COMMENTARY

Human genomics and microbiomics: the post-genomic scenario

J. Rajendhran and P. Gunasekaran

The human genome has been sequenced and the task was named as the 'human genome project'. The ultimate goal of the human genome project was to understand the human physiology and health, and to fight against infectious and non-infectious diseases. Variations in the human genome sequence, such as single nucleotide polymorphisms (SNPs) and haplotypes, associated with various diseases have been identified through 'the HapMap project'. Recently, 'the 1000 genomes project' has been initiated to catalogue the chromosomal structural variations, which might be missed by the SNP approaches. In addition, the role of endogenous microorganisms, which live in and on humans and their genomes in human health has gained much attention in recent years. Hence 'the international human microbiome project' has been recently launched, which may reveal the role of endogenous microorganisms in maintaining the human health and susceptibility to diseases.

Soon after the human genome was sequenced, scientists and physicians worldwide started debating on the consequence of the human genome project[1]. In general, any two humans are more than 99.9% similar with respect to their genome sequences. Only a small fraction (0.1%) of sequences that varies among the people is responsible for the individual differences in susceptibility to disease or response to drugs[2]. Therefore, human genome research is now mainly focused on the 0.1% variation associated with different populations. The most common type of variant is the single nucleotide polymorphism (SNP), which is the single base difference at a particular site. Even before the human genome was completely sequenced, the need for cataloging of SNPs associated with various diseases was proposed[3]. In this direction, the SNP consortium was established in 1999 and more than 1.4 million SNPs have been reported[3] by 2001. The comprehensive database, dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/), has information on the SNP allele frequencies, the type of SNP (either coding or non-coding), and if coding, either synonymous or non-synonymous[4].

The international HapMap project[2] (www.hapmap.org) was launched in 2002. The HapMap is a catalogue of common genetic variants that occur in human beings. It describes what these variants are, where they occur in the genome, and how they are distributed among people within populations and among populations in different parts of the world. The first phase of the HapMap project was completed in 2005, where 1.3 million genotyping assays with more than one million SNPs were reported[5]. The second phase was completed in 2007 and has increased the total to more than 3.1 million SNPs[6]. The HapMap data consist of SNPs, SNP haplotypes, genotypes and tag SNPs. Variations organized on a single chromosome or part of a chromosome are referred to as haplotypes, which are usually inherited as intact blocks of information. New haplotypes are formed by additional mutations or by recombination. The coinheritance of SNP alleles on these haplotypes leads to associations between these alleles in the population, which is known as linkage disequilibrium (LD). Many empirical studies have reported significant levels of LD and often strong associations between nearby SNPs in the human genome. Therefore, there are only a few haplotypes in many chromosome regions, which account for most of the variations among people in those regions. As a result, only a few of these SNPs, referred to as 'tag SNPs', are sufficient to identify each of the common haplotypes in a region[2].

Recently, 'the 1000 genomes project' (www.1000genomes.org) has been initiated to create the most comprehensive map of human genetic variation, which involves sequencing the genomes of a thousand people from around the world[7]. The project builds on the human haplotype map developed by the international HapMap project. Going a major step beyond the HapMap, the 1000 genomes project will map not only the SNPs, but will also produce a high-resolution map of larger differences in genome structure called structural variants, such as rearrangements, deletions or duplications of genome segments. The catalogue of genetic variations developed in this project will be useful in genome-wide association studies on people with particular diseases.

Thus, human geneticists and epidemiologists are concerned about the genetic basis of susceptibility and resistance to diseases. On the other hand, microbiologists are more concerned about the microorganisms that live inside and on humans (human microbiota). The human microbiota has co-evolved with its human host and plays an important role in human physiology. For example, the human gastrointestinal tract itself harbours approximately 10 times more bacterial cells than there are cells in the entire human body. The gut microbiota is composed of >1000 species and it has been estimated that the human microbiome (total genomes of the microbiota) contains roughly 100 times more genes than the human genome. Thus, human beings are considered as 'superorganisms' and the metabolism of superorganism involves the integration of indigenous metabolic processes coded in the human genome with those of the associated microbiome[8]. Current estimates indicate that more than 99% of the microorganisms present in many natural environmental niches are not readily culturable and therefore not accessible. Analysis of nucleic acids directly extracted from the associated niche, the metagenomic approach, allows researchers to study natural microbiota without the need for cultivation[9]. Each organism in an environmental niche has a unique set of genes in its genome; the combined genomes of all the community members make up the 'metagenome'.

