Huiying Li
Institute for Genomics and Proteomics
University of California, Los Angeles

Teaching Statement

Teaching experience

I have worked as Teaching Assistant at two universities. I taught the course of Experimental Nutrition Laboratory for Drs. Barry Shane and Greg Aponte at the Department of Nutrition at UC, Berkeley in 1999, and Biochemical Methods I, a biochemistry laboratory course, at the Department of Chemistry and Biochemistry at UCLA for three quarters from 1999 to 2000. I have also mentored several graduate students and undergraduate students in research laboratories. I received an ‘A’ for the teaching course at UC, Berkeley in 1999, and the Excellence in Teaching Award at UCLA in 2000.

Teaching philosophy and interests

From my own learning and teaching experience, I realized that getting interested is the first step leading to successful learning. Therefore, my teaching philosophy is to make the subject of the course interesting and attractive to students with the help of a careful design of the course and stimulating discussions in class.

Based on my background, I am interested in teaching courses in biochemistry, molecular biology, and bioinformatics. In the past I have loved teaching very much. In my opinion, teaching is one of the most rewarding careers in the world. Therefore I look forward to the opportunity of teaching in the future.

Huiying Li
Institute for Genomics and Proteomics
University of California, Los Angeles

Previous Research Summary

During my Ph.D. and postdoctoral training, my research has focused on three main areas with the common theme of protein functions and protein-protein interactions: comparative genomics, structural biology using X-ray crystallography, and computational analysis of gene expression data.

Comparative genomics

In the laboratory of Dr. David Eisenberg, I have studied protein functions and interactions using bioinformatic approaches. I developed a four-step method to discover parallel pathways from genome sequences using comparative genomics. From ten genomes, we identified 37 cellular systems that consist of parallel functional modules, including known parallel complexes and pathways, and new ones that conventional homology-based methods did not previously reveal[1]. I also applied genome-context-based methods, such as phylogenetic profile, rosetta stone, gene neighbor, and gene cluster, to identify functionally associated proteins. Based on co-evolution relationship, I constructed protein interaction networks of bacterial organisms with an emphasis on metabolic pathways. One of the discoveries from these studies is potential new nitrogenases in the genome of a metabolically versatile bacterium, *Rhodopseudomonas palustris* (Fig 1)[1].

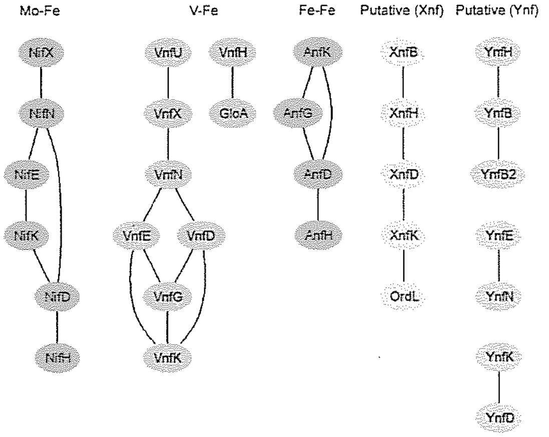

Fig 1. Five parallel nitrogenase pathways were discovered from *Rhodopseudomonas palustris* genome. The right two pathways are putative new nitrogenases.

To further explore protein sequence space, I have taken part in the analysis of Global Ocean Survey sequences generated by the Sorcerer II Expedition, a metagenomic sequencing project of the microorganisms in the ocean. We analyzed ~17 million sequences from thousands of oceanic organisms[2]. We have inferred functional linkages between protein families that have many diverse examples of conserved same-strand adjacency in sequencing reads and partial assemblies. We are able to assign functional associations for protein pairs without sequence homology and infer functions for novel protein sequences. We also studied the abundance, sequence diversity, structure conservation and geographic distributions of two protein families that are involved in metabolism[2] *(collaboration with J. Craig Venter Institute).*

