Whole-body systems approaches for gut microbiota-targeted, preventive healthcare

Liping Zhaoa,*, Jian Shena,b

a School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai Center for Systems Biomedicine, Shanghai 200240, China
b Bioinformatics Center, Key Lab of Molecular Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

Article info

Article history:
Received 24 July 2009
Received in revised form 5 January 2010
Accepted 15 February 2010

Keywords:
Whole-body approaches
Systems biology
Gut microbiota
Preventive healthcare

Abstract

Humans are superorganisms with two genomes that dictate phenotype, the genetically inherited human genome (25,000 genes) and the environmentally acquired human microbiome (over 1 million genes). The two genomes must work in harmonious integration as a hologenome to maintain health. Nutrition plays a crucial role in directly modulating our microbiomes and health phenotypes. Poorly balanced diets can turn the gut microbiome from a partner for health to a “pathogen” in chronic diseases, e.g. accumulating evidence supports the new hypothesis that obesity and related metabolic diseases develop because of low-grade, systemic and chronic inflammation induced by diet-disrupted gut microbiota. Due to the tight integration of gut microbiota into human global metabolism, molecular profiling of urine metabolites can provide a new window for reflecting physiological functions of gut microbiomes. Changes of gut microbiota and urine metabolites can thus be employed as new systems approaches for quantitative assessment and monitoring of health at the whole-body level with the advantage of measuring human health based on the results of interactions between the two genomes and the environment rather than just host genomic information. Large-scale population-based studies in conjunction with these whole-body level systems methods will generate pre-disease biomarkers with predictive power, thus making preventive health management of populations with rapidly changing disease spectrums possible through re-engineering of the imbalanced gut microbiomes with specially designed foods/diets.

© 2010 Elsevier B.V. All rights reserved.

1. Preventive healthcare and new health assessment technologies

It seems that most countries on this planet are facing a devastating epidemic of chronic diseases, particularly diet and lifestyle-related disorders such as obesity, diabetes, and cardiovascular diseases (Consultation, 2000). These diseases are characterized by deranged metabolism, which is progressive over a long period of time and which appears to be non-reversible (i.e., patients are affected with a chronic condition.) The endpoint is often death that results from system failure due to various complications. Thus, these diseases are life-long and economically costly. In terms of metabolic diseases, heart disease, stroke and diabetes alone will cost China 558 billion US$ by 2015 (WHO, 2005); soon this type of disease prevalence alone will drag the public healthcare systems of even the most affluent countries into total bankruptcy, let alone in rapidly developing countries such as China and India.

Prevention is the only viable solution to this public healthcare disaster. Epidemiological evidence has indicated that more than 90% of chronic diseases are actually preventable given proper and in-time diet and lifestyle interventions (Campbell and Campbell, 2005; Willett, 2008). However, diet and lifestyle management for “clinically healthy” populations are very difficult to implement because the general public tend not to take the advice of healthcare providers until real diseases occur.

To face the challenges of chronic diseases, biomarkers of the pre-disease transition stage must be used instead of biomarkers related to the disease stage. Yet, before these biomarkers are identified and the foundation of the pathological process of chronic diseases such as metabolic diseases is elucidated, the health assessment technol-
ogy should have the capacity to identify those who are moving away from the health trajectory and toward a disease attractor (German et al., 2005).

What is an ideal system for health assessment in preventive healthcare?

An ideal health assessment system must work at both the individual and population levels.

(1) The sampling process should be non-invasive or minimally invasive so that “healthy people” will comply with sampling.
(2) The measurements should be systemic in that they reflect the health condition of every part of the body, not simply one particular organ;
(3) The technologies should be sufficiently low cost and high-throughput to make dynamic monitoring of large populations possible.
(4) The measurements should be molecular-based so that the data can be integrated with and supported by pre-existing knowledge in molecular medicine.
(5) The assessment should be non-targeted and discovery-based. Current health assessment technologies are mostly if not all targeted analysis, i.e. one measurement that focuses on one particular pathology or condition to reveal the presence and progression of one or a related group of diseases and disorders. For example, the glucose tolerance test can only identify individuals experiencing insulin resistance. As human diseases are highly diverse, attempts to monitor human health with all possible tests would significantly increase the costs of testing to a formidable level.
(6) The readouts should be digital, quantitative, and reproducible so that all data collected based on identical protocols can be integrated into one database for modeling, simulation, and pattern recognition.

Systems understanding of human biology at the whole-body level is the foundation for developing this type of novel health assessment technologies. What specific biological properties of human organisms can be monitored for the ideal health assessment in prevention and what kinds of analytical technologies can be employed to satisfy these criteria?

