Funct Integr Genomics (2002) 1:331–344
DOI 10.1007/s10142-002-0048-4

REVIEW

Sheng-He Huang · Timothy Triche · Ambrose Y. Jong

Infectomics: genomics and proteomics of microbial infections

Received: 2 November 2001 / Accepted: 13 December 2001 / Published online: 1 March 2002

Abstract The completion of genomic sequences is the greatest triumph of molecular reductionism since the discovery of the DNA double helix in 1953. However, the utility of reductionism is becoming limited and holistic approaches, including theories and techniques, are desperately needed in the postgenomic era. In the field of infectious diseases there is an urgent need for global approaches that can efficiently, precisely and integratively study structural and functional genomics and proteomics of microbial infections (infectomics). The combination of new (e.g. DNA and protein microarrays) and traditional approaches (e.g. cloning, PCR, gene knockout and knockin, and antisense) will help overcome the challenges we are facing today. We assume that the global phenotypic changes (infectomes) in microbes and their host during infections are encoded by the genomes of microbial pathogens and their hosts, expressed in certain environmental conditions devoted to specific microbe-host interactions. Global drug responses (pharmacomes) in microbes and their host can be detected by genomic and proteomic approaches. Genome-wide approaches to genotyping and phenotyping or expression profiling will eventually lead to global dissection of microbial pathogenesis, efficient and rapid diagnosis of infectious diseases, and the development of novel strategies to control infections. The key fundamental issue of infectious diseases is how to globally and integratively understand the interactions between microbial pathogens and their hosts by using infectomics. In this review, we focus on the events that are considered important in infectomics.

Keywords Infectomics · Infectomes · Microbial infections · Genomics

S.-H. Huang (✉) · T. Triche · A.Y. Jong
Childrens Hospital Los Angeles
and the University of Southern California, School of Medicine,
4650 Sunset Blvd., Los Angeles, CA 90027, USA
e-mail: shhuang@hsc.usc.edu
Tel.: +1-323-6694160, Fax: +1-323-6602661

Introduction

Infectious diseases caused by bacterial, viral, fungal or parasitic pathogens continue to be the leading cause of morbidity and mortality worldwide despite the availability of effective anti-microbial agents and vaccines over the last 50 years (Fauci 2001; Huang and Jong 2001). According to the World Health Report 2000, there are ten top infectious diseases causing death worldwide (Fauci 2001). These include acute lower respiratory tract infections, human immunodeficiency virus/acquired immunodeficient disease (HIV/AIDS), diarrheal diseases, tuberculosis, malaria, measles, tetanus, pertussis, sexually transmitted diseases (excluding HIV) and meningitis. The continual emergence of previously undescribed new infectious diseases and reemergence of old pathogens will certainly heighten the global impact of infectious diseases in the twenty-first century. Another significant problem in medicine is the development of microbial resistance to antimicrobial drugs, due to the widespread and often inappropriate use of these antimicrobials. There is an emergence of resistant strains of a number of important microbes, including pneumococci, enterococci, staphylococci, *Plasmodium falciparum*, and *Mycobacterium tuberculosis*. As clinical practice trends towards great use of invasive interventions and patients live longer, there is a continually growing proportion of older and immunocompromised patients. These groups are predisposed to opportunistic infections caused by nonpathogenic microbes such as yeast. How to prevent and deal with bioterrorism has become a very serious issue in the twenty-first century. The development of new anti-infective agents against resistant or mutated microbial pathogens has emerged as an urgent defect in modern medicine.

The completion of the human and many other organisms' genomic sequences present both a tremendous opportunity and a vast challenge to us. In the past half-century molecular biology and medicine have been dominated by the reductionist approaches. The completion of genomic sequences is the most important masterpiece of

molecular reductionism. Focusing research on individual virulence genes and the important pathogens has been the traditional approach to human infectious diseases. The large size and complexity of the human genome have dampened the identification and functional characterization of components of the host defense system against invading microorganisms. Currently there is an urgent need for global approaches that can efficiently, precisely and integratively study structural and functional genomics and proteomics of microbial infections (infectomics). Combination of new (e.g. DNA microarrays and proteome) and traditional approaches (e.g. cloning, PCR, gene knockout and knockin, antisense) will be a solution to the challenges we are facing today. The key fundamental issue of infectious diseases is how to globally and integratively understand the interactions between microbial pathogens and their hosts during infection by using infectomics. On one hand we need to identify and characterize the virulence factors, antimicrobial targets, and in vivo survival mechanisms of the invading microorganism; on the other hand we have to dissect the components of the host response that lead to elimination of the invading pathogen and resolution of the disease.

This article highlights the fundamental issues of infectomics (here defined as structural and functional genomics and proteomics of infectious diseases) and infectome (referred to as dynamic expression profile changes in microbes and their hosts during microbial infections). We discuss data mining from human and several microbial genome initiatives, the global approaches to infectome and infectomics, potential novel therapeutic targets for infectious diseases and research shifts from genomic data generation to postgenomic analysis. Reference is made to information sources on the Internet (Table 1).

Table 1 Selected Web sources for infectomics

Host and model organism genomes
Human Genome Project:
http://www.ncbi.nlm.nih.gov/genomes/seq/
Mouse Genome Project:
http://www.nih.gov/science/models/mouse
Microbial pathogen genomes:
http://www-fp.mes.anl.gov/~gaasterland/genomes.html
http://www.tigr.org/tdb/mdb/mdbcomplete.html
http://www.wehi.edu.au
Parasite genome WWW site:
http://www.ebi.ac.uk/parasites/parasite-genome.html
http://www.who.ch/tdr/workplan/genome.htm
Genomics and proteomics
DNA microarrays:
http://www.nhgri.nih.gov/DIR/LCG/15 K/HTML
http://cmgm.standford.edu/pbrown/index.html
http://www.ebi.ac.uk/arrayexpress/
Proteomics:
http://www.expasy.ch/ch2d/
Special issues of infectious diseases
Antimicrobial resistance:
http://www.who.int/emc/amr.html; http://www.earss.rivm.nl;
http://www.cdc.gov/ncidod/dbmd/antibioticresistance
Emerging infectious diseases:
http://www.cdc.gov/ncidod/emergplan

