



Metagenomics of the Human Intestinal Tract

A detailed understanding of human biology will require not only knowledge of the human genome but also of the human microbial metagenome. Humans live in constant association with microbes that are present on surfaces and in cavities of the human body, and even within our cells. The number of our microbial companions exceeds by at least ten-fold those of cells of our own body and the number of unique genes they encode is predicted to be at least 100-fold greater than the number of genes in our own genome. This complex and dynamic microbiota has a profound influence on human physiology, nutrition, and immunity. Disruptions in these human-associated microbial communities are a significant factor in many diseases. The ensemble of the genomes of human-associated microorganisms represents the human metagenome. Understanding the dynamic and variable nature of human microbial communities is a critical aspect of the challenge before us. Defining this dynamic diversity represents the next frontier of genomics. To progress towards this ambitious goal we focus on the microbiota of the intestinal tract, which plays a particularly important role in human health and well-being.

Microbes in our intestine

The microbiota of the human intestine is composed of a very high number of microbial cells which reach up to ten trillion within a few days after birth and can represent a weight of two kilograms. The intestinal microbes not only help digest food and synthesize vitamins that are needed by our body, they also protect us, by educating the immune system to distinguish friends from foes, and controlling production of various drug-like metabolites. Many different diseases originate from microbial disorders. This is naturally the case of infectious diseases affecting the digestive system. But non-infectious diseases such as obesity have also been associated with unusual changes in microbiota. Even apparently psychological disorders such as autism may be related to an overgrowth of neurotoxin-producing bacteria living in the intestine.

Genes of intestinal microbes, human health and disease – the project objectives

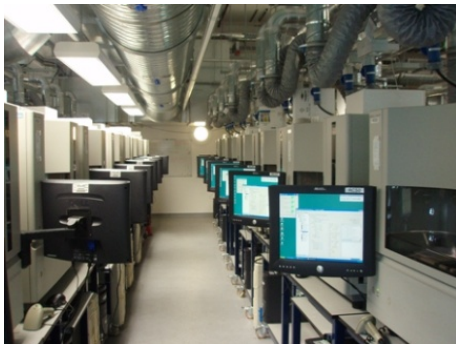
The central objective of our project is to establish associations between the genes of the human intestinal microbiota and our health and disease. We focus on two disorders of increasing importance in Europe, Inflammatory Bowel Disease (IBD) and obesity. The incidence of IBD has been growing steadily during the past five decades in Western Europe, and is now increasing dramatically in Eastern Europe. IBD is becoming an important burden on the young populations. The global epidemic of obesity is well recognized and imposes a huge and rapidly growing challenge for the public health services.

To reach our central objective a number of activities must be carried out in an integrated way. First, we need to establish an extensive reference catalog of microbial genes and genomes present in the human intestine. The second challenge is to develop tools to determine which genes and genomes of the reference catalog are present in different individuals and at what frequency. Third, it is necessary to gather cohorts of individuals, some sick and some healthy, and determine which genes and genomes they carry. Also, it is crucial to develop bioinformatic tools to store, organize and interpret this information, and thus establish associations between intestinal microbiota and health and disease. In addition, the function of bacterial genes associated with disease need to be studied to help understand the underlying mechanisms and host/microbe interactions.

Our project must be integrated in the world we live in. For this purpose, the necessary international cooperation and coordination in the human Metagenome field will be facilitated, by active participation in the International Human Microbiome Consortium (IHMC), the transfer of technology to industry will be promoted, via an appropriate Stakeholder platform, and the information about the project will be presented and made easily accessible to the general public.



Catalog of genes and genomes of intestinal microbes



We follow a two-pronged approach to generate the catalog of intestinal genes and genomes. One is to determine the complete sequence of genomes from bacteria that we know how to cultivate. As they represent only a minority, about 20 %, of all intestinal bacteria, we will also sequence the total bacterial DNA prepared from stool samples. This provides information about genes of bacteria that we cannot cultivate.

Our objective is to sequence full genomes of 100 bacterial species and the total stool DNA of a cumulative length corresponding to some 1000 bacterial species. The first thirty cultivable species have

been chosen, the genomes of the first three have been sequenced and fifteen more are under way. The choice of species has been coordinated within IHMC, and in particular the USA National Institutes of Health, which supports sequencing of intestinal bacterial species carried out within the Human Microbiome Project.

We plan to sequence total stool DNA from different individuals, in order to maximize the reference gene catalog, as different individuals carry different microbiota. Bacterial diversity of 35 individuals from the IBD and obesity cohorts was examined and the 8 individuals with the most divergent microbiota were identified. The cumulative length of total stool DNA sequence determined thus far from three of these individuals corresponds to about 200 bacterial genomes.

To establish the most extensive reference catalog of intestinal genes and genomes, which will be of universal value, we plan to include sequence information generated not only in MetaHIT but also in other large metagenomic programs.

Microbial genes in different individuals

Two technologies are being used to determine the presence and frequency of microbial genes in different individuals, which we term gene profiles. One is based on high density micro-arrays and the other on very high throughput DNA sequencing. Both will use information from the reference gene catalog.

A prototype procedure to create high density arrays from a complex gene catalog has been developed and is being tested. The present array should allow the detection of over 850 000 microbial genes.

High throughput DNA sequencing has been calibrated on DNA samples of known composition, and has been used to analyze samples from 124 individuals. Procedures to compute and compare gene profiles for different individuals are being tested.

Patient cohorts and microbial profiling

IBD includes two diseases, Crohn's (CD) and Ulcerative colitis (UC).