2. Emergent functions and systems health monitoring

Human bodies are complex biological systems, within which individual cells/tissues/organs integrate and coordinate with each other to achieve the global functions. The most important feature of a complex system is its emergent properties (i.e., “the whole is bigger than the sum of the parts”), and the formation of one emergent property is the result of the integrated functions of many if not all parts of the system. Understanding the structure and function of each part of the body will not automatically lead to an understanding of the body’s behavior as a whole. Instead, “to understand the whole, one must study the whole” (Kell, 2004). Ludwig von Bertalanffy is one of the founders of general systems theory and the primary proponent for the concept of “emergent functions”. In his opinion, the ultimate goal of general systems theory was to develop methods to describe emergent functions and understand why functions of the whole are bigger than the summation of those of all the parts (Hammond, 2003). In light of this important concept, systems biology should be the science behind developing methods for measuring emergent functions of biological systems, be it the human body or the cell, and for understanding why functions of the whole body (or cell) are more than simply the summation of those of all component organs (or organelles). Focusing on emergent functions at the systems level will warrant the independent position of systems biology and distinguish this new science from other branches of biological disciplines.

The foundation of biomedical science and the healthcare industry is to stratify individuals quantitatively and predictively according to their health status. In other words, classification is the very fundamental task of health assessment technology. Classification and identification of natural complex systems such as human bodies can be based on the measurements of their emergent functions before each component of the system has been understood.

What kinds of emergent functions at the whole-body level can be measured with molecular methods in a non-invasive, global, and dynamic way still be digital, quantitative, and highly reproducible to support preventive, predictive, and personalized healthcare? Biological information in extracellular spaces, such as metabolites in urine, bacteria in fecal matter, and proteins in blood, can be such emergent functions assessable with molecular profiling approaches. Among these emergent functions, those that are related with the composition and activity of gut microbiota may play a pivotal role for human health assessment thanks to the recent developments in the understanding of the roles of gut microbiota in human health; particularly in the onset and progression of chronic diseases such as metabolic syndromes.

3. Superorganism concept and gut origin of chronic diseases

Human beings are “Superorganisms” or “Ecosystem man” (Lederberg, 2000). Proportionally, the human body consists of only 10% human cells and the remaining 90% are cells of microbial origin. Thus, there are two genomes within the human body, the human genome, genetically inherited from parents and the human microbiome, acquired from the environment after birth. The human genome and human microbiome work together via enterohepatic circulation and other anatomic and physiological connections to form the hologenome (Zilber-Rosenberg and Rosenberg, 2008) for controlling human-microbial metabolic axis within the human body (Nicholson et al., 2005). Therefore, our global metabolism at the whole-body level is the result of interactions between the two integrated genomes and the environment. Diet and drugs taken into our body will be co-processed and co-metabolized by functional genes of both the human genome and the microbiome. In contrast to the human genome, the gene composition of the human microbiome is rather flexible and can be modulated by foods and drugs. Changes of the composition of gut microbiota will affect the host metabolism since the global metabolism of the human body consists of functions encoded by both genomes. These two genomes must work in harmony to maintain the health of human bodies under various environmental conditions (Jia et al., 2008). This superorganism concept is a new, complete view on human organisms for managing health and wellness.

Our global metabolism at the whole-body level is the integration between the activities of our genome and the microbiome. “As nutrients are brought to cells and tissues by the circulatory system, so are the metabolic products from our microbial symbionts” (McFall-Ngai, 2008). In other words, no human cell can escape the fate of being dosed by metabolites produced by the gut microbiota. Microbial products produced by members of the gut microbiota can directly affect human physiology and immunity. Fecalibacterium prausnitzii has been identified as an anti-inflammatory commensal bacterium that can excrete anti-inflammatory compounds to block NF-κB activation and IL-8 production (Sokol et al., 2008). Lactobacillus acidophilus can modulate intestinal pain and induce opioid and cannabinoid receptors with the possible production of pain-relieving substances (Rousseaux et al., 2007). Chemical chaperones that can reduce ER stress and restore glucose home-
Ostasis are actually produced or co-metabolized by gut bacteria (Ozcan et al., 2006). In addition to these drug-like compounds, some gut bacteria produce various toxins related with chronic diseases. Overgrowth of neurotoxin producers in the gut may be responsible for progression of children’s autism (Parracho et al., 2005; Sandler et al., 2000). Genotoxic compounds produced by the gut microbiota have been connected with incidence of colon cancer, and perhaps some other types of cancers as well (Hughes and Rowland, 2000). More intriguingly, immunotoxins such as endotoxins, produced by opportunistic pathogens in gut microbiota have been demonstrated in animal models to be the primary trigger of metabolic diseases (Cani et al., 2007a).

Energy imbalance has been long proposed as the fundamental cause of obesity. However, two recent reports pose challenges to this established theory. The first report found that germ-free mice are resistant to high fat diet-induced obesity (Backhed et al., 2007). The second report showed that addition of a prebiotic oligofructose into the high fat diet can protect mice from developing obesity (Cani et al., 2007b). These data indicate that gut microbiota is the possible fundamental mediators between extra calories and obesity. Diet-disrupted gut microbiota may play an indispensable role in mediating the etiology of obesity and other related metabolic diseases in two distinct but related mechanisms.