Structural biology

In the laboratory of Dr. David Eisenberg, I was also interested in studying protein functions and interactions from the structural aspect. Using X-ray crystallography, I determined the subunit structure of eukaryotic glutamine synthetases and solved the structures of RuBisCO-like proteins (RLP) from two bacterial organisms. The biochemical function of RLP was not previously known. We studied the active site of RLP and compared its structure with three other forms of RuBisCO. We also analyzed the functional linkages of the protein using bioinformatic approach. By combining X-ray crystallography and bioinformatics, we suggested that the RLP from *Chlorobium tepidum* is capable of catalyzing enolization, but not carboxylation, and that it coevolved with enzymes in the bacteriochlorophyll biosynthesis pathway[3] *(collaboration with Dr. F. Robert Tabita, the Ohio State University).*

Gene expression analysis

In collaboration with Dr. Robert Modlin at UCLA, I have studied protein functions and interactions using gene expression data. I applied several statistical methods and pathway-based analysis to study mycobacterium infection diseases, including classification of disease type[4], predicting clinical courses of diseases and outcomes of treatment. By using supervised and unsupervised analyses of the microarray data from patient samples, we were able to correctly classify the clinical types of leprosy[4]. From studying the most differentiated genes in two types of leprosy, we revealed the functions of a family of previously uncharacterized receptor molecules, the leukocyte immunoglobulin-like receptors (LIR), in human innate immune response. Experimental results showed that LIR-7 shifts cytokine production, therefore suppresses innate host defense mechanism in leprosy patients with severe infection and high load of bacteria in the cell.

To discover the signaling pathway via which Toll-like receptors (TLR) instruct adaptive immune response in monocytes, I analyzed correlated gene expression of known protein ligand-receptor pairs. My analysis revealed that the expressions of two cytokines IL-15 and GM-CSF were well correlated with their receptors across different samples, and suggested that IL-15 and GM-CSF may function as autocrine and paracrine signaling molecules[5]. Follow-up experiments demonstrated that IL-15 and GM-CSF function as the signaling molecules to mediate TLR induced differentiation of monocytes into macrophages and dendritic cells (Fig 2). Lack of TLR-induced GM-CSF signaling may explain why some patients with the pathogen-susceptible form, lepromatous leprosy (L-lep) do not upgrade towards the pathogen-resistant form, tuberculoid leprosy (T-lep).

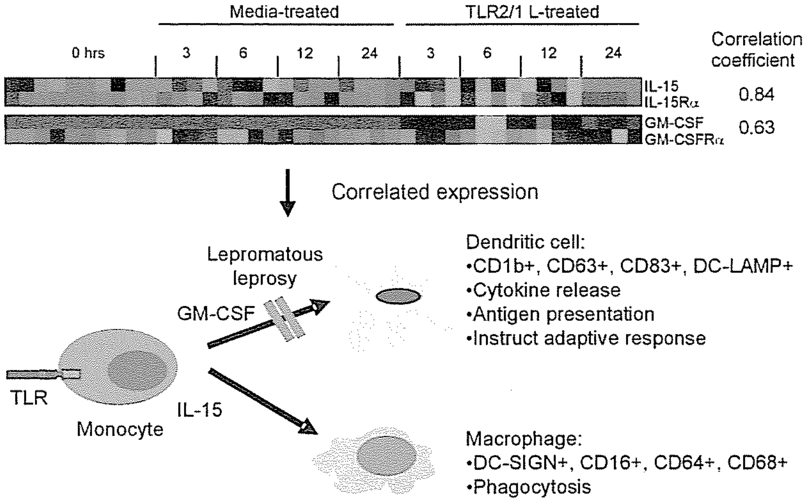

Fig 2. By analyzing correlated expression of protein ligand-receptor pairs, we discovered the pathway via which Toll-like receptor activation leads to the differentiation of monocytes into dendritic cells, with antigen presentation function, and macrophages, with phagocytic capacity. In patients with the lepromatous form of leprosy, TLR-induced monocyte differentiation into dendritic cells is absent.

