and Martina Butorac for helping me when I was almost lost in the lab and needed help with all sort of practicalities. There are many more collaborators, students, colleagues and friends that could be included here by name. Know that my gratitude extends to all of you. In addition, I would like to thank the people whose minds and wits and laughs have made this whole thing the fantastic experience that it has been: Amir, Pouyan, Aida, Siavash, Ali, Adil, Antonio, Martina, Luis, Christoph, Tobias, and many more.

At the onset, I dedicate this doctoral thesis to my love, Azadeh, my family, Maman, Baba and Sahar. Maman & Baba I know I have been away for more than 8 years and you had tough days without me, I can only hope that this accomplishment reaffirms the values you have instilled in me. Azadeh, thanks for your unwavering support, constant love and understanding. I know we have had stressful moments couple of last months, but all is going to be alright, alright, alright!

References

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Comput Biol *8*, e1002518.

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., and Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput Biol *9*, e1002980.

Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., and Nielsen, J. (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol *10*, 721.

Aiba, S., and Matsuoka, M. (1979). Identification of Metabolic Model - Citrate Production from Glucose by Candida-Lipolytica. Biotechnol Bioeng *21*, 1373-1386.

Aminov, R.I., Walker, A.W., Duncan, S.H., Harmsen, H.J.M., Welling, G.W., and Flint, H.J. (2006). Molecular diversity, cultivation, and improved detection by fluorescent in situ hybridization of a dominant group of human gut bacteria related to Roseburia spp. or Eubacterium rectale. Appl Environ Microb *72*, 6371-6376.

Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y., and Tomita, M. (2006). GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. BMC Bioinformatics *7*, 168.

Armougom, F., Henry, M., Vialettes, B., Raccah, D., and Raoult, D. (2009). Monitoring Bacterial Community of Human Gut Microbiota Reveals an Increase in Lactobacillus in Obese Patients and Methanogens in Anorexic Patients. Plos One *4*.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., et al. (2011). Enterotypes of the human gut microbiome. Nature *473*, 174-180.

Atherton, H.J., Bailey, N.J., Zhang, W., Taylor, J., Major, H., Shockcor, J., Clarke, K., and Griffin, J.L. (2006). A combined 1H-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR-alpha null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome. Physiological genomics *27*, 178-186.

Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. Science *307*, 1915-1920.

Backhed, F., Manchester, J.K., Semenkovich, C.F., and Gordon, J.I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. P Natl Acad Sci USA *104*, 979-984.

Baldrian, P., Kolarik, M., Stursova, M., Kopecky, J., Valaskova, V., Vetrovsky, T., Zifcakova, L., Snajdr, J., Ridl, J., Vlcek, C., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. Isme J *6*, 248-258.

Beard, D.A., Liang, S.D., and Qian, H. (2002). Energy balance for analysis of complex metabolic networks. Biophys J *83*, 79-86.

Becker, S.A., and Palsson, B.O. (2008). Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol *4*, e1000082.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., Nikkila, J., Monti, D., Satokari, R., Franceschi, C., et al. (2010). Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians. Plos One *5*.

Bianucci, M., Mannella, R., West, B.J., and Grigolini, P. (1995). From Dynamics to Thermodynamics - Linear-Response and Statistical-Mechanics. Phys Rev E *51*, 3002-3022.

Bingham, S.A., Pignatelli, B., Pollock, J.R., Ellul, A., Malaveille, C., Gross, G., Runswick, S., Cummings, J.H., and O'Neill, I.K. (1996). Does increased endogenous formation of N-nitroso compounds in the human colon explain the association between red meat and colon cancer? Carcinogenesis *17*, 515-523.

Bordbar, A., Feist, A.M., Usaite-Black, R., Woodcock, J., Palsson, B.O., and Famili, I. (2011). A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. BMC systems biology *5*, 180.