In the first mechanism, gut microbiota directly participate in host energy metabolism in three ways: (1) gut microbiota ferment otherwise indigestible food components into extra calories; (2) gut microbiota can suppress the expression of the fasting-induced adipogenic factor (fasf) gene. Fasf is a circulating lipoprotein lipase inhibitor whose expression is required for mobilizing genes involved in fatty acid oxidation, the process for burning stored fat. Suppression of this regulator gene by some members of the gut microbiota makes calorie restriction (fasting) much less efficient for losing weight; (3) suppression of fasf expression by gut microbiota also is essential for the microbiota-induced deposition of triglycerides in adipocytes. Gut microbiota in obese animals have the capacity to control expression of host genes, affecting the two sides of the energy equation and essentially transforming the host organism into a highly efficient fat-making and storage machine (Backhed et al., 2004, 2007).

In the second mechanism, the gut microbiota plays a pivotal role in inducing low-grade, systemic and chronic inflammation in obese subjects. This inflammation has long been proposed as the important pathological condition responsible for onset and progression of metabolic diseases such as type II diabetes and cardiovascular diseases (Wellen and Hotamisligil, 2005; Yudkin et al., 2004). For example, elevated levels of inflammatory cytokines such as TNF-α can promote serine phosphorylation of insulin receptor substrate 1. After this modification, the receptor becomes inactivated so that it no longer responds to insulin signaling. This molecular mechanism for insulin resistance development is the important pathological condition for type II diabetes (Hotamisligil et al., 1996). However, the origin and primary mediator of this inflammatory condition have been controversial. Recently, the gut microbiota has been proposed to be responsible for inducing this inflammation in obese individuals. In one important animal study, it was found that LPS, the endotoxin produced by gram-negative opportunistic pathogens in the gut, is increased 2–3 times in the blood of high-fat diet fed animals, which showed low-grade systemic chronic inflammation comparable to what has been found in human subjects. The mild increase of endotoxin in the blood (termed metabolic endotoxemia) is the result of high-fat diet promoted overgrowth of LPS-producers and the diminishing of the gut barrier-protecting bacteria such as bifidobacteria. This diet-disrupted gut microbiota releases more endotoxin into blood across a damaged gut barrier that activates the host innate immune system and induces inflammation (Cani et al., 2007a). Maintaining a normal population level of bifido bacteria by including of oligofructose in the high-fat diet prevented the animals from developing obesity and early diabetes symptoms by protecting the gut barrier integrity and reducing endotoxemia and thus inflammation (Cani et al., 2007b).

Low-grade, systemic and chronic inflammation have been also implicated in the progression of aging, age-related diseases (Chung et al., 2009) and many forms of cancer (Vasto et al., 2009). It is possible that most, if not all, diet-related chronic diseases have their root in the gut microbiota. In this new hypothesis on the gut origin of chronic diseases, an imbalanced diet disrupts the gut microbiota leading to increased population levels of various toxins producers and diminished levels of gut barrier-protecting bacteria to release abnormal levels of various toxins including cytokotixins, genotoxins, and immunotoxins from the gut into the blood through a damaged gut barrier. Gut microbiota structure can be stable over many years if individuals do not change the diet pattern (Zoetendal et al., 2006). This stable microbiota would assure a constant level of toxins released into the blood system to circulate throughout the body over many years. A long-term increase of various toxins in the host’s blood eventually will lead to various forms of chronic diseases, in which inflammation serves as a primary pathological promoter.

Dysregulated metabolism, therefore, results from abnormal host immune function induced by disintegrated harmony between the human host and its microbiome via an unhealthy diet. This concept opens vast possibilities for prevention and intervention of chronic diseases by modulating gut microbiota with selected foods based on their known effects on growth of different members of the ecosystem. Due to this central mediating role of the gut microbiota in the onset and progression of chronic diseases, particularly metabolic diseases, emergent functions at the whole-body level as related to the gut microbiota thus can be used for monitoring human health in both individuals and populations for preventive healthcare (Nicholson, 2006).

For the human body, the production of urine and fecal matter is one of the most important emergent functions. These types of products are present outside of human cells (i.e., extracellular space) but reveal biological information reflecting functional status of human body: the urine contains abundant small-molecular metabolites—“the urine metabonome” consisting of metabolites from both host and gut microbiota, and the fecal matter contains large quantity of microorganisms—“the fecal microbiome”. Urine and fecal matter carry important biological information about the metabolism of the body, and any malfunction within the human body can lead to changes in the molecular composition of these samples (Nicholson et al., 1999). Global systems biology may help capture these changes for health assessment and monitoring (Nicholson and Wilson, 2003).