Two studies are planned for each. One will compare patients in remission with healthy relatives (30 for each disease), the other will follow the 30 patients over a 2 year period, during which about 50 % are expected to relapse. The first aims to identify bacterial genes associated with the disease, while the second aims to identify genes associated with relapse. This information could be used to predict risk of disease and of relapse, which would be of great value for diagnosis and preventive treatment. These studies have

been approved by the ethical committee of the Val d'Hebron Hospital in Barcelona, Spain, where the patients are followed. Currently the cohorts are of 15 patients and 9 family matched controls for the UC





studies, and 8 patients and 9 family matched controls for the CD studies. Microbial DNA from 24 individuals has been analyzed by the high throughput sequencing technology.

Obesity is associated with a number of co-morbidities, such as diabetes, hypertension and cardiovascular disease. These co-morbidities are found typically in individuals with visceral rather than subcutaneous obesity. We plan to compare 60 individuals of each type with 60 healthy individuals, aiming to detect bacterial genes associated with obesity which could be used as risk predictors. The study has been approved by the ethical committee of Copenhagen, where the individuals are being followed at the Novo Nordisk Steno Diabetes Center. The cohort consists of 46 lean individuals and 40 obese individuals at present. Microbial DNA from all of them has been analyzed by the high throughput sequencing technology.

Information organization and analysis



Sequence information must be interpreted (annotated) in terms of genes, proteins and the functions they perform. An automated sequence assembly and annotation pipeline was developed to analyze sequence data from the human intestinal metagenome. This allows the efficient assembly of individual sequences into longer "contigs" and their efficient annotation. The function of up to 75 % of the genes can be deduced in this way, which is similar to the proportion of genes in individual genomes that it is possible to annotate automatically. The automated annotation is very

consistent, which is of great importance for comparing different metagenomes. The pipeline was used to process sequences generated within MetaHIT, as well as those from other worldwide projects targeting the human intestinal metagenome, published previously or conducted currently. In total, bacterial sequences equivalent to over 500 full genomes have been integrated successfully. Comparative analysis of this data, which is on-going, will provide unprecedented detail of the human intestinal microbiota. Annotated sequences will be used as the reference gene catalog for microbial profiling in MetaHIT and will also provide a resource for other large projects seeking to understand the impact of various parameters, such as age, alimentation or environment on our microbiome, and in turn on our health and well-being.

Function of bacterial genes associated with disease

Two approaches are planned to uncover the function of bacterial genes associated with disease. One addresses the role of genes in bacteria while the other focuses on the effect of gene products on the host. The former involves inactivation of the genes in appropriate model bacteria and studying the effects of inactivation on the bacterial phenotypes. The latter is centered on over-expression of proteins in standard bacterial hosts and determining the effects of the proteins on appropriate human cell lines. The focus will be on two types of effects, cell proliferation and activation of inflammatory pathways. Twelve different screening procedures have been established for this purpose.



As the genes associated with disease are not yet known, the project will focus initially on the identification of genes involved in the interaction of the microbiota with their human host. These genes will be identified using the screening methods already developed by the MetaHIT project and will incorporate the bioinformatic analysis of metagenome and genome sequence data. This analysis will make use of previously published data as well as the greater body of data generated by the MetaHIT project.



Technology transfer

Two clinical studies of IBD are being carried out at the val d'Hebron Hospital in Barcelona, in association with MetaHIT industrial partners. The first concerns effects of a fermented milk product on the stability of microbiota in ulcerative colitis patients, while the second compares microbiota of patients that respond or do not respond to a particular drug treatment. Both are based on the profiling methodology developed within our project. The former is under way; 40 out of 48 patients have been enrolled and microbial DNA from 15 of them has been analyzed by the high throughput sequencing technology. Approval for the second clinical study from the ethical committee of the Hospital is pending.

Communication and Coordination in the Human Metagenome field

To promote the necessary international cooperation and coordination in the human Metagenome field, MetaHIT took an active role in the establishment of the International Human Microbiome Consortium (IHMC). IHMC was presented at the Kick-off meeting of MetaHIT, (Jouy en Josas, France, April 2008) and its constitution was publicly announced at the second MetaHIT meeting (Heidelberg, Germany, October 2008).



To support and accompany the technology transfer policy of MetaHIT, and communicate with patients' associations, a Stakeholder Platform is being designed. The first action in this direction was to participate in the EuroBioForum organized in Strasbourg in September 2008, in association with the Nestlé Research Centre. This enabled us to present the project to a new audience, introduce potential development and future projects. A first meeting of the stakeholders' platform is organized in conjunction with the MetaHIT 3rd meeting (Barcelona, Spain, March 2009). Participants from various industries, including human health, food and biotechnology are taking part, as are patient associations.

To inform the general public about the field, the project and its achievements, a MetaHIT website (<http://www.metahit.eu>) has been created. In conjunction with MetaHIT meetings, press releases have been issued, resulting in a number of articles in the national and international press, and some radio and TV coverage. In addition, two short human metagenomic documentary films oriented towards general public have been produced and are accessible on the MetaHIT web site.

Expected achievements and their impact

The expected final achievements are the establishment of the methodology to characterize individual intestinal metagenomes and the discovery of associations between bacterial genes and human disease. This should pave the way for the development of novel diagnostic and prognostic tools, based on microbial genes, aiding preventive and personalized medicine. Furthermore, the unprecedented description of the human intestinal microbiota, its dynamics and its interaction with the human host that we aim to provide will lead to a much more complete understanding of global human biology.

Last but not least, by combining methodological and conceptual advances our project will result in, we expect to open avenues for reasoned modulation of our microbiota. The ability to characterize individual intestinal microbiota and follow its evolution with time, coupled to the understanding of the microbe/host interactions, should allow to explore the effects of factors such as food, environment and age on the dynamics of our microbial populations, and to develop interventions to optimize these populations. This should open novel possibilities to improve human health and well-being.