The metagenomics-based microbial diversity analysis largely depends on the sequence analysis of small subunit ribo-

somal RNA (16S rRNA) genes. The 16S rRNA genes are amplified using broad-range PCR primers based on the sequences that are well-conserved among prokaryotes, corresponding to positions 8 to 1513 of the *Escherichia coli* 16S rRNA gene[10]. The 16S rRNA genes amplified from metagenome are used to make libraries of clones, where each clone represents a 16S rRNA gene from a prokaryotic species. Individual clones are sequenced and the 16S rRNA gene sequence similarity analysis is used to identify the species-level phylogenetic types (phylotypes). The human metagenome-based 16S rRNA gene analyses have also been used to detect uncultivated organisms that cause diseases[11]. For this, DNA extracted from the site of infection was subjected to 16S rRNA gene-based microbial diversity analysis. The first novel pathogen to be identified by sequence-based method was *Rochalimaea henselae* (redesignated as *Bartonella henselae*), the organism responsible for bacillary angiomatosis (cat scratch disease)[12]. *Ehrlichia chaffeensis* causing a febrile illness associated with tick bites and *Tropheryma whipplei* causing the Whipple's disease are other examples of pathogens identified using this approach[11].

The health and predisposition to various non-infectious diseases of humans are also determined by the genes coded by the microbiome. For example, the influence of human microbiota in obesity has been reported[13]. The metagenomic sequence data of faecal specimens from obese and lean individuals revealed a strong association between the microbiota and obesity. In humans and mice, more than 90% of the colonic bacteria belong to just two divisions in the domain bacteria: the Firmicutes and the Bacteroidetes. The obese (ob/ob) mice have a 50% reduction in the abundance of Bacteroidetes and a proportional increase in Firmicutes compared with lean mice. Further, the role of microbiota in obesity has also been demonstrated by the 'microbiota transplantation' experiments. Turnbaugh *et al.*[14] have colonized the adult germ-free C57BL/6J mice with a microbiota harvested from the caecum of obese (ob/ob) or lean (+/+) donors. Colonization of germ-free mice with an obese (ob/ob) microbiota resulted in a significantly greater increase in total body fat than colonization with a lean (+/+) microbiota. The metagenomics-based shot-gun sequencing analyses of gut specimens from obese mice revealed that the obese microbiome has an increased capacity to harvest energy from the diet[14].

The international human microbiome project (HMP) has recently been launched to explore the role of microbiota in human health and disease[15]. The HMP is another rational extension of the human genome project. Ultimately, the goal is to associate differences in microbiome with those in metabolic function and/or disease. Initially, researchers will sequence 600 reference microbial genomes to make a collection of 1000 microbial genomes. Sequencing of more reference genomes would provide further insight into the role of human microbiota and its diverse metabolic functions. In addition, shotgun sequencing approach will be employed to generate more sequence data from the human microbiome. The datasets produced by HMP will be made accessible through a public repository database (www.hmpdacc.org). Initially, the microbiome profiles of healthy volunteers will be generated. These data will then be compared with microbiome from volunteers with specific diseases.

To conclude, the human genome project revealed the complete human genome sequence, which is 99.9% similar in all humans. The other genome-variation studies are aimed to characterize the common patterns within 0.1% of the genome structure, where humans differ from each other. The SNP consortium, the international HapMap project and the 1000 genomes project are aimed at identifying the underlying genetic variations in various human populations. The HapMap project has contributed much to our understanding of the underlying genetic variation in diverse human populations and has facilitated the discovery of many loci associated with common human diseases, such as diabetes, obesity, breast cancer, etc. The recently launched 1000 genomes project aims to discover much of the existing common variation, including both SNPs and the less-explored structural variants. The superorganism concept is an important paradigm shift in understanding human biology. The HMP is aimed at understanding the role of microbiome in human physiology and health. Thus, the integration of human genomics and microbiomics is essential for strategies to develop food or drugs.

1. Vasan, S. S., *Curr. Sci.*, 2001, **80**, 723–724.
2. The International HapMap Consortium, *Nature*, 2003, **426**, 789–796.
3. Sachidanandam, R. *et al.*, *Nature*, 2001, **409**, 928–933.
4. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K., *Nucleic Acids Res.*, 2001, **29**, 308–311.
5. The International HapMap Consortium, *Nature*, 2005, **437**, 1299–1320.
6. Frazer, K. *et al.*, *Nature*, 2007, **449**, 851–861.
7. Kuehn, B. M., *JAMA*, 2008, **300**, 2715.
8. Li, M. *et al.*, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 2117–2122.
9. Sharma, R., Ranjan, R., Kapardar, R. K. and Grover, A., *Curr. Sci.*, 2005, **89**, 72–77.
10. Gao, Z., Tseng, C. H., Pei, Z. and Blaser, M. J., *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 2927–2932.
11. Weng, L., Rubin, E. M. and Bristow, J., *Genome Res.*, 2006, **16**, 316–322.
12. Relman, D. A., Loutit, J. S., Schmidt, T. M., Falkow, S. and Tompkins, L. S., *N. Engl. J. Med.*, 1990, **323**, 1573–1580.
13. Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. and Gordon, J. I., *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 11070–11075.
14. Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. and Gordon, J. I., *Nature*, 2006, **444**, 1027–1031.
15. Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R. and Gordon, J. I., *Nature*, 2007, **449**, 804–810.

J. Rajendhran and P. Gunasekaran are in the Department of Genetics, Centre for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, India.*
**e-mail: gunagenomics@gmail.com*