To distinguish the TLR2/1 signaling pathways in different cell types upon activation, I applied two-way ANOVA in microarray data analysis. We discovered that the expression of the vitamin D receptor (VDR) is significantly up-regulated in monocytes but not in dendritic cells upon TLR2/1 activation[6]. Follow-up experiments showed that up-regulation of VDR leads to induction of an antimicrobial peptide, LL37, in monocytes and macrophages, but not in dendritic cell. Clinical studies also showed that the induction of LL37 was correlated with serum vitamin D levels in different populations. These data support the model that in human monocytes and macrophages, the antimicrobial activity induced by TLR2/1 is mediated by VDR. The discovery of the function of vitamin D in immune defense provides a molecular explanation for the clinical application of UV light treatment to tuberculosis patients in the late 1800's and early 1900's. The discovery also has clinical significance in reducing mycobacterium infection, which is one of the major causes of death in Africa, Southeast Asia, and South America.

1. **Li, H.**, Pellegrini, M. and Eisenberg, D. (2005) Detection of parallel functional modules by comparative analysis of genome sequences. *Nat Biotechnol* 23, 253-60.
2. Yooseph, S., ..., **Li, H.**, et al. The global proteome: leveraging environmental survey data to build a more comprehensive view of protein space. *Manuscript in preparation.*
3. **Li, H.**, Sawaya, M.R., Tabita, F.R., et al. (2005) Crystal structure of a RuBisCO-like protein from the green sulfur bacterium Chlorobium tepidum. *Structure (Camb)* 13, 779-89.
4. Bleharski, J.R.*, **Li, H.***, Meinken, C.*, Graeber, T.G.*, et al. (2003) Use of genetic profiling in leprosy to discriminate clinical forms of the disease. *Science* 301, 1527-30 (*equal contributions).
5. Krutzik, S.R., Tan, B., **Li, H.**, et al. (2005) TLR activation triggers the rapid differentiation of monocytes into macrophages and dendritic cells. *Nat Med* 11, 653-60.
6. Liu, P.T.*, Stenger, S.*, **Li, H.**, et al. Activation of human TLR2/1 triggers a vitamin D receptor-dependent antimicrobial response. *Submitted to Science* (*equal contributions).

Huiying Li
Institute for Genomics and Proteomics
University of California, Los Angeles

Research Proposal

My research interest is the molecular mechanism of oncogenesis in relation to the immune response in pathogen-related cancers, in particular, gastric and colon cancers. I propose research projects aiming at three levels: protein-protein interactions, cell-cell interactions, and interactions between host and microorganisms in the community, using genomic, bioinformatic and metagenomic approaches. I plan to analyze gene expression data from cancer patients to study protein-protein interactions and signaling pathways in gastric and colon cancers, and to develop computational tools to aid cancer diagnosis and prognosis. To understand cell-cell interactions between host and pathogen in cancer development, I plan to study correlated expression of host and pathogen genes using DNA microarrays. As a long-term project, I propose to study host-pathogen-flora interactions in the gastrointestinal tract by metagenomic sequencing and gene expression profiling.

Biological background

Infection of human cells by bacteria and viruses and the consequent defense of human immune response against microorganisms have been shown to play a key role in the development of cancers. Approximately 15% of cancers worldwide can be attributed to viral, bacterial, and other pathogens[1]. Gastric carcinoma, colon cancer and hepatocellular carcinoma are a few examples. Gastric cancer is a major health problem and remains a leading cause of cancer mortality worldwide. *Helicobacter pylori* infection is the most important risk factor for gastric cancer, and exists in at least 50% of the population worldwide[2]. Colon cancer is the third most frequent type of cancer in developed countries and remains the second leading cause of cancer deaths in the US[3]. Epidemiologic differences between dietary habits and socioeconomic diverse populations are evident in the incidence of both gastric and colon cancers. This suggests that the formation and development of these cancers are influenced by nutrition, bacterial infection, host immune response, and the micro-environment in the host gastrointestinal tract.

To study the molecular mechanism of oncogenesis in relation to the immune response in gastric and colon cancers, I propose research projects that will focus on the following three levels: (1) protein-protein interactions and signaling in host cells, (2) host-pathogen interactions, and (3) host-pathogen-flora interactions (Fig 1). Studying the mechanism of interactions at these three levels is important for two main reasons: the function of biological molecules may only be understood in the context of interactions with other molecules, and to gain a systematic view of the roles and regulation of proteins and signaling pathways one must consider the state of the cell and its environment.