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. PLoS Comput Biol *6*, e1000859.
Borenstein, E., Kupiec, M., Feldman, M.W., and Ruppin, E. (2008). Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci U S A *105*, 14482-14487.
Burgard, A.P., and Maranas, C.D. (2003). Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng *82*, 670-677.
Byrd, A.L., and Segre, J.A. (2015). Integrating host gene expression and the microbiome to explore disease pathogenesis. Genome Biol *16*, 70.
Chaffron, S., Rehrauer, H., Pernthaler, J., and von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res *20*, 947-959.
Chambers, A.F., Groom, A.C., and MacDonald, I.C. (2002). Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer *2*, 563-572.
Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011). Bacterial communities of diverse Drosophila species: ecological context of a host-microbe model system. PLoS Genet *7*, e1002272.
Christl, S.U., Murgatroyd, P.R., Gibson, G.R., and Cummings, J.H. (1992). Production, metabolism, and excretion of hydrogen in the large intestine. Gastroenterology *102*, 1269-1277.
Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M.B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., et al. (2012). Gut microbiota composition correlates with diet and health in the elderly. Nature *488*, 178-+.
Claus, S.P., Tsang, T.M., Wang, Y., Cloarec, O., Skordi, E., Martin, F.P., Rezzi, S., Ross, A., Kochhar, S., Holmes, E., et al. (2008). Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. Mol Syst Biol *4*, 219.
Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandezgarayzabal, J., Garcia, P., Cai, J., Hippe, H., and Farrow, J.A.E. (1994). The Phylogeny of the Genus Clostridium - Proposal of 5 New Genera and 11 New Species Combinations. Int J Syst Bacteriol *44*, 812-826.
Comalada, M., Bailon, E., de Haro, O., Lara-Villoslada, F., Xaus, J., Zarzuelo, A., and Galvez, J. (2006). The effects of short-chain fatty acids on colon epithelial proliferation and survival depend on the cellular phenotype. J Cancer Res Clin Oncol *132*, 487-497.
Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial Community Variation in Human Body Habitats Across Space and Time. Science *326*, 1694-1697.
Cotillard, A., Kennedy, S.P., Kong, L.C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M., Quinquis, B., Levenez, F., Galleron, N., et al. (2013). Dietary intervention impact on gut microbial gene richness. Nature *500*, 585-588.
Cottret, L., Milreu, P.V., Acuna, V., Marchetti-Spaccamela, A., Stougie, L., Charles, H., and Sagot, M.F. (2010). Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, Baumannia cicadellinicola and Sulcia muelleri, with their insect host, Homalodisca coagulata. PLoS Comput Biol *6*.
Covert, M.W., Schilling, C.H., and Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. J Theor Biol *213*, 73-88.
Cummings, J.H., and Englyst, H.N. (1987). Fermentation in the human large intestine and the available substrates. Am J Clin Nutr *45*, 1243-1255.
Cummings, J.H., Hill, M.J., Jivraj, T., Houston, H., Branch, W.J., and Jenkins, D.J. (1979). The effect of meat protein and dietary fiber on colonic function and metabolism. I. Changes in bowel habit, bile acid excretion, and calcium absorption. Am J Clin Nutr *32*, 2086-2093.
Davie, J.R. (2003). Inhibition of histone deacetylase activity by butyrate. J Nutr *133*, 2485S-2493S.
Dawson, P.A., Lan, T., and Rao, A. (2009). Bile acid transporters. J Lipid Res *50*, 2340-2357.
De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A *107*, 14691-14696.

Derrien, M., Vaughan, E., Plugge, C., and de Vos, W. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol *54*, 1469-1476.
Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. P Natl Acad Sci USA *108*, 4554-4561.
Donohoe, D.R., Garge, N., Zhang, X., Sun, W., O'Connell, T.M., Bunger, M.K., and Bultman, S.J. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell Metab *13*, 517-526.
Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., and Palsson, B.O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A *104*, 1777-1782.
Dumas, M.E., Barton, R.H., Toye, A., Cloarec, O., Blancher, C., Rothwell, A., Fearnside, J., Tatoud, R., Blanc, V., Lindon, J.C., et al. (2006). Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. Proc Natl Acad Sci U S A *103*, 12511-12516.
Duncan, S., and Flint, H. (2008). Proposal of a neotype strain (A1-86) for Eubacterium rectale. Request for an Opinion. Int J Syst Evol Microbiol *58*, 1735-1736.
Duncan, S.H., Aminov, R.I., Scott, K.P., Louis, P., Stanton, T.B., and Flint, H.J. (2006). Proposal of Roseburia faecis sp. nov., Roseburia hominis sp. nov. and Roseburia inulinivorans sp. nov., based on isolates from human faeces. Int J Syst Evol Microbiol *56*, 2437-2441.
Duncan, S.H., Louis, P., and Flint, H.J. (2004). Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl Environ Microbiol *70*, 5810-5817.
Durot, M., Bourguignon, P.Y., and Schachter, V. (2009). Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev *33*, 164-190.
Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. Science *308*, 1635-1638.
El-Semman, I.E., Karlsson, F.H., Shoaie, S., Nookaew, I., Soliman, T.H., and Nielsen, J. (2014). Genome-scale metabolic reconstructions of Bifidobacterium adolescentis L2-32 and Faecalibacterium prausnitzii A2-165 and their interaction. BMC systems biology *8*, 41.
El Aidy, S., Derrien, M., Merrifield, C.A., Levenez, F., Dore, J., Boekschoten, M.V., Dekker, J., Holmes, E., Zoetendal, E.G., van Baarlen, P., et al. (2013). Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon. The ISME journal *7*, 743-755.
Fagerberg, L., Hallstrom, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics *13*, 397-406.
Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J.Y., and Gordon, J.I. (2014). Identifying Gut Microbe-Host Phenotype Relationships Using Combinatorial Communities in Gnotobiotic Mice. Sci Transl Med *6*.
Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The Long-Term Stability of the Human Gut Microbiota. Science *341*, 44-+.
Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. Science *333*, 101-104.
Fell, D.A., and Small, J.R. (1986). Fat Synthesis in Adipose-Tissue - an Examination of Stoichiometric Constraints. Biochemical Journal *238*, 781-786.
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science *269*, 496-512.
Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nature reviews. Microbiology *6*, 121-131.