Genomic features of microbial pathogens

Because of the smaller sizes of most microbial genomes, the complete genomic sequencing of many microbial pathogens has been achieved. This has paved the way for global understanding of genomic features and organization of bacterial and fungal pathogens (Strauss and Falkow 1997; McNicholl et al. 2000). Most of parasitic pathogens have relatively larger genomes that make analysis of gene expression and functions more difficult (Johnston et al. 1999). Of 59 microbial genomes that have been completely sequenced, about 24 are those of microbes causing infectious diseases in humans. However, estimates of bacterial diversity from various sources suggest that only very small portions of microbial species are human pathogens (Ochman and Moran 2001). In the evolutionary process of microbial pathogenesis, gene acquisition and deletion, and point mutations are the major events leading to the emergence and evolution of microbial pathogens or commensals. Therefore, microbial pathogens causing infectious diseases usually possess traits or sequence signatures that distinguish them from commensal strains.

Pathogenicity or genetic islands: acquiring genes for microbial pathogenicity

Gene acquisition is prevalent in the evolutionary process of virulence within lineages. Pathogen-specific chromosomal regions are termed pathogenicity or genetic islands. Pathogenicity islands (PAIs) are generally large chromosomal regions (50–200 kb) and usually have a different GC-content from the rest of the chromosome. PAIs can be genetically unstable due to flanking repetitive sequences or IS elements and tRNA loci usually serve as targets for their integration and excision (Ochman and Moran 2001). PAIs have been defined in a number of pathogenic bacteria. These include pathogenic *Escherichia coli*, *Citrobacter freundii*, *Helicobacter pylori*, *Salmonella typhimurium* and *Yersinia pestis* (Strauss and Falkow 1997; Huang et al. 2001). Recently, pathogen-specific genomic islands have been identified from a number of meningitic bacteria including *E. coli* K1 (Huang et al. 2001) and *Neisseria meningitidis* (Klee et al. 2000). However, none of these meningitic bacteria-specific islands fulfill the criteria for PAIs. Eight genetic islands have been identified in *N. meningitidis* (Klee et al. 2000). One island (region 8) is needed to induce bacteremia in an infant rat model of meningococcal infection. However, none of these islands are required for the pathogen's ability to interact with endothelial cells. Recently, we have identified a genetic island of meningitic *E. coli* containing *ibeA* (GimA) that contributes to invasion of the blood-brain barrier (BBB; Huang et al. 2001). GimA encodes 14 novel genes in addition to *ibeA* and consists of four operons, *ptnIPKC*, *cglDTEC*, *gcxKRCI* and *ibeRAT*.

"Black holes": losing genes and degrading genomes for microbial pathogenicity

As described above, the genomic addition of pathogenicity or genetic islands is a major mechanism in pathogen evolution. However, microbial virulence is often multifactorial and mechanisms other than genomic addition may also contribute to the evolution of microbial pathogens. Recently, a number of studies demonstrated that deletion of genes or degradation of certain genomic regions that are detrimental to microbial virulence, i.e. the formation of "black holes", is a complementary but opposite pathway that allows commensal microbes to adapt to a pathogenic lifestyle (Maurelli et al. 1998; Kato-Maeda et al. 2001; Ochman and Moran 2001). For example, *Shigella* spp., the causative agents of bacillary dysentery, differ from the closely related commensal *E. coli* K-12 in the absence of certain genes in the genome of *Shigella* that enhance virulence functions. Lysine decarboxylase (LDC) and OmpT are present in *E. coli* K-12 strains but are uniformly absent in *Shigella* strains. When the genes *cadA* encoding LDC and *ompT* were introduced into *S. flexneri* 2a, virulence became attenuated, and enterotoxin activity and intracellular spread were greatly reduced. Large genomic deletions are also detected in *M. tuberculosis* (Kato-Maeda et al. 2001) and uropathogenic *E. coli* strain 536 (Dobrindt et al. 2001). Recent genome sequencing of *E. coli* 0157:H7 strain EDL933 and comparison with the genome of *E. coli* K-12 strain MG1655 indicate that there are 234 K-islands (0.53 Mb), which are present in MG1655 but not in EDL933, and 177 O-islands (1.34 Mb), which are unique to *E. coli* 0157:H7 (Perna et al. 2001). These results suggest that microbes not only acquire virulence genes but also shed genes via deletions during the evolution of pathogens. Furthermore, the demonstration that genomic deletions can inhibit microbial virulence may lead to new models about microbial pathogenicity. This may also yield clues to new treatments for infectious diseases.

Microvariations: intragenic point mutations and small deletions or insertions for microbial pathogenicity

In addition to virulence evolution through large-scale genomic addition and deletion, genetic polymorphism derived from microvariations (point mutation and small deletions or insertions) in non-virulence genes also contributes to the adaptation of the commensal strains to pathogens (Sokurenko et al. 1998). Recent studies show that there are a number of virulence proteins that share high sequence homology (more than 90%) with the corresponding homologues in the commensal strains. These include FimH (Sokurenko et al. 1998) and several bacterial proteins (IbeB, YijP, AslA and OmpA) contributing to *E. coli* K1 invasion of brain endothelial cells (Huang and Jong 2001). Up to 95% of all isolates of *E. coli* express type 1 fimbriae, which are also called mannose-sensitive (MS) fimbriae. FimH, a 30-kDa lectin-like protein on the tip of type 1 fimbriae, is responsible for the MS adhesive phenotype of *E. coli*. All *fimH* alleles studied so far encode subunits that mediate high levels of binding to tri-mannose structures, but binding to mono-mannose residues among the FimH variants can differ up to 15-fold (Sokurenko et al. 1998). These adhesive variations are solely dependent on their structural differences in FimH that affect receptor specificity of the lectin but do not affect fimbrial morphology or level of fimbriation (Sokurenko et al. 1998). Genetic variations in the FimH lectin of type 1 fimbriae can shift the tropism of *E. coli* toward a urovirulent phenotype. These studies indicate that efficient selection of these very small differences may take place in virulence evolution in certain environmental compartments where a commensal strain becomes adapted to a virulence life-style.