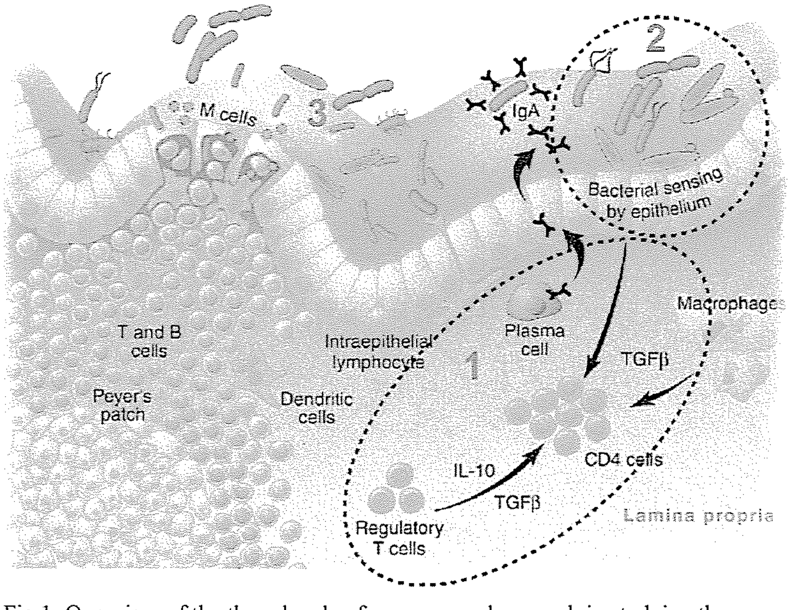

Fig 1. Overview of the three levels of my proposed research in studying the molecular mechanism of oncogenesis in relation to the immune response in gastric and colon cancers. 1. protein-protein interactions and signaling in host cells. 2. host-pathogen interactions in the development of cancer. 3. host-pathogen-flora interactions in the development of cancer – **the full picture.** (Modified from MacDonald T.T., et al., *Science*, 2005, 307: 1920-1925).

Project 1. Using gene expression data to study protein-protein interactions and signaling in cancer with the ultimate goal towards personalized medicine

Gene expression profiling is a valuable tool to aid diagnosis and prognosis of cancer, as well as improve personalized therapies. My proposed study will include developing computational tools for detection and classification of cancer, outcome prediction of treatment, and pathway-based analysis of cellular signaling events using genomic-scale data.

1.1. Detection and classification of cancer - To achieve personalized medicine, it requires first the ability to correctly detect and classify cancer type based on each individual's unique profiles. Several studies have demonstrated the ability of gene expression profiling to distinguish between tumor tissue and adjacent normal mucosa in gastric and colon cancers. However, tumor stages were not all classified correctly, indicating the need for further methodological improvements. In gastric carcinoma the exact preoperative staging is of great importance in choosing the best therapy options. Many different systems for histological classification have been proposed, however, none of them are completely satisfactory. I plan to improve the analysis tools for comprehensive cancer profiling, including large-scale profiling across the spectrum of tumor class, stage, and grade, with the aim of developing a reliable classification method and enhancing prognostic stratification of patients based on gene expression.

Previously, we studied the gene expression profiles of leprosy patients. Although the patients were infected with the same pathogen, *Mycobacterium leprae*, their symptoms represent a broad spectrum, from limited, self-curing with very few lesions (T-lep type) to many lesions with high load of bacteria (L-lep type). To understand the underlying mechanism of different immune responses seen in patients, we analyzed the gene expression data from patients of T-lep and L-lep types using statistical methods, including principal component analysis, hierarchical clustering, permutation, leave-one-out cross validation and weighted gene-voting. We were able to correctly classify the clinical types of leprosy of all patients (Fig 2). A misdiagnosed patient was also discovered from the study[4] *(collaboration with Dr. Robert L. Modlin, UCLA).* The computational tools that we developed in this study can be applied and improved in the future study of cancer samples.