Frank, D.N., St Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A *104*, 13780-13785.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., and Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res *38*, 3857-3868.

Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., Gophna, U., Sharan, R., and Ruppin, E. (2011). Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun *2*, 589.

Gabert, L., Vors, C., Louche-Pelissier, C., Sauvinet, V., Lambert-Porcheron, S., Drai, J., Laville, M., Desage, M., and Michalski, M.C. (2011). 13C tracer recovery in human stools after digestion of a fat-rich meal labelled with [1,1,1-13C3]tripalmitin and [1,1,1-13C3]triolein. Rapid Commun Mass Spectrom *25*, 2697-2703.

Geiger, T., Velic, A., Macek, B., Lundberg, E., Kampf, C., Nagaraj, N., Uhlen, M., Cox, J., and Mann, M. (2013). Initial Quantitative Proteomic Map of 28 Mouse Tissues Using the SILAC Mouse. Molecular & Cellular Proteomics *12*, 1709-1722.

Gibson, G.R., Macfarlane, G.T., and Cummings, J.H. (1988). Occurrence of sulphate-reducing bacteria in human faeces and the relationship of dissimilatory sulphate reduction to methanogenesis in the large gut. J Appl Bacteriol *65*, 103-111.

Gibson, G.R., and Roberfroid, M.B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. The Journal of nutrition *125*, 1401-1412.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., and Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. Plos One *3*, e3042.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006). Metagenomic analysis of the human distal gut microbiome. Science *312*, 1355-1359.

Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. Cell *160*, 583-594.

Greenblum, S., Turnbaugh, P.J., and 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 U S A *109*, 594-599.

Groh, H., Schade, K., and Horhold-Schubert, C. (1993). Steroid metabolism with intestinal microorganisms. J Basic Microbiol *33*, 59-72.

Hamer, H.M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F.J., and Brummer, R.J. (2008). Review article: the role of butyrate on colonic function. Aliment Pharmacol Ther *27*, 104-119.

Hanfrey, C.C., Pearson, B.M., Hazeldine, S., Lee, J., Gaskin, D.J., Woster, P.M., Phillips, M.A., and Michael, A.J. (2011). Alternative spermidine biosynthetic route is critical for growth of Campylobacter jejuni and is the dominant polyamine pathway in human gut microbiota. J Biol Chem *286*, 43301-43312.

Hartman, A.L., Lough, D.M., Barupal, D.K., Fiehn, O., Fishbein, T., Zasloff, M., and Eisen, J.A. (2009). Human gut microbiome adopts an alternative state following small bowel transplantation. Proc Natl Acad Sci U S A *106*, 17187-17192.

Hayashi, H., Shibata, K., Sakamoto, M., Tomita, S., and Benno, Y. (2007). Prevotella copri sp. nov. and Prevotella stercorea sp. nov., isolated from human faeces. Int J Syst Evol Microbiol *57*, 941-946.

Heinken, A., Sahoo, S., Fleming, R.M., and Thiele, I. (2013). Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. Gut microbes *4*, 28-40.