Genomic implications for the host defense systems

With the human genome is a repository of all of the genes controlling immune responses against microbial pathogens. However, in contrast to microbial pathogens, there have been relatively limited advances in our understanding of the molecular basis of host defense. These types of studies are exceedingly complex due to the larger genome, multidimensional and dynamic process involved, and the limited opportunities for controlled observation and experimental manipulation. Host resistance gene discovery can be accelerated by genetic analysis of available data for model organisms. This will lead to identification of human homologues contributing to the microbial pathogenesis. We summarize several examples of human host defense factors in the following section.

Predicting components of innate and adaptive immunity from studies on insects

Innate immunity refers to the first-line host defense against the early phase of microbial infection, which is an evolutionary and ancient defense mechanism. Only quite recently has this first-line defense received renewed attention. This is in contrast to the overwhelming research on immunity in the past few decades that has focused on the adaptive (also called acquired or specific) immunity, which is normally stimulated when an individual is exposed to a microbial pathogen. Clues to the molecular components involved in host defense may be found from genetic or genomic studies of innate immunity in *Drosophila melanogaster*. This organism has been the workhorse for genetic manipulation in eukaryotes in the past century. Genetic analysis in *Drosophila* can easily allow one to directly correlate a mutant phenotype with a specific gene or genes. Comparative analysis of genomics between *Drosophila* and humans may help us identify human homologues contributing to the host

defense responses against microbial pathogens. Furthermore, *Drosophila*, like all other invertebrates, depends entirely on innate immunity against invading microbes (Medzhitov and Janeway 1998).

The detection of common microbial products such as endotoxin and peptidoglycans is one of the basic functions of the innate defense system. These products can be recognized by a class of receptors known as pattern-recognition receptors (PRRs). It has recently been shown that a big family of genes encoding for homologous receptors involved in microbial recognition in other organisms have been identified (Rubin et al. 2000). These include 2 novel homologues of the *Drosophila* scavenger receptors (dSR-CI), 9 members of the CD36 family, 11 members of the peptidoglycan recognition protein (PGRP) family, 3 Gram-negative binding protein (GNBP) homologues, and a number of lectins (Rubin et al. 2000). Genetic studies have demonstrated that Toll signaling pathways are important mediators of immune responses against fungi and bacteria in *Drosophila* and mammals (Hoffmann et al. 1999). Since the report of the first human homologue in 1997, a family of ten related molecules called Toll-like receptors (TLR) have been identified in humans by comparative genomic analysis of *Drosophila* and humans (Medzhitov et al. 1997; Aravind et al. 2001). TLR proteins appear to represent a conserved family of innate immune recognition receptors. They function as the PRRs in mammals, and play an essential role in the recognition of microbial components. A variety of bacterial and fungal products such as lipoproteins have been identified as TLR ligands. TLRs share extensive homology with receptors for the cytokines interleukin 1 (IL-1R) and interleukin 18 (IL-18). These receptors are coupled to a signaling pathway that is conserved in mammals, insects, and plants, resulting in cellular activation, thereby stimulating innate immune defenses. TLRs may also recognize endogenous ligands induced during the inflammatory response (Aravind et al. 2001). Similar cytoplasmic domains allow TLRs to use the same signaling molecules as those used by the IL-1Rs: these include MyD88, IL-1R-associated protein kinase and tumor necrosis factor receptor-activated factor 6.

Compared to the genome of *Drosophila*, the presence of genes encoding specific immunity is one of the most striking differences that has been found in the human genome (Fahrer et al. 2001; Venter et al. 2001). The human genome contains 22 class I and 22 class II major histocompatibility complex (MHC) antigen genes, 114 ORFs encoding immunoglobulins and 59 genes encoding the cognate immunoglobulin receptors. Vertebrate-specific proteins also include the paracrine immune regulator family of secreted 4-alpha helical bundle proteins (cytokines and chemokines). Some of these signal transduction molecules associated with cytokine receptor signal transduction are rarely present in the fly and worm. These include protein domains found in the suppressors of cytokine signaling (SOCS), the signal transducer and activator of transcription (STATs), and protein inhibitors of activated STATs (PIAS; Venter et al. 2001).

Genetic features of human and mouse

Due to the limited opportunities for controlled experimental manipulation in humans, animal models of human infectious diseases are crucial for studies on infectomics. Among the model organisms amenable to genetic analysis, the mouse has been by far the most well-developed and physiologically relevant system for study of human host defense (Paigen 1995; Bedell et al.1997). Despite millions of years of evolutionary separation, close homologous regions are present in many mouse and human gene sequences. Many extended chromosomal regions have maintained the same genes in the same order (Peltonen and McKusick 2001). Data sets of this mouse-human synteny are deposited in the major human and mouse databases and are becoming even more comprehensive as sequencing of the mouse genome advances (http://www.ncbi.nlm.nih.gov/Homology). The comparative map of the mouse and human genomes is the most well developed of all species. A comprehensive summary of mouse/human homology has placed 1,416 loci on both maps by using human physical mapping data and mouse genetic maps (Qureshi et al. 1999). One hundred and eighty one conserved linkage groups have been established. The fine chromosomal and genetic relationships between mouse and human will be defined by comparative analysis of whole genome sequences of both mouse and human.

Mouse models have been successfully used for studies on pathogenesis of various microbial infections. Since mice have short generation times and high breeding efficiency this animal model is extremely useful for dissecting and identifying the most important individual genetic elements that govern the host response to important pathogens (Qureshi et al. 1999). The first requirement for functional genomic study of host defense systems is to identify commercially available inbred strains of mice that can show differential responses to a well-defined infectious challenge. This provides shortcuts to gene identification, unequivocal proof that a mutation in that gene makes the host resistant or susceptible to infection, and rapid dissection of the molecular pathway which the mutant gene codes for. Once different phenotypes are identified, controlled breeding will be performed to determine the mode of inheritance of the phenotype (simple or complex). Linkage analysis can be carried out to correlate the inheritance of resistance or susceptibility to a specific infectious challenge with one or more chromosomal locations. Knowledge of the genomic structure of human and mouse can be used to facilitate localization and identification of the human orthologues of resistance or susceptibility genes identified through animal experiments.