4. Metabolic profiling of the urine metabonome for health assessment

The kidney is a key organ for the maintenance of physiological homeostasis by eliminating unwanted metabolites from the body through urine. Hence, urine contains various metabolite species arising from endogenous metabolism, gut microbiota activities and the co-metabolism of both host and microbiome (Nicholson and Wilson, 2003). The biochemical composition of the urine reflects the physiological processes occurring in the superorganism (Nicholson and Wilson, 1989).

The concept of metabonomics was first defined by Jeremy Nicholson as the “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to patho-
physiological stimuli or genetic modification" (Nicholson et al., 1999; Nicholson and Wilson, 2003). Metabolites in body fluids such as blood, urine, and saliva are globally profiled and dynamic changes in the metabolite composition are monitored with high-throughput and high-sensitivity techniques such as GC–MS or NMR spectroscopy. Multivariate statistical approaches can help to determine the physiological or pathological state of the human body, reveal the relationship of the phenotypes, environments and genes, and identify biomarkers (German et al., 2005; Goodacre, 2007).

NMR spectroscopy allows profiling of the metabolite composition in biofluids including blood, urine and feces, etc. (Martin et al., 2007; Martin et al., 2008b), and has been applied widely in metabolomic studies. NMR spectroscopy is non-invasive, cost effective, high-throughput, and reproducible. It does not require complex sample preparation such as deproteinization or derivatization, and typically requires only a few minutes to generate a complete metabolite profile. Both biofluids and intact tissues can be analyzed directly. NMR spectra provide both metabolite composition and molecular structure information, which is needed for the identification of unknown substances that may be used as biomarkers. NMR spectroscopy-based metabolomics has been used widely in studies on toxicity, nutrition, epidemiology, etc. (Rezzi et al., 2007a).

Dumas et al. evaluated the analytical reproducibility and inter-instrument variability of NMR spectroscopy by metabolotyping 24-h urine specimens collected from approximately 800 persons of different racial/ethnic origins. Multivariate analyses of the spectra data showed that the analytical reproducibility of the NMR screening platform was >98% and that most classification errors were due to inconsistencies in the urine specimen handling procedures. In addition, cross-population differences in urinary metabolites could be related to genetic, dietary, and gut microbial factors. Results of the study suggested that NMR spectroscopic data from different labs could be integrated and analyzed together because of the high reproducibility. This makes NMR spectroscopy-based metabolomics a powerful tool for large-scale population phenotyping in epidemiological studies and for identifying biomarkers for early detection of chronic diseases threatening the public health (Dumas et al., 2006b; Holmes et al., 2008).

Robosky et al. found genetically identical Sprague-Dawley [Crl:CD(SD)] rats that were kept in two different environments showed different urinary metabolic phenotypes. Statistical analyses showed that the most pronounced difference between the two phenotypes was the relative amount of hippuric acid versus other aromatic acid metabolites of chlorogenic acid, both of which are derived from gut microbial metabolism (Robosky et al., 2005). Similarly, Nicholson’s group characterized the urinary metabolic phenotypes of rats of identical genetic background with 1H NMR spectroscopy-based metabolic methods, and a few rats were significantly different from other animals in the metabolites produced by the gut microbiome (Nicholson et al., 2005). These results indicate that the activities of the gut microbiome significantly affect the metabolic phenotype of the host. NMR spectroscopy-based metabolic profiling can capture the metabolic activities of both host and gut microbiota and thus has the potential for assessing the health status of the superorganism hosts.

Nonalcoholic fatty liver disease (NAFLD) is one type of metabolic syndrome. Using 1H NMR spectroscopy and multivariate statistical modeling, Dumas et al. characterized the plasma and urine metabolite profiles of 12956 mice that are genetically susceptible to dietary-induced impaired glucose homeostasis and NAFLD, and found that the mice have a disruption of choline metabolism: low circulating levels of plasma phosphatidylcholine and high urinary excretion of methylamines. Because dietary choline is necessary for fat removal from the liver and urinary methylamines are the products of the gut microbiota metabolism on dietary choline, the authors concluded that the gut microbiota of 12956 mice had a strong capacity to convert dietary choline into methylamines and thus reduce the bioavailability of choline to predispose the mice to NAFLD. Their data also indicate that gut microbiota may play an active role in the development of insulin resistance (Dumas et al., 2006a).

Hypertension is another metabolic syndrome that is a major risk factor for coronary heart disease and stroke. Holmes et al. applied 1H NMR spectroscopy-based metabonomic methods to investigate the 24-h urinary metabolic phenotype variation of over 4000 people from 17 groups form China, Japan, the UK, and the USA. The urinary metabolite patterns for human populations with different diets, coronary heart disease/stroke incidence, and blood pressure, were significantly differentiated, and the discriminatory metabolites were identified and quantified. Subsequently, researchers correlated the excretion amounts of these metabolites in 24-h urine with blood pressure for individuals, and showed that alanine and hippurate, which reflect diet and gut microbial activities, are associated with blood pressure of individuals. This study also demonstrated that the application of metabolic phenotyping techniques to epidemiological data has a high potential for the discovery of biomarkers related to cardiovascular disease risk (Holmes et al., 2008). These biomarkers often have their root in the gut microbiota.