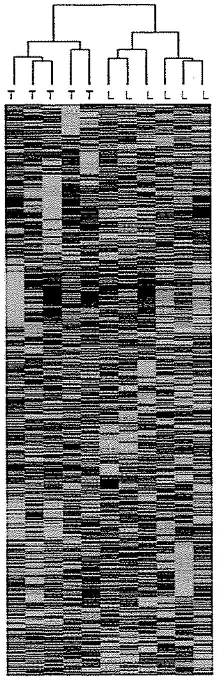

Fig 2. Leprosy patients can be classified based on distinct gene expression pattern.

1.2. Outcome prediction of treatment - One of the toughest tasks faced by physicians is assessing which cancer patients are likely to respond to treatment and which are not. The ability to predict the clinical course of cancer and the outcome of treatment based on gene expression profiling will improve tailored cancer treatment and prognosis. A number of studies in colon cancer cell lines highlight the potential of gene expression profiling for prediction of response of colon cancer to chemotherapeutic agents. It is limited, however, by the fact that the experiments were performed *in vitro*. Therefore, the immediate challenge remains the demonstration of the utility of gene expression profiling for the prediction of response in patients.

Currently, we are studying the progression states of leprosy in patients after multidrug treatment. When treated with multidrug therapy, 15-50% of L-lep patients develop erythema nodosum leprosum (ENL), an immunological serious complication, within the first year. We are comparing the gene expression data of L-lep patients before and after ENL reaction to study the dynamic changes in gene regulation and signaling in the two states. By doing so, we hope to minimize individual differences among patients and to obtain the molecular response profiles of therapeutic agents, which will assist in understanding the molecular mechanism of the disease, development of better prognostics, and improvement of therapies *(collaboration with Dr. Robert L. Modlin, UCLA).*

1.3. Pathway-based analysis - Pathway-based analysis allows us to interpret the observed expression patterns in terms of biological processes, and offers the possibility of identifying specific pathways that are altered in patient's sample and predicting the likelihood of response to a given therapeutic intervention. We attempt to describe the difference in gene expression as the sum of a few *basis states*. Basis states represent biological processes, such as gene ontology terms, known metabolic pathways and cellular signaling pathways. In the case of cancers, they can be distinct oncogenic pathways. If we can determine the major basis states, whose linear combinations accurately capture the changes or differences in expression seen in data, they will dramatically simplify and enhance our interpretation of gene expression data and guide targeted therapies. Using pathway-based analysis, we are able to distinguish TLR2/1 and TLR2/6 signaling pathways in monocytes, which previously were thought to be identical. The different basis states in these two pathways may explain the differences seen in antimicrobial activities triggered by TLR2/1 and TLR2/6 ligands. As methods for pathway-based analyses mature and become more accessible to biologists, expression profiling will become a more powerful tool *(collaboration with Dr. Matteo Pellegrini, UCLA).*

Project 2. Using genomic approaches to study host-pathogen interactions in cancer development

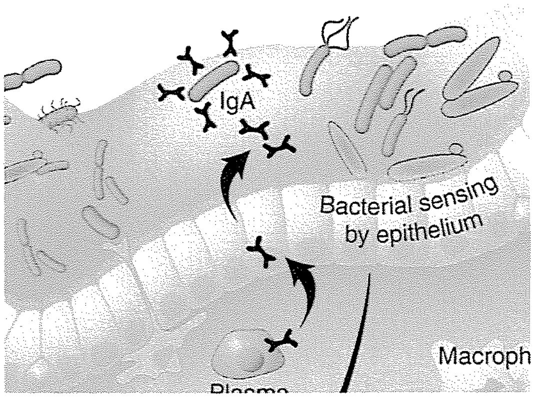

Fig 3. Studying cell-cell interactions between host epithelium and bacteria will help elucidate the development of cancers caused by infection. (Modified from MacDonald T.T., et al., Science, 307: 1920-1925).