Heinrich, R., Rapoport, S.M., and Rapoport, T.A. (1977). Metabolic regulation and mathematical models. Prog Biophys Mol Biol *32*, 1-82.

Henao-Mejia, J., Elinav, E., Jin, C., Hao, L., Mehal, W.Z., Strowig, T., Thaiss, C.A., Kau, A.L., Eisenbarth, S.C., Jurczak, M.J., et al. (2012). Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature *482*, 179-185.

Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., and Stevens, R.L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol *28*, 977-982.

Hughes, R., Magee, E.A., and Bingham, S. (2000). Protein degradation in the large intestine: relevance to colorectal cancer. Curr Issues Intest Microbiol *1*, 51-58.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207-214.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. Nature *407*, 651-654.

Jindou, S., Brulc, J.M., Levy-Assaraf, M., Rincon, M.T., Flint, H.J., Berg, M.E., Wilson, M.K., White, B.A., Bayer, E.A., Lamed, R., et al. (2008). Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens. FEMS Microbiol Lett *285*, 188-194.

Jorth, P., Turner, K.H., Gumus, P., Nizam, N., Buduneli, N., and Whiteley, M. (2014). Metatranscriptomics of the human oral microbiome during health and disease. MBio *5*, e01012-01014.

Kageyama, A., and Benno, Y. (2000). Emendation of genus Collinsella and proposal of Collinsella stercoris sp. nov. and Collinsella intestinalis sp. nov. Int J Syst Evol Microbiol *50*, 1767-1774.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. Nucleic Acids Res *32*, D277-280.

Karlsson, F.H., Fak, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., Backhed, F., and Nielsen, J. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun *3*, 1245.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Backhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature *498*, 99-103.

Karlsson, F.H., Ussery, D.W., Nielsen, J., and Nookaew, I. (2011). A Closer Look at Bacteroides: Phylogenetic Relationship and Genomic Implications of a Life in the Human Gut. Microb Ecol *61*, 473-485.

Karp, R.M. (2008). George Dantzig's impact on the theory of computation. Discrete Optim *5*, 174-185.

Kinross, J.M., Darzi, A.W., and Nicholson, J.K. (2011). Gut microbiome-host interactions in health and disease. Genome medicine *3*, 14.

Kitano, H. (2002a). Computational systems biology. Nature *420*, 206-210.

Kitano, H. (2002b). Systems biology: a brief overview. Science *295*, 1662-1664.

Knorr, A.L., Jain, R., and Srivastava, R. (2007). Bayesian-based selection of metabolic objective functions. Bioinformatics *23*, 351-357.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A *108 Suppl 1*, 4578-4585.

Kovatcheva-Datchary, P., Tremaroli, V., and Bäckhed, F. (2013). The Gut Microbiota. In The Prokaryotes. E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Springer Berlin Heidelberg), pp. 3-24.

Lakhan, S.E., and Kirchgessner, A. (2010). Gut inflammation in chronic fatigue syndrome. Nutr Metab (Lond) *7*, 79.

Larsson, E., Tremaroli, V., Lee, Y.S., Koren, O., Nookaew, I., Fricker, A., Nielsen, J., Ley, R.E., and Backhed, F. (2012). Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. Gut *61*, 1124-1131.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541-546.

Lewis, N.E., Nagarajan, H., and Palsson, B.O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature Reviews Microbiology *10*, 291-305.

Levy, R., and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc Natl Acad Sci U S A *110*, 12804-12809.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Prifti, E., Nielsen, T., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol *32*, 834-841.

Louis, P., Duncan, S.H., McCrae, S.I., Millar, J., Jackson, M.S., and Flint, H.J. (2004). Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. J Bacteriol *186*, 2099-2106.

Lun, D.S., Rockwell, G., Guido, N.J., Baym, M., Kelner, J.A., Berger, B., Galagan, J.E., and Church, G.M. (2009). Large-scale identification of genetic design strategies using local search. Mol Syst Biol *5*.

Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., and Goryanin, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol *3*, 135.

Macfarlane, G.T., Cummings, J.H., and Allison, C. (1986). Protein degradation by human intestinal bacteria. J Gen Microbiol *132*, 1647-1656.

Macfarlane, G.T., Gibson, G.R., and Cummings, J.H. (1992). Comparison of fermentation reactions in different regions of the human colon. J Appl Bacteriol *72*, 57-64.