The 1H NMR spectroscopy-based metabonomic study showed that people consuming varied diets have different urinary metabolotypes: high-meat consumption elevated urine excretion of creatine, carnitine, acetylcarnitine, and trimethylamine-N-oxide (a microbial-mammalian co-metabolite), while the vegetarian diet increased p-hydroxyphenylacetate (a microbial-mammalian co-metabolite) in urine compared to meat diets, which suggests an alteration in composition or metabolism of the gut microbiota in response to diet (Stella et al., 2006). Rezzi et al. developed a “nutrimetabonomic” approach in which urine and plasma metabolotypes measured by 1H NMR spectroscopy were correlated with individuals’ behavioral/psychological dietary preference, namely “chocolate desiring” or “chocolate indifferent”. Their data showed that postprandial lipoprotein profile and gut microbial co-metabolism are the discriminatory factors for the differentiation of “chocolate lovers” and “chocolate haters”, suggesting that specific dietary preferences have a root in basal metabolic state and gut microbiome activity that in turn may have long-term health consequences for the host (Rezzi et al., 2007b). The data also indicate the possibility that modulation of the gut microbiota with dietary intervention may improve host metabolism and health, which makes metabolomics a promising tool for personalized nutritional management.

Clayton et al. developed a “pharmaco-metabonomic” approach to predict how different individuals will respond to a particular drug/dose combination. Rats were treated with paracetamol (acetaminophen), and the animals’ pre-dose urine metabolotypes were profiled. Chemometric modeling showed animals with different pre-dose urinary metabolites responded to the drug differently and the discriminatory metabolites originated from the metabolism of the gut microbiota. Hence, the pre-dose urinary composition can be used to predict an individual’s response to drugs, which is profoundly influenced by the gut microbiota (Clayton et al., 2006). Indeed, we have pointed out that the gut microbiota can be a potential drug target, which could become a new research direction in personalized medicine (Jia et al., 2008).

In summary, by phenotyping biofluid metabolite composition, the above metabolomic studies demonstrate that the gut microbiota play an essential role in host metabolism and profoundly affect the health and disease of the humans. However, the composition of the gut microbiota was not analyzed, which makes the investigations incomplete. Indeed, to illustrate the symbiotic interaction between the gut microbiota and host, it is necessary to
perform experiments on both metabolomics and the gut microbiota structure and correlate the results.

5. Metagenomic profiling of gut microbiota for health assessment

Evidence is accumulating rapidly that the balance within the gut microbiota is an important determinant for the host health (Guarnier and Malagelada, 2003). An imbalance within the gut microbiota is involved in the etiology of diseases not only of the bowel but also other organs. Since the microbial composition in feces is a collection of all symbiotic microorganisms from the gut, fecal samples are used in most studies on the relationship between the gut microbiota composition and disease. Molecular profiling of the composition of species and the functional gene pool in fecal samples has the potential to be used for health assessment analogous to using exhaust gas for monitoring engine health.

Moore et al. compared the fecal bacterial composition of different human populations (polyp patients, Japanese-Hawaiians, North American Caucasians, rural native Japanese, and rural native Africans) with high, intermediate and low risk of colon cancer, respectively. The population levels of 15 bacterial species including Bacteroides vulgatus, B. stercoris and Ruminococcus albus were positively correlated with the high risk of colon cancer, and the proportion of lactic acid-producing bacteria such as Lactobacillus spp. and Eubacterium aerofaciens was higher in populations with a low risk (Moore and Moore, 1995). However, this type of work relied on culture-based technology for profiling gut microbiota. Due to the fact that less than 20% of gut bacteria are culturable with current technology, molecular methods targeting the diversity of bacterial DNAs have become the primary driving force for gut microbiota research in recent years.

DNA sequencing has become the gold standard for comprehensive and quantitative characterization of the structure of microbial communities. Thanks to the improvement of sequencing techniques, the throughput of sequencing is continuously increasing while the cost is significantly dropping, which lay a solid foundation for the structural analyses of the gut microbiota.

The 16S rRNA gene clone library has become one of the most widely used techniques in studies on the diversity of microbial communities. In this approach, the 16S rRNA gene of all bacteria in a community is amplified with universal primer pairs, and subsequently cloned to construct a library. Clones are picked randomly for sequencing, and each unique sequence is aligned to those deposited in public databases (such NCBI and RDP) to identify the microbial species that the sequence originated from and to determine its phylogenetic position. Clones with identical or similar sequences are classified as one operational taxonomic unit (OTU). The number of OTUs represents the number of different bacterial species (or populations) in the community, and the proportion of each OTU represents the richness of the corresponding bacterial species (or population), which are the parameters reflecting the structure of the bacterial community. Eckburg et al. applied the 16S rRNA gene clone library technique to profile the microbial diversity of the gut microbiota of three healthy persons (Eckburg et al., 2005). Twenty-one clone libraries were constructed with the genomic DNA extracted from fecal samples, as well as from mucosal samples of different regions along the large bowel including the cecum, ascending colon, transverse colon, descending colon, sigmoid colon and rectum. Altogether, 11,831 bacterial ribosomal RNA gene sequences were examined, which then were classified into 395 phylotypes at the 99% sequence similarity as the threshold; 244 of phylotypes were from unknown species. This work, for the first time, provided an integral overview of the composition of the human gut microbiota, and revealed the complexity of the human gut microbiome.