These candidate genes can further be tested for human infection susceptibility through genetic analysis. For example, studies on inbred strains of mice have demonstrated unambiguously that Nramp1 (natural resistance-associated macrophage protein 1) plays an impor-

tant role in host resistance to mycobacterial infections. Based on this information, the human homologue of Nramp1 (NRAMP1) has been cloned and characterized. Mouse Nramp1 and human NRAMP1 share significant sequence homology (85% identity and 92% similarity). The high degree of sequence homology, the presence of homologous regulatory elements within the promoter regions, and similar tissue expression profiles suggest that the NRAMP1 has similar functions in both mouse and human (Qureshi et al. 1999).

Global approaches to microbial pathogenesis

Two major global approaches, genomics and proteomics, have been used to study microbial pathogenesis. Combination of the strengths and advantages of both the genome-based global approaches and conventional tools will allow us to attain a global picture of microbial pathogenesis.

Genomic approaches

The genomes of human and 59 microbes, including those of more than 24 human pathogens, have been completely sequenced. The challenge in this post-genomic era is now how to "mine" these data and translate them into a functional whole. Among the new high-throughput methods, DNA microarrays are prominent for their simplicity, comprehensiveness, and data consistency. Instead of detecting and studying one or a few genes at a time, microarrays allow thousands or tens of thousands of specific DNA or RNA sequences to be detected simultaneously on a small piece of glass or silica only 1–2 cm in lateral dimensions.

Two approaches, cDNA and oligonucleotide microarrays, are commonly used for high-density microarray studies (Triche et al. 2001). In one, cDNA microarrays are constructed by physically attaching DNA fragments (PCR products) to a solid phase. By using a robotic arrayer and capillary printing tips, one can print more than 30,000 DNA fragments onto a microscope slide. Another method is to construct the arrays by use of single-stranded oligonucleotides in silico using photolithographic techniques, primarily used by the commercial company, Affymetrix. In order to measure relative gene expression by using cDNA microarrays, RNA is prepared from the two samples to be compared, and labeled cDNA is made by reverse transcription, incorporating either Cy3 (green) or Cy5 (red) fluorescent dye. The two labeled cDNAs are mixed and hybridized to the microarray, and the slide is scanned. In cases where the green Cy3 and red Cy5 signals are overlaid, yellow spots indicate equal intensity for the dyes. With the use of image analysis software, signal intensities are determined for each dye at each element of the array, and the logarithm of the ratio of Cy5 intensity to Cy3 intensity is calculated. Positive log (Cy5/Cy3) ratios indicate relative excess of the transcript in the Cy5-labeled sample, and negative log (Cy5/Cy3) ratios indicate relative excess of the transcript in the Cy3-labeled sample. A series of cRNA biotinylated samples is prepared to measure the gene expression profiles individually under various conditions, such as disease stages. After several such experiments have been performed, the data set can be analyzed by various clustering or other computational analyses to identify the overall gene expression profile or lists of up- or down-regulated genes. This rapidly developing field is generically termed bioinformatics, and will be essential to understanding the patterns of gene expression generated by microarray studies.

Microarray is an extremely effective way to monitor infectomes in microorganisms and their hosts during microbial infection. The ability to measure the expression level of thousands of genes in any tissue sample allows exploration of biological function on a global scale that was previously impossible. For example, a microarray containing 1,534 ORF of *H. pylori* was used to explore phenotypic changes under acidic stress (Ang et al. 2001). Eighty genes increase their expression levels significantly during acid stress. Among them, *Omp11* encodes a member of the proton-translocating ATPase family. This is in agreement with its function that *Omp11* plays a role in pH regulation by extruding protons from cytoplasm. Monitoring the response of *H. pylori* genes under acid stress may help to understand the pathogenesis of gastric diseases. Similar microarray approaches have been performed to monitor and dissect the infectomes of host cells in response to various infectious agents. These include HIV (Vahey et al. 1999; Geiss et al. 2000), cytomegalovirus (CMV; Zhu et al. 1998), hepatitis C virus (Bigger et al. 2001), coxsackievirus B3 (Taylor et al. 2000), *Pseudomonas aeruginosa* (Ichikawa et al. 2000), *S. typhimurium* (Rosenberger al. 2000); *Toxoplasma gondii* (Blader et al. 2001), and *Listeria monocytogenes* (Cohen at el. 2000). We have recently performed a time-course study of *Cryptococcus neoformans* infection of human brain microvascular endothelial cells (HBMEC), using oligonucleotide microarrays to monitor the infectomes of 12,558 human genes (our unpublished data). Infectome at each time point was compared by scattering analyses (Fig. 1A). An ontology (gene functional classification) analysis reveals gene expression patterns of different subsets of genes within the same functional class. For example, among the 7 time point samples the expression levels of the 29 MHC class II-related genes were compared with one another (Fig. 1B). The changes in expression profiles of MHC class II genes suggest that *C. neoformans* may contain superantigens stimulating the immune system. DNA microarray can also be used to monitor infectomes of pathogens during infection. *P. falciparum* is the parasite pathogen responsible for the most severe forms of human malaria. The parasite at the asexual blood stages is most infective. DNA microarray has been used to monitor infectomes of the parasite during asexual intraerythrocytic development (Ben Mamoun et al. 2001). Co-ordinated upregulation of specific mRNAs