Mahowald, M.A., Rey, F.E., Seedorf, H., Turnbaugh, P.J., Fulton, R.S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R.K., et al. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci U S A *106*, 5859-5864.

Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., and Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun *5*, 3083.

Mardinoglu, A., and Nielsen, J. (2015). New paradigms for metabolic modeling of human cells. Curr Opin Biotechnol *34*, 91-97.

Marquet, P., Duncan, S.H., Chassard, C., Bernalier-Donadille, A., and Flint, H.J. (2009). Lactate has the potential to promote hydrogen sulphide formation in the human colon. FEMS Microbiology Letters *299*, 128-134.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host & Microbe *4*, 447-457.

Martin, F.P., Sprenger, N., Montoliu, I., Rezzi, S., Kochhar, S., and Nicholson, J.K. (2010). Dietary modulation of gut functional ecology studied by fecal metabonomics. J Proteome Res *9*, 5284-5295.

Matsumoto, M., and Benno, Y. (2007). The relationship between microbiota and polyamine concentration in the human intestine: a pilot study. Microbiol Immunol *51*, 25-35.

McNeil, N.I. (1984). The contribution of the large intestine to energy supplies in man. Am J Clin Nutr *39*, 338-342.

McNeil, N.I., Cummings, J.H., and James, W.P. (1978). Short chain fatty acid absorption by the human large intestine. Gut *19*, 819-822.

Million, M., Maraninchi, M., Henry, M., Armougom, F., Richet, H., Carrieri, P., Valero, R., Raccah, D., Vialettes, B., and Raoult, D. (2012). Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. Int J Obesity *36*, 817-825.

Mouse Genome Sequencing, C., Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Muir, J.G., Yeow, E.G., Keogh, J., Pizzey, C., Bird, A.R., Sharpe, K., O'Dea, K., and Macrae, F.A. (2004). Combining wheat bran with resistant starch has more beneficial effects on fecal indexes than does wheat bran alone. Am J Clin Nutr *79*, 1020-1028.

Notebaart, R.A., van Enckevort, F.H., Francke, C., Siezen, R.J., and Teusink, B. (2006). Accelerating the reconstruction of genome-scale metabolic networks. BMC Bioinformatics *7*, 296.

Oliveira, A.P., Nielsen, J., and Forster, J. (2005). Modeling Lactococcus lactis using a genome-scale flux model. Bmc Microbiol *5*.

Othmer, H.G. (1976). The qualitative dynamics of a class of biochemical control circuits. J Math Biol *3*, 53-78.

Ouwehand, A.C., Derrien, M., de Vos, W., Tiihonen, K., and Rautonen, N. (2005). Prebiotics and other microbial substrates for gut functionality. Curr Opin Biotechnol *16*, 212-217.

Patil, K.R., and Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci U S A *102*, 2685-2689.

Patil, K.R., Rocha, I., Forster, J., and Nielsen, J. (2005). Evolutionary programming as a platform for in silico metabolic engineering. Bmc Bioinformatics *6*.

Pharkya, P., Burgard, A.P., and Maranas, C.D. (2004). OptStrain: a computational framework for redesign of microbial production systems. Genome Res *14*, 2367-2376.

Pharkya, P., and Maranas, C.D. (2006). An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng *8*, 1-13.

Qin, J.J., Li, R.Q., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59-U70.

Ramakrishna, R., Edwards, J.S., McCulloch, A., and Palsson, B.O. (2001). Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am J Physiol Regul Integr Comp Physiol *280*, R695-704.

Rey, F.E., Gonzalez, M.D., Cheng, J., Wu, M., Ahern, P.P., and Gordon, J.I. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. P Natl Acad Sci USA *110*, 13582-13587.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Ridlon, J.M., Kang, D.J., and Hylemon, P.B. (2006). Bile salt biotransformations by human intestinal bacteria. J Lipid Res *47*, 241-259.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K., and Rawls, J.F. (2011). Evidence for a core gut microbiota in the zebrafish. Isme J *5*, 1595-1608.

Rossi, M., Corradini, C., Amaretti, A., Nicolini, M., Pompei, A., Zanoni, S., and Matteuzzi, D. (2005). Fermentation of fructooligosaccharides and inulin by bifidobacteria: a comparative study of pure and fecal cultures. Appl Environ Microbiol *71*, 6150-6158.