Next-generation sequencing techniques that are rapid, high-throughput and low cost are being developed. By combining pyrosequencing, single molecule emulsion PCR and microfluidic techniques, “454 pyrosequencing” has become the most powerful, robust and promising new-generation sequencing approach for gut microbiota profiling (Margulies et al., 2005). It is highly automatic without requirements for clone library construction, fluorescence-labeled nucleotide primers/probes or electrophoresis. Moreover, it is a high-throughput system capable of sequencing over 100 megabases of DNA in 7.5 h with 200–300 base pairs in average for each read. Over 400,000 reads can be obtained in one run, which reduces the cost significantly compared to conventional Sanger sequencing. The accuracy for one read longer than 200 base pairs can reach to 99.5%. These unique advantages of 454 sequencing have made this technique employed extensively in genomic studies (Sogin et al., 2006; Turnbaugh et al., 2006). McKenna et al. used barcode 454 pyrosequencing to study the diversity of the vertebrate gut microbiota. Approximately 141,000 sequences of 16S rRNA genes were obtained from 100 gastrointestinal bacterial samples of feces, colonic contents and mucosa collected from healthy macaque and macaque with colitis induced by HIV infection and examined; it was found that communities from diseased and healthy animals differed significantly in composition (McKenna et al., 2008). Andersson also utilized bar-coded 454 sequencing of 16S rRNA gene fragments to compare the gut microbiota of healthy people and patients, and demonstrated the applicability of bar-coded pyrosequencing as a high-throughput method for comparative microbial ecology (Andersson et al., 2008). With the improvement of the 454 sequencing technique with longer reads and the advances in phylogenetic analytical methods of short DNA sequences, 454 pyrosequencing will become more robust and accurate, and has great application potential for microbial ecological studies.

Metagenome represents the genomes of all constituent microbes in one microbial community (Schloss and Handelsman, 2003). One main strategy of metagenomic techniques is multiplex cloning. The total genomic DNA is extracted from the microbial communities in environmental samples, and the resultant DNA fragments are inserted into vectors to construct bacterial artificial chromosomes (BACs) or fosmid libraries. By sequencing clones selected with gene markers reflecting the evolution of bacteria species such as 16S rRNA gene, microbial species present in the community can be identified even if they are unculturable, and other fragments adjacent to the marker gene in their genomes can be discovered. In addition, new genes encoding new functional enzymes can be discovered by sequencing clones selected with bioactive assays. Hence, metagenomic techniques are the new strategy to explore genes in microbial communities without cultivating microbial members, which allows a more comprehensive and deeper understanding of the composition and structure of a microbial community, and enables the discovery of new genes and bacteria with previously unknown functions. Manichanh et al. used a comprehensive metagenomic approach to characterize the diversity of the human gut microbiome. Two fosmid libraries of genomic DNA extracted from fecal samples of healthy donors and Crohn’s disease patients were constructed. Sequencing of clones that contained 16S rRNA gene from the two libraries demonstrated that the diversity of the bacterial phylum Firmicutes was reduced in the fecal microbiota of Crohn’s disease patients when compared to healthy individuals (Manichanh et al., 2006).

The other technical strategy in metagenomics is the whole-genome shotgun sequencing approach, which sequences the genomic DNA fragments isolated from environmental samples in a random manner (Venter et al., 2004). Gill et al. carried out the whole genomic shotgun sequencing of DNA isolated from human
distal gut microbiome. Approximately 78 million base pairs of unique DNA sequence were analyzed and metabolic function analyses of identified genes were performed. The data showed that the human microbiome has a significantly enriched metabolism of glycans, amino acids, and xenobiotics; methanogenesis; and 2-methyl-n-erythritol 4-phosphate pathway-mediated biosynthesis of vitamins and isoprenoids (Gill et al., 2006).

However, the cost of conventional metagenomic techniques is relatively high, and analysis of a single sample is time consuming. The advent of the 454 pyrosequencing technique partially overcomes the disadvantages and advances the capability for metagenomic studies. Turnbaugh et al. analyzed the gut microbiome of obese and lean mice with metagenomic techniques combining conventional shotgun sequencing with 454 sequencing. The data indicated that the obese microbiome has an increased capacity to harvest energy from the diet (Turnbaugh et al., 2006).