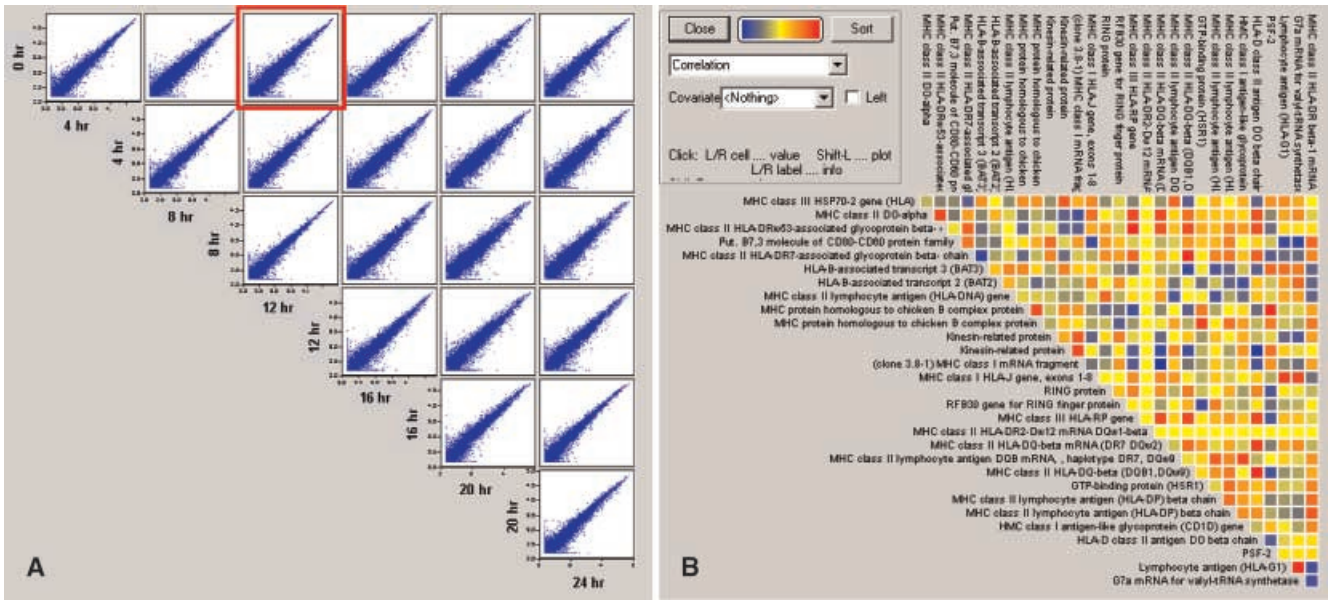

Fig. 1A, B Microarray analyses of infectomes of human brain microvascular endothelial cells infected with *Cryptococcus neoformans* (Jong et al., unpublished data). **A** Infectomes of human brain microvascular endothelial cells (HBMEC) were examined at 0, 4, 8, 12, 16, 20, 24 h postinfection with *C. neoformans* strain B3501. Scatter plot analysis was used to compare 12,558 human genes from the Affymetirx HU95A biochip at different time points, based on the GENETRIX computer program. *Red box* showed an example of the scatter plot from the 0 h and 12 h-sample comparison. The ontology program in GENETRIX sorted the human genes into different assigned functional groups. As an example, among the seven time point samples, the expression levels of the 29 major histocompatibility complex (MHC) -related genes were compared with one another. The *yellow*, *red* and *blue* indicated different corrected distances of the selected paired genes

was found to cluster into functional groups. Nearly all genes showed some regulation over the course of development. These data will help to identify genes with a critical role in parasite progression and multiplication in the human host. Through the use of microarrays for monitoring gene expression profiles, infectomes of microbial and host cells during infection provide global and accurate information for building a comprehensive framework to interpret pathogenic processes.

Proteomic approaches

The ultimate goal of genomics is the global elucidation of the functional partners of genes and genomes (proteins and proteome; Greenbaum et al. 2001). Therefore, as an alternative and complimentary approach to genomic-based technologies, proteomics is essential for the identification and validation of proteins and for the global monitoring of infectomic changes in protein expression during infections. Proteomics to date is based on two-dimensional gel electrophoresis (2-DE) with a combination of mass spectrometry (MS) and computer technologies. It is able to separate and detect several thousand protein spots in a good gel. Commercial robots are now available for staining gels, spot excision, and subsequent proteolysis before MS. Proteins can be further identified by matrix-assisted laser desorption ionization (MALDI), MS and special softwares such as Melanie 3 (http://www.expasy.ch/melanie/). Proteomics has been successfully used for comparative analysis of protein expression profiles of pathogens and the infected host cells. For example, the protein expression patterns of virulent *M. tuberculosis* strains and attenuated vaccine strains were compared using a combination of 2-DE (Fig. 2A) and MS (Fig. 2B; Jungblut et al. 2001). Among 1,800 protein spots isolated, 6 new gene products, not previously predicted by the genomic study of *M. tuberculosis*, were identified with proteomics. A combination of proteomics and genomics has been used to identify unknown regulons and proteins in *Bacillus subtilis* and gram-positive pathogens (Hecker and Engelmann 2000). The data demonstrate that proteomics can be used as a complementary approach to identifying gene products undetected by the genomic tools. The proteome represents the functional status of a cell in response to environmental stimuli and thus provide more direct information on functional changes.

Other tools

As soon as candidate genes are identified by a genomic or proteomic approach, their role in pathogenesis must be confirmed by mutagenesis and subsequent assays that measure virulence. Some methods directly identify virulence genes through genomic library screening, i.e. in vivo expression technology (IVET) and differential fluorescence induction (DFI), in which expression of certain genes induced during infection is required for virulence (Chiang et al. 1999; Cotter and Miller 1998). Differential display is another widely used method for cloning differ-