Ruppin, H., Bar-Meir, S., Soergel, K.H., Wood, C.M., and Schmitt, M.G., Jr. (1980). Absorption of short-chain fatty acids by the colon. Gastroenterology *78*, 1500-1507.

Sahoo, S., and Thiele, I. (2013). Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells. Hum Mol Genet *22*, 2705-2722.

Said, H.M. (2011). Intestinal absorption of water-soluble vitamins in health and disease. Biochem J *437*, 357-372.

Salonen, A., Nikkila, J., Jalanka-Tuovinen, J., Immonen, O., Rajilic-Stojanovic, M., Kekkonen, R.A., Palva, A., and de Vos, W.M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods *81*, 127-134.

Samuel, B.S., and Gordon, J.I. (2006). A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci U S A *103*, 10011-10016.
Samuel, B.S., Hansen, E.E., Manchester, J.K., Coutinho, P.M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R.K., and Gordon, J.I. (2007). Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. Proc Natl Acad Sci U S A *104*, 10643-10648.
Samuel, B.S., Shaito, A., Motoike, T., Rey, F.E., Backhed, F., Manchester, J.K., Hammer, R.E., Williams, S.C., Crowley, J., Yanagisawa, M., et al. (2008). Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. Proc Natl Acad Sci U S A *105*, 16767-16772.
Sauer, U., Heinemann, M., and Zamboni, N. (2007). Genetics. Getting closer to the whole picture. Science *316*, 550-551.
Scheppach, W. (1994). Effects of short chain fatty acids on gut morphology and function. Gut *35*, S35-38.
Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012). Multidimensional optimality of microbial metabolism. Science *336*, 601-604.
Scott, K.P., Gratz, S.W., Sheridan, P.O., Flint, H.J., and Duncan, S.H. (2013). The influence of diet on the gut microbiota. Pharmacol Res *69*, 52-60.
Scott, K.P., Martin, J.C., Campbell, G., Mayer, C.D., and Flint, H.J. (2006). Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans". J Bacteriol *188*, 4340-4349.
Segre, D., Vitkup, D., and Church, G.M. (2002). Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A *99*, 15112-15117.
Shlomi, T., Berkman, O., and Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci U S A *102*, 7695-7700.
Shoaie, S., Karlsson, F., Mardinoglu, A., Nookaew, I., Bordel, S., and Nielsen, J. (2013). Understanding the interactions between bacteria in the human gut through metabolic modeling. Scientific reports *3*, 2532.
Slack, E., Hapfelmeier, S., Stecher, B., Velykoredko, Y., Stoel, M., Lawson, M.A.E., Geuking, M.B., Beutler, B., Tedder, T.F., Hardt, W.D., et al. (2009). Innate and Adaptive Immunity Cooperate Flexibly to Maintain Host-Microbiota Mutualism. Science *325*, 617-620.
Smith, M.I., Yatsunenko, T., Manary, M.J., Trehan, I., Mkakosya, R., Cheng, J., Kau, A.L., Rich, S.S., Concannon, P., Mychaleckyj, J.C., et al. (2013). Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science *339*, 548-554.
Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nature reviews. Microbiology *9*, 279-290.
Stappenbeck, T.S., Hooper, L.V., and Gordon, J.I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. P Natl Acad Sci USA *99*, 15451-15455.
Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J.A., and Stahl, D.A. (2007). Metabolic modeling of a mutualistic microbial community. Mol Syst Biol *3*, 92.
Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M.A., Benezra, A., DeStefano, J., Meier, M.F., Muegge, B.D., et al. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature *510*, 417-421.
Sun, J., and Zeng, A.P. (2004). IdentiCS--identification of coding sequence and in silico reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence. BMC Bioinformatics *5*, 112.
Swann, J.R., Want, E.J., Geier, F.M., Spagou, K., Wilson, I.D., Sidaway, J.E., Nicholson, J.K., and Holmes, E. (2011). Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. Proc Natl Acad Sci U S A *108 Suppl 1*, 4523-4530.
Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.P., Ugarte, E., Munoz-Tamayo, R., Paslier, D.L.E., Nalin, R., et al. (2009). Towards the human intestinal microbiota phylogenetic core. Environ Microbiol *11*, 2574-2584.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics *4*, 41.

Thiele, I., and Palsson, B.O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc *5*, 93-121.