Given the development of these DNA-based profiling technologies, accumulating evidence has demonstrated that structural changes within the gut microbiota can be used for health assessment and monitoring.

Ott et al. investigated the mucosa-associated colonic microbiota of patients with inflammatory bowel disease (IBD) and healthy controls using a clone library combined with fingerprinting techniques. Diversity within the microflora in Crohn’s disease patients was decreased to 50% of that of controls and to 30% in ulcerative colitis patients; the reduction in diversity in IBD was due to loss of normal anaerobic bacteria, such as Bacteroides species, Eubacterium species, and Lactobacillus species. Therefore, the author concluded that mucosal inflammation in IBD is associated with loss of normal anaerobic bacteria (Ott et al., 2004). Manichanh et al. employed metagenomic approaches to further characterize the gut microbiome in Crohn’s disease, and found that the diversity within the Firmicutes phylum was significantly reduced in the patients (Manichanh et al., 2006). Our group used denaturing gradient gel electrophoresis to compare the microbiota of the ulcerated and non-ulcerated gut mucosa from individuals with ulcerative colitis. The data revealed that Lactobacilli and the Clostridium leptum subgroup were significantly different between the ulcerated and the non-ulcerated regions, suggesting that localized dysbiosis of the mucosa-associated lactobacilli and the C. leptum subgroup may be closely related to ulcerative colitis (Zhang et al., 2007).

The structure of the gut microbiota of obese people is different from lean individuals. The relative proportion of Bacteroidetes to Firmicutes decreases in the gut of obese people, and this proportion increases with weight loss on low-calorie diets, indicating that the gut microbiota might have potential therapeutic implications for obesity (Ley et al., 2006). Similar differences in the gut microbiota were observed in obese and normal-weight mice, and a functional genomics study revealed that the gut microbiome in obese mice had a stronger capability to harvest calories from foods (Turnbaugh et al., 2006). With this mouse model, Turnbaugh et al. found that diet-induced obesity (DIO) stimulated an uncultured clade within the Mollicutes class of the Firmicutes, which was decreased by following dietary manipulations that limit weight gain (Turnbaugh et al., 2008). In addition, microbiota transplanted from mice with DIO to lean germ-free recipients promoted greater fat deposition than transplants from lean donors, indicating that the structured gut microbiota contributed significantly to the obesity phenotype (Turnbaugh et al., 2008, 2006). Massive pyrosequencing also revealed that the obese gut may have a functional gene core rather than a phylogenetic core of commonly shared species (Turnbaugh et al., 2009).

Both genetic variations and diet-disrupted gut microbiota can predispose animals to metabolic syndromes (MS). To assess the relative contributions of host genetics and diet in shaping the gut microbiota and modulating MS-relevant phenotypes in mice, we used mice of two genotypes (a high-density lipoprotein knock-out mouse and its wild-type counterpart) on two differing diets (high fat or normal chow) for six months as a research model to ensure stabilized physiological integration between gut microbiota and host. Using 454 pyrosequencing and appropriate multivariate statistics as top-down molecular profiling tools, we were able to show, for the first time, that diet plays a much more profound role than host genotypes in shaping gut microbiota in a way relevant to MS development. We were able to further identify key phyotypes with defined responses to diet change, genotype diversity and host physiology (with or without obesity/insulin resistance). Most notably, the diminishing of gut barrier-protecting bacteria, and increasing numbers of sulfate-reducing and endotoxin-producing bacteria have been identified as key alterations associated with long-term, unlimited intake of high fat diet. Despite a complete host genome, wild-type animals on a high fat diet developed the worst MS phenotypes, likely due to a much higher caloric intake, and more profound alterations of gut microbiota. The finding that high-fat diet has a strong overriding role to host genome in altering gut microbiota for MS development reveals vast possibilities for combating metabolic diseases via diet interventions, with the disrupted gut microbiota as a primary target (Zhang et al., 2009a).

Kirjavainen et al. studied the composition of fecal microflora in infants with early onset atopic eczema, and found that the gut microbiota was different in atopic infants that varied with the extent of atopic sensitization. Infants with high sensitization had higher numbers of lactobacilli/enterococci, and Escherichia coli and Bacteroides counts are associated with the extent of atopic sensitization (Kirjavainen et al., 2002). Kalliomaki and co-workers’ investigation demonstrated that differences in the neonatal gut microflora precede the development of atopy, which suggested that the balance of indigenous intestinal bacteria plays a crucial role in the maturation of human immunity to a nonatopic phenotype (Kalliomaki et al., 2001).

Finegold et al. compared the gut microflora of autistic children and healthy children. The number of fecal clostridial species was higher in children with autism than in healthy children, and nine clostridial species were found only in autistic children. In gastrotaxic and duodenal specimens, non-spore-forming anaerobes and microaerophilic bacteria were totally absent in healthy children but present in significant numbers in autistic individuals (Finegold et al., 2002). Parracho et al. found that the fecal flora of children with autistic spectrum disorders (ASDs) contained a higher incidence of the Clostridium histolyticum group of bacteria than that of healthy children. Members of the C. histolyticum group are toxin-producers and may lead to gut dysfunction and systemic effects. These studies provide new insights into the role of the gut microflora in the development of autism, and indicate that dietary modulation of the gut microflora may help to alleviate ASDs in these patients (Knivsberg et al., 2002; Parracho et al., 2005).