Many common cancers develop as a consequence of years of chronic inflammation. Expression profiling of both host and pathogen simultaneously during infection will help elucidate the complex process of infection and consequent development of cancer. Microbial characteristics as well as the type of the inflammatory response have been suggested in dictating the variable outcome of diseases. It has been shown that *H. pylori*-induced persistent gastric inflammation nearly always precedes the development of cancer and is instrumental in initiating a multi-step process leading to carcinogenesis. Therefore, elucidation of the intimate relationship between pathogen and host response is essential in understanding carcinogenesis (Fig 3).

My approach for studying host-pathogen interactions will utilize DNA microarrays to detect simultaneous expression of host and pathogen genes. In studying gastric cancer, I will detect gene expression using both host (human) microarray and pathogen (*H. pylori*) microarray simultaneously at the time of infection and subsequent processes. The gene expression of human gastric cancer cells cultured with *H. pylori* knockout strains was examined previously[5]. The method, however, requires prior knowledge of the genes involved in infection in order to make knockout strains, and it tests bacterial genes only one at a time. If we examine the gene expression of the bacteria at the same time as we examine the host, we will be able to reveal coordinated expression and interplay between the molecules from the two organisms at the entire transcriptome level. With the genome sequences of many human pathogens available, it is feasible to design and manufacture DNA microarrays for individual pathogens. Previously, I have designed DNA oligonucleotide microarrays for two bacterial organisms, *Rhodopseudomonas palustris* and *Fusobacterium nucleatum (collaboration with Dr. James Liao and Dr. Susan Haake, UCLA).* These experiences will help me in future studies of host-pathogen interactions.

Project 3. Using metagenomic approaches to study host-pathogen-flora interactions in the development of cancer

The interaction between host and pathogen usually is not a simple one-to-one battle. The environment, which consists of not only small chemical molecules, but also other living organisms, so-called flora, plays an important role. Disturbances in the delicate balance between host and the environment influence the outcomes of pathogen-host interaction and the development of cancer.

As a long-term project, I plan to study host-pathogen-flora interactions in the gastrointestinal (GI) tract. The GI tract is the site where divergent needs of nutrient absorption, maintenance of a balanced population of microorganisms, and host immune defense collide. Diet appears to influence colonic flora, and differences in dietary habits have been implicated in the risk of developing colon cancer. Studying host-pathogen-flora interactions in the GI tract may permit future utilization of fecal flora as markers for screening and diagnosis of colon cancer.

My proposed future study of the GI tract flora includes (1) shotgun sequencing the microorganisms in the GI tract, (2) studying gene expression of the flora using metagenomic microarrays, and (3) comparing the flora of healthy donors and cancer patients for developing diagnostic markers and therapies.

3.1. Sequencing the gut - It has been estimated from 16S rRNA sequences that there are more than 800 species and over 7000 strains in the human GI tract (Fig 4)[6]. The abundant presence of bacteria plays a role in the maintenance of human health, as well as in the induction of chronic inflammatory diseases of the GI tract. Research in this field is, however, considerably hampered by the abundance of bacterial species, many of which have not even been characterized, and are difficult to culture. In the last few years, several environmental metagenomic projects based on whole-genome shotgun sequencing have taken place. These studies are not constrained by the presence of unculturable organisms, and provide a global view of the important players in a community. Shotgun sequencing of the gut will offer insights into human GI-tract-specific organism abundance, species diversity and spatial distributions.

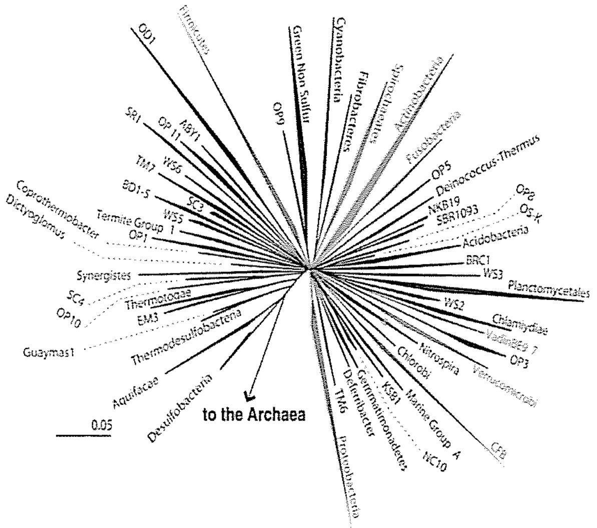

Fig 4. Phylogenetic diversity of bacteria in the human intestine[6].