Thiele, I., Swainston, N., Fleming, R.M., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. Nat Biotechnol *31*, 419-425.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009a). A core gut microbiome in obese and lean twins. Nature *457*, 480-484.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027-1031.

Turnbaugh, P.J., Quince, C., Faith, J.J., McHardy, A.C., Yatsunenko, T., Niazi, F., Affourtit, J., Egholm, M., Henrissat, B., Knight, R., et al. (2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A *107*, 7503-7508.

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009b). The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Sci Transl Med *1*.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. Nat Biotechnol *28*, 1248-1250.

Wall, R., Ross, R.P., Ryan, C.A., Hussey, S., Murphy, B., Fitzgerald, G.F., and Stanton, C. (2009). Role of gut microbiota in early infant development. Clin Med Pediatr *3*, 45-54.

van Passel, M.W.J., Kant, R., Zoetendal, E.G., Plugge, C.M., Derrien, M., Malfatti, S.A., Chain, P.S.G., Woyke, T., Palva, A., de Vos, W.M., et al. (2011). The Genome of <italic>Akkermansia muciniphila</italic>, a Dedicated Intestinal Mucin Degrader, and Its Use in Exploring Intestinal Metagenomes. Plos One *6*, e16876.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., Dugar, B., Feldstein, A.E., Britt, E.B., Fu, X., Chung, Y.M., et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature *472*, 57-63.

Vangulik, W.M., and Heijnen, J.J. (1995). A Metabolic Network Stoichiometry Analysis of Microbial-Growth and Product Formation. Biotechnol Bioeng *48*, 681-698.

Varma, A., and Palsson, B.O. (1994a). Metabolic Flux Balancing - Basic Concepts, Scientific and Practical Use. Bio-Technol *12*, 994-998.

Varma, A., and Palsson, B.O. (1994b). Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic by-Product Secretion in Wild-Type Escherichia-Coli W3110. Appl Environ Microb *60*, 3724-3731.

Watrous, J.D., and Dorrestein, P.C. (2011). Imaging mass spectrometry in microbiology. Nature reviews. Microbiology *9*, 683-694.

Wichmann, A., Allahyar, A., Greiner, Thomas U., Plovier, H., Lundén, Gunnel Ö., Larsson, T., Drucker, Daniel J., Delzenne, Nathalie M., Cani, Patrice D., and Bäckhed, F. (2013). Microbial Modulation of Energy Availability in the Colon Regulates Intestinal Transit. Cell Host & Microbe *14*, 582-590.

Wong, J.M., de Souza, R., Kendall, C.W., Emam, A., and Jenkins, D.J. (2006). Colonic health: fermentation and short chain fatty acids. J Clin Gastroenterol *40*, 235-243.

Wos-Oxley, M., Bleich, A., Oxley, A., Kahl, S., Janus, L., Smoczek, A., Nahrstedt, H., Pils, M., Taudien, S., Platzer, M., et al. (2012). Comparative evaluation of establishing a human gut microbial community within rodent models. Gut microbes *3*, 234-249.

Wostmann, B.S. (1981). The germfree animal in nutritional studies. Annu Rev Nutr *1*, 257-279.

Wu, H.J., and Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. Gut microbes *3*, 4-14.

Xiong, Y., Miyamoto, N., Shibata, K., Valasek, M.A., Motoike, T., Kedzierski, R.M., and Yanagisawa, M. (2004). Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41. Proc Natl Acad Sci U S A *101*, 1045-1050.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., and Gordon, J.I. (2003). A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. Science *299*, 2074-2076.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., et al. (2007). Evolution of Symbiotic Bacteria in the Distal Human Intestine. PLoS Biol *5*, e156.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222-227.

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., and Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics *26*, i255-260.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. Isme J *6*, 1535-1543.

Zhang, Z.G., Geng, J.W., Tang, X.D., Fan, H., Xu, J.C., Wen, X.J., Ma, Z.S., and Shi, P. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. Isme J *8*, 881-893.

Zhao, L. (2013). The gut microbiota and obesity: from correlation to causality. Nature reviews. Microbiology *11*, 639-647.

Zomorrodi, A.R., and Maranas, C.D. (2012). OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. Plos Computational Biology *8*.

Zomorrodi, A.R., Suthers, P.F., Ranganathan, S., and Maranas, C.D. (2012). Mathematical optimization applications in metabolic networks. Metab Eng *14*, 672-686.