6. Multivariate statistics for data mining in health assessment

Large amounts of multivariate data are produced in metabonomic and microbiomic studies, thus, multivariate statistical analytical methods must be used to model the data of different sample groups and to identify biomarkers or drug targets from the complex information on the host metabolism and gut microbiota (Nicholson et al., 1999). Multivariate statistical approaches include unsupervised and supervised methods. Principal Components Analysis (PCA) is an unsupervised method, in which no priori knowledge of the class of the samples is required, and samples are mapped according to the inherent composition of metabolites/gut microbiota. The clus-
tering of sample coordinates in PCA trajectories reflects an intrinsic similarity in biochemical/gut microbial composition. Stanley performed PCA analysis on the 1H NMR spectra of urine samples collected from Han Wistar rats, and found gender-dependent metabolic variation in these experimental animals. Sulfate conjugates of m-hydroxyphenylpropionic acid, trimethylamine-N-oxide, N,N’-dimethylglycine, m-hydroxyphenylpropionic acid, N-acetylglucosamine, and cholate in the urine were identified that allowed differentiation of male and female animals (Stanley et al., 2005). Partial Least Squares Discriminant Analysis (PLS-DA) and Orthogonal Partial Least Squares Discriminant Analysis (O-PLS-DA) are supervised methods in which sample class information is used to maximize the separation between classes, hence information that relates to biomarkers can be maximally extracted from the data and help us understand the response of the host metabolism/gut microbiota to physiological variations/interventions. We constructed *Bacteroides* spp. clone libraries with fecal samples from healthy and rotavirus-infected children, and analyzed the data from the two groups with PLS-DA. The results indicated that rotavirus infection significantly changed the *Bacteroides* composition, and *B. vulgatus, B. stercoris*, and *B. fragilis* were the key species of the infection (Zhang et al., 2009b). The multivariate statistical approaches are powerful in mining the relevant biomarkers from massive amounts of complex data.

Another important application of multivariate analysis in global systems biology research is to reveal the symbiotic relationships between the human/mammalian host and the gut microbiome by correlating the gut microbes with the host metabolic phenotypes (that is, a transgenomic approach). With a combination of NMR spectroscopic, microbiomic, and multivariate statistical tools, we profiled the fecal bacteria and urinary metabolites from seven Chinese individuals and modeled the microbial-host metabolic connectivities. For example, *Faecalibacterium prausnitzii* population variation is associated with modulation of eight urinary metabolites of diverse structure, indicating that this species is a highly functionally active member of the microbiome, and influences numerous host pathways. Another nine species were identified to have different and varied metabolic interactions with the host. Therefore, we introduced the concept of functional metagenomics, defined as “the characterization of key functional members of the microbiome that most influence host metabolism and hence health.” Our approach for understanding the dynamic basis of host-microbiome synergy provides a foundation for the development of functional metagenomics as a probe of systemic effects of drugs and diet that are of relevance to personal and public health (Li et al., 2008).

In a mouse model, Martin et al. performed parallel microbiological profiling, metabolic profiling by 1H NMR spectroscopy of liver, plasma, urine and ileal flushes, and bile acid profiling by ultra performance liquid chromatography–mass spectrometry. They also derived a transgenomic graph model using a multivariate statistical tool, showing that gut bacteria have a significant association with host-specific metabolotypes and impact directly on the host’s ability to metabolize lipids (Martin et al., 2007, 2008a,b).

7. The future of gut microbiota-targeted analysis in health assessment

As whole-body systems approaches, gut microbiota-targeted analysis (GMTA) can be applied in large-scale epidemiological research to discover pre-disease biomarkers. Urine and feces are two windows to reflect the health status of the human body; the comprehensive and precise profiling of the metabolites and microbes in these samples with metabolomic and metagenomic techniques can be used to characterize and monitor the health status of the whole human body at the molecular level. Analysis and modeling of data from large-scale populations with changes in health status using multivariate statistical methods and pattern recognition will allow biomarkers associated with the transition period from health to disease onset to be discovered and identified. Based on these epidemiological data, a preventative, predictive and personalized healthcare system finally will be established.

Conflict of interest

None.

Acknowledgements

The authors would like to acknowledge the support by the National Natural Science Foundation of China Program Grants 30730005 and 20875061; 973 Program Grants 2007CB513002 and 2004CB518600; 863 Program Grant 2008AA02Z315; International Cooperation Program Grants 2007DFC30450 and 075407001, and Chinese Academy of Sciences (Knowledge Innovation Program KSX1-YW-02).

References