In the future, I plan to analyze the genome sequences of the microbes in the GI tract to gain insights on novel protein families and functions, pathogenicity islands, interactions and co-evolution of microorganisms in the gut. My previous experience in the study of Global Ocean Survey sequences will help me in handling the large amount of data. At present, there are ~20 complete genome sequences of the microbes in the GI tract available. Comparative analysis of the available genome sequences of the pathogens and microbes in the GI tract will be performed and will help to decode the metagenomic sequencing data of the gut in the future.

3.2. Gene expression profiling of the flora using metagenomic microarrays - Although human GI tract contains all three domains of life – bacteria, archaea, and eukarya, diversity at the division level is among the lowest for any ecosystem[6]. Only 8 of the 55 known bacterial divisions have been identified in the gut. The dominant divisions comprise ~60% of bacteria in feces and the mucus overlying the intestinal epithelium[6]. This makes it possible to design genotyping DNA oligonucleotide microarrays[7], which represent the genes from the major bacterial divisions in the GI tract. The gene expression data will give us an overall view of the major bacteria species and the growth balance between them in the gut.

3.3. Comparison of the flora of healthy donors and cancer patients - Metagenomic profiling will enable us to monitor the growth of microorganisms in the GI tract and to compare the flora of different populations, such as healthy donors versus cancer patients. This will help develop screening and diagnostic markers for colon cancer, monitor cancer progression, and improve therapies.

Career goals

Over the next few years I plan to pursue research projects focusing on oncogenesis, in particular, the molecular mechanism of oncogenesis in relation to the immune response in gastric and colon cancers. Gene expression data from patients with gastric and colon cancers will be used to identify aberrant signaling pathways and molecular markers for diagnosis and prognosis. Simultaneous gene expression profiling of both host and pathogen will be performed to study host-pathogen interactions in cancer development. Computational analysis of the genome sequences and gene expression of the microbes in the GI tract will be used to identify pathogenic genes and to study the host-pathogen-flora interactions in relation to carcinogenesis.

My long-term goal is to establish a research group that combines computation and wet-lab experimentation to understand biological systems. As my previous experience has taught me, computational tools can greatly facilitate biologically meaningful discoveries, and enhance the design and interpretation of biological experiments. Experimental approaches will help validate computational discoveries and generate new data, which will stimulate further development of new computational methods.

References

1. Srivastava, S., Verma, M. and Gopal-Srivastava, R. (2005) Proteomic maps of the cancer-associated infectious agents. *J Proteome Res* 4, 1171-80.
2. Pinto-Santini, D. and Salama, N.R. (2005) The biology of Helicobacter pylori infection, a major risk factor for gastric adenocarcinoma. *Cancer Epidemiol Biomarkers Prev* 14, 1853-8.
3. McGarr, S.E., Ridlon, J.M. and Hylemon, P.B. (2005) Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J Clin Gastroenterol* 39, 98-109.
4. Bleharski, J.R.*, **Li, H.***, Meinken, C.*, Graeber, T.G.*, et al. (2003) Use of genetic profiling in leprosy to discriminate clinical forms of the disease. *Science* 301, 1527-30 (*equal contributions).
5. Maeda, S., Otsuka, M., Hirata, Y., et al. (2001) cDNA microarray analysis of Helicobacter pylori-mediated alteration of gene expression in gastric cancer cells. *Biochem Biophys Res Commun* 284, 443-9.
6. Backhed, F., Ley, R.E., Sonnenburg, J.L., et al. (2005) Host-bacterial mutualism in the human intestine. *Science* 307, 1915-20.
7. Wang, D., Coscoy, L., Zylberberg, M., et al. (2002) Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* 99, 15687-92.