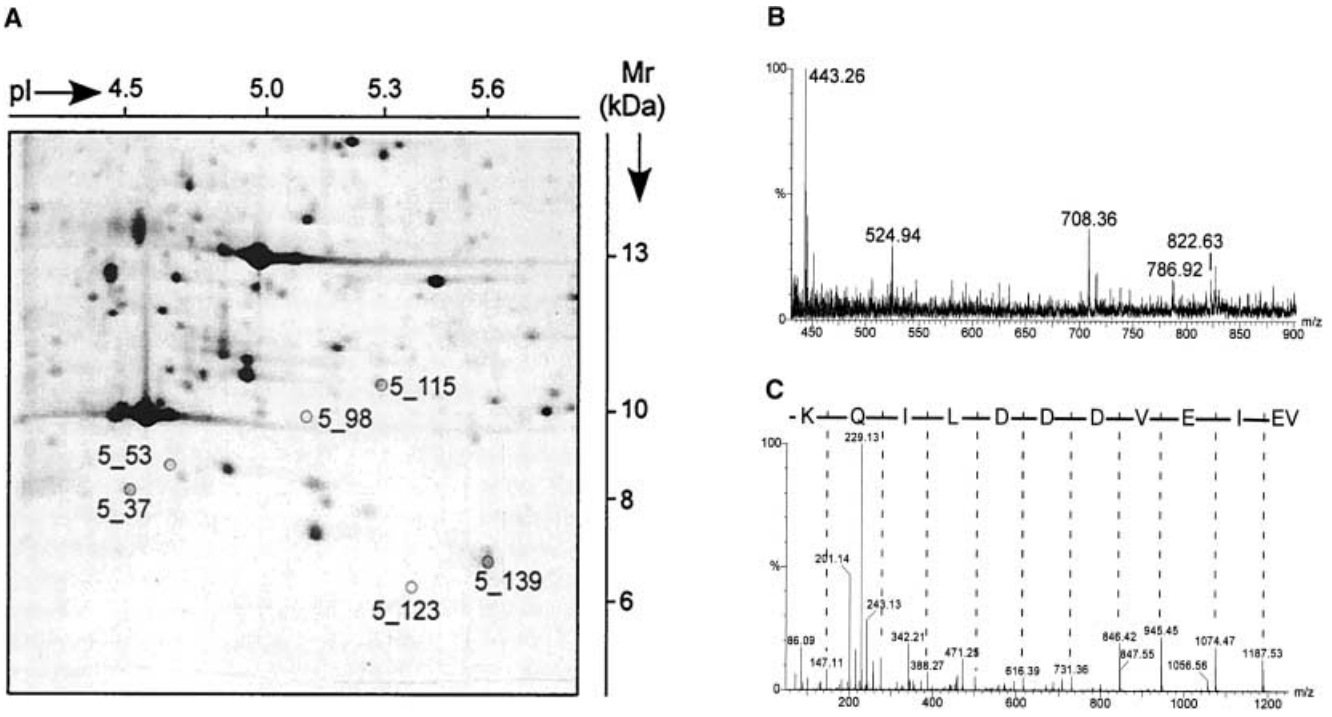

Fig. 2A–C A proteomic analysis showing open reading frames (ORF) in *Mycobacterium tuberculosis* H37Rv that were not predicted by genomics. **A** Identification of previously unpredicted ORFs from *M. tuberculosis* H37Rv by 2-dimensional electrophoresis. Proteins were stained with silver nitrate. An M_r *range* between 6 and 15 kDa and *pI range* between 4 and 6 are indicated. The six de novo spots indicated by *numbers* were sequenced by nanospray mannose-sensitive (MS)/MS. Results revealed that ORFs had not been predicted previously. **B**, **C** MS analysis of spot 5_98. **B** Spectrum of the trypsinized protein. Peptides for sequencing were labeled and fragmented. **C** Fragmentation pattern of the peptide with an *m/z* of 708.36 identified as VEIEVDDDLIQK [reprinted with permission from Infection and Immunity (Jungblut et al. 2001)]

entially expressed genes. In principal, different cDNA is made and displayed by denaturing polyacrylamide gel electrophoresis. Side-by-side comparison of the resulting cDNA patterns between relevant RNA samples (for example, normal and infected specimen or different disease stages) would reveal differences which may represent changes at mRNA levels. The differential display method has to be simple, sensitive, systematic, and reliable. Several modified high-throughput methods for differential gene expression, including SAGE (serial analysis of gene expression) and differential display, are widely used for genomic gene hunting. These include cDNA RDA (representational difference analysis), SSH (suppression subtractive hybridization), SAGE, and TPEA (three prime end amplification; Livesey and Hunt 2000). A challenge in all approaches based on differential RNA display is to isolate high-quality bacterial mRNA. With the combination of genomic approaches with differential display, we can focus on genome-wide searching of differentially expressed genes rather than relying on subsets of genes amplified by PCR (McDaniel and Valdivia 2000).

Diagnostic infectomics

Almost all currently available diagnostic tests for identification of microbial pathogens depend on the techniques by which the microorganisms are obtained, manipulated, and analyzed in the laboratory. These traditional approaches that have been developed based on Koch's postulate have made great contributions to fighting infectious diseases (Casadevall and Pirofski 2000). However, a number of recent studies suggest that there are certain inherent limitations in current diagnostic methods (Relman 1999). These include a lack of a genome-wide survey of infectious agents and host responses, and ignorance of the host environmental conditions that are required by certain microbial pathogens and can not be duplicated in the laboratory. The availability of complete human and numerous microbial genome sequences has profoundly revolutionized the ways of prediction, prognosis and diagnosis of diseases. We assume that infectomes of microbial infections are encoded by the genomes of microbial pathogens and their hosts, expressed in certain environmental conditions devoted to specific microbe-host interactions, and can be used as diagnostic indications to identify specific microbial pathogens. Genome-wide approaches to genotyping and phenotyping facilitate rapid translation of molecular discoveries to diagnostics.

Infectomes of microbial pathogens

Microbial pathogens are exposed to multifactoral and dynamic environmental conditions during an infection cycle and have to globally regulate their gene expression

at both RNA and protein levels (infectomes) accordingly. Presumably, different microbial pathogens have their distinct infectomes that are induced in their hosts during microbial infections. The same pathogen may have distinct infectomes induced in different tissues of the same host. These characteristic features of infectomes can be used for diagnosis. To date, most of the advances in our understanding of human infectious diseases come from analysis of a single or a small number of genes. Recently, DNA microarray and proteomic analyses have been used to globally monitor microbial gene expression profiles under certain environmental conditions or in their hosts.

For example, up to 90% of the primary *T. gondii* infections in pregnant women are not detected by the current methods that are exclusively based on serological screening and PCR (Jungblut et al. 1999). In AIDS patients, *T. gondii* is the major reason for death due to intracerebral lesions. There is an emerging need for early diagnosis of this disease in order that an effective treatment can be given. By using proteome analysis on 2-DE, about 300 spots were resolved in silver-stained gels (7×8 cm). Comparative immunoblotting analyses were carried out on the serum samples from pregnant and non-pregnant women with acute toxoplasmosis, and patients with latent infection. Seven antigens from seven spots have been shown to be potential diagnostic markers that may help to distinguish between acute and latent infections (Jungblut et al. 1999). The same approach has also been used to identify *H. pylori* antigens that may have potential diagnostic or therapeutic applications (Jungblut et al. 2000).

Genotyping of microbial pathogens

Based on genomic features of microbial pathogens, genotyping analyses provide complementary, and sometimes more reliable and accurate information for detecting and identifying microbial pathogens. Among the well-established genotyping techniques are broad-range PCR, representational difference analysis (RDA) and various DNA microarray approaches (Relman 1999; Cummings and Relman 2000). Conserved priming sites among broad groups of microbes are used in the broad-range PCR to amplify genetic loci that provide accurate phylogenetic information for a given microbe. The genes encoding the small and large subunit rRNAs have been commonly used. The broad-range PCR method has been successfully used to detect and identify fungi such as *Candida albicans* directly from human venous blood (Evertsson et al. 2000). The broad-range amplification in combination with the PCR-ELISA kit has been shown to be a sensitive and specific approach for the detection of agents causing reactive arthritis, meningitis or other diseases associated with a limited number of different bacteria (Fischer-Romero et al. 2000). RDA is an alternative approach complementing the advantages and disadvantages of the broad-range PCR. By using RDA, *E. coli* 0157 specific sequences and a new virus (TT virus) have been identified (Springfeld et al. 2000; Allen et al. 2001). Genomewide DNA microarray is potentially the most powerful technique for genotyping microbial pathogens. A GeneChip containing a set of 82 polymorphic oligonucleotides derived from a 16S rRNA gene has been successfully used to accurately identify *Mycobacterium* species, suggesting the potential power of this approach for molecular diagnostics (Cummings and Relman 2000). As more and more microbial genome DNA microarrays become available, global genotyping of microbial pathogens will be feasible and crucial for diagnostic infectomics.

Infectomes of the host

The infectomic changes, including mRNA and protein expression profiles, in the host infected by pathogens are believed to be patterned and stereotyped. These infectomes can be used to distinguish among different infectious agents or different pathogenic mechanisms. It has been proposed that the host gene expression signatures can be used as potential diagnostic markers for infectious diseases (Cummings and Relman 2000). DNA and protein microarray techniques permit us to obtain global diagnostic information on gene expression profiles inside cells. Human cDNA microarrays have been used to globally monitor the host response to viruses (e.g. HIV, and coxsackievirus), bacteria (e.g. *S. typhimurium*, *P. aeruginosa*, and *L. monocytogenes*), and parasites (e.g. *T. gondii*; Blader et al. 2001). Infectomes of human foreskin fibroblasts in response to *T. gondii* infection have been monitored by using human cDNA microarrays consisting of around 22,000 known genes and uncharacterized expressed sequence tags. These studies suggest that the early response does not require parasite invasion and expression of the late phase genes is mainly dependent on the direct presence of the parasite. The difference in the early and late phase gene expression may not only facilitate dissecting pathogenesis of this parasitic disease but also provide important information on the progress of the disease. It is anticipated that the global monitoring of host infectomes induced by pathogens should be much more specific than the detecting of the traditional markers of inflammation, such as cytokines.

Preventive and therapeutic infectomics

Host-pathogen interactions in the development of infection and disease, like other important issues in life, are continuous, complex and multidimensional processes. We propose that the development of microbial infections is determined by the nature of host-microbe relationships. These include host-pathogen, host-commensal, and pathogen-commensal interactions. A holistic balance of these relationships is essential to our health.

However, this balance remains poorly defined, and little attention has been paid to commensal microbes that may be beneficial to the host defense systems. The availability of genomic and proteomic approaches allows for global study of preventive and therapeutic infectomics that leads to holistic solutions to infectious diseases.

Pharmacogenomics

There is a high degree of heterogeneity in the way we respond to medications including antimicrobial therapy. This requires holistic strategies to find the optimal drug therapy for individuals who may have variations in drug transport, drug metabolism, cellular targets, and cellular response pathways. Unfortunately, optimal and individualized therapy for each patient is impossible for many diseases and medications due to the theoretic and technical limitations. In order to solve this problem, a burgeoning field, pharmacogenomics, is born by the combination of pharmacology and genomics. Using the advanced global approaches, including genomics and proteomics, it is feasible to monitor and predict holistic drug responsiveness (pharmacome) of the optimal and individualized drug therapy for each patient. This approach can be used in both drug development and clinical treatment to predict either good or adverse responses to individual drugs (McLeod and Evans 2001). Correlation of microbial and host pharmacomes with drug activity may suggest molecular details of the drug's action. A DNA microarray was used to monitor pharmacomes of *M. tuberculosis* in response to the anti-tuberculous drug isoniazid, a drug that blocks mycolic biosynthesis (Wilson et al. 1999). The drug elicited a unique pharmacome, characterized by pronounced transcription induction of five adjacent genes encoding fatty acid biosynthesis enzymes. Because a known isoniazid target, *KasA*, was among these genes, it was predicted that the adjacent, coregulated loci might be targets for new tuberculosis drugs. Gene expression profiles in drug treated and untreated cells may reveal mechanisms for sensitivity and resistance. As mentioned above, pharmacogenomics has major implications both for drug development and clinical management of infectious diseases.

Vaccine

Prophylactic immunization against infectious diseases by vaccination continues to be an important strategy to enhance the host defense system. Genomic information provides a tremendous opportunity for development of vaccines. The use of high-throughput cloning and expression of candidate genes permits a comprehensive evaluation of all predicted gene products. The candidate proteins can then be directly tested in animals for protection against challenge or time to resolution of infection. The advantages of a genomic approach are that the method does not rely on knowledge of any protein function and it conducts a global search (Zagursky and Russell 2001). An example of using genomics for bacterial vaccine development is the mining of the *N. meningitidis* serogroup B (strain NmB) genome to identify possible vaccine candidates (Pizza et al. 2000). Bioinformatic analyses were first used to survey the genome for prediction of surface-associated proteins. These include transmembrane domains, leader peptides, homologies to known surface proteins, lipoprotein signatures, outer membrane anchoring motives, and host-cell-binding domains such as the tripeptide, arginine-glycine-aspartic acid (RGD). Out of a total of 2,158 ORFs, 7 representative proteins were identified as vaccine candidates (Fig. 3). Interestingly, these 7 surface-exposed proteins are conserved among all 31 *N. meningitidis* strains tested. These vaccine candidates will be moving into the pre-clinical phase of vaccine development. Use of a whole genome approach to the development of new vaccines has also been applied to *M. tuberculosis* (Gomez et al. 2000), *S. pneumoniae* (Wizemann et al. 2001), *Porphyromonas gingivalis* (Ross et al. 2001), and *H. pylori* (Chakravarti et al. 2000). These few examples highlight current approaches whereby genomic analysis and screening can be done in silico before any wet laboratory research.

Anti-bioterrorism

The threat of attack posed by bioterrorism has recently been recognized by the public. Much more attention is being paid to the counter measures against bioterrorism. DNA-based approaches, including DNA vaccines and immunostimulatory synthetic oligodeoxynucleotides (ODN), have been used to develop novel and effective pathogen-specific vaccines and immunostimulatory agents capable of boosting the hosts resistance to bioterroristic agents (Klinman et al. 1999). DNA vaccines require a plasmid carrying the antigen-encoding gene whose expression is regulated by a strong mammalian promoter (Dubensky et al. 2000). A major disadvantage of DNA vaccines is the matter of time. It may take more than a decade to produce, test, and license a new vaccine, whereas mutant pathogens capable of circumventing the efficacy of such vaccines can be generated in a few months (Klinman et al. 1999). The immunomodulatory agents such as ODN can be used as an alternative strategy to broadly stimulate the innate immune system and then improve host resistance against bioterroristic agents. It has been demonstrated that ODN was able to trigger an immune stimulation in mice (Klinman et al. 1999). In addition, gene therapy can be investigated as an alternative approach to controlling microbial infections including diseases caused by bioterroristic agents (Bunnell and Morgan 1998). Mining of genomic data of human and microbes will have a great potential to develop much more powerful counter measures against bioterrorism.

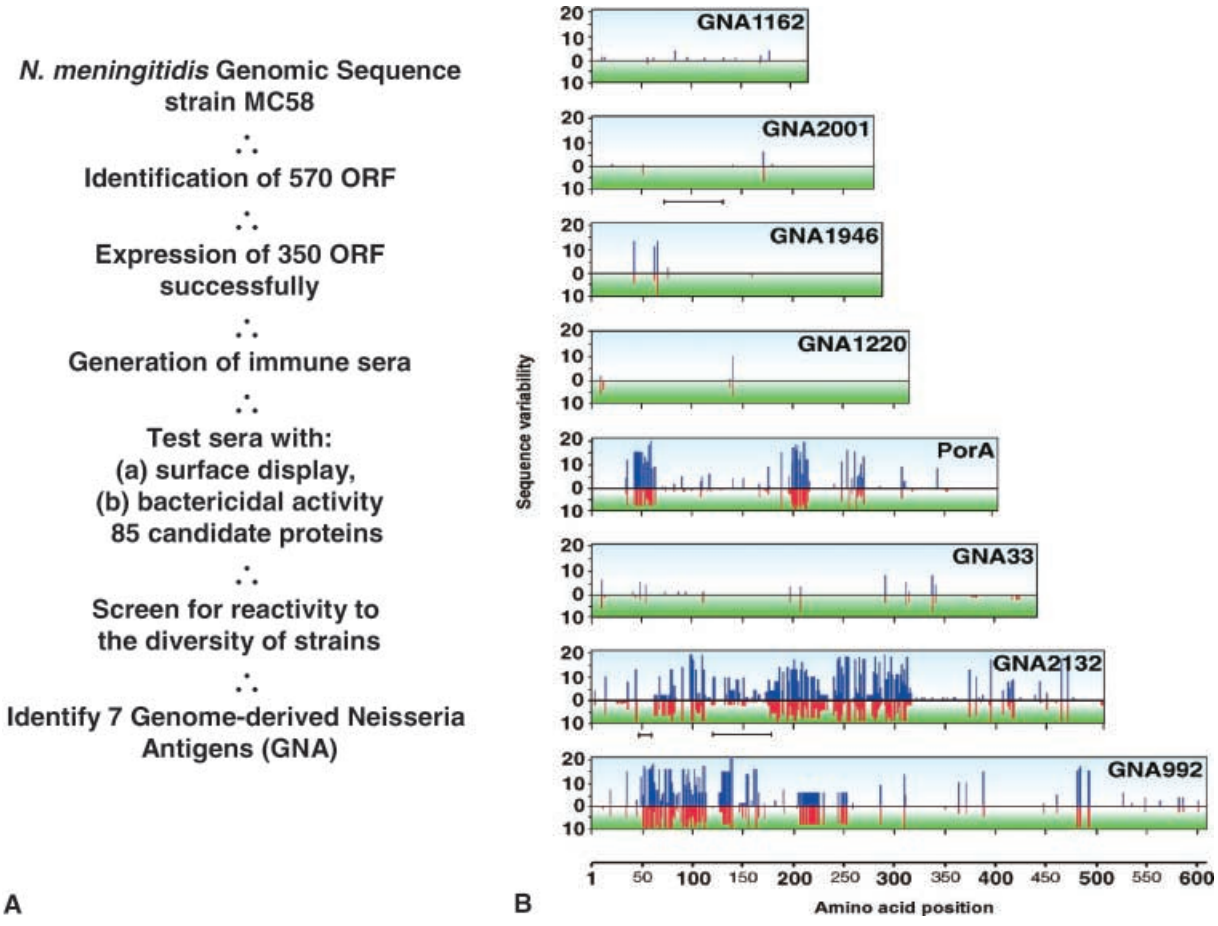

Fig. 3A, B Schematic representation of identification of vaccine candidates against serogroup B meningococcus by a whole-genome approach. **A** The flow-chart of identification of the potential vaccine antigens. **B** Amino acid sequence variability of seven vaccine candidates from *Neisseria meningitidis* (*abscissa* amino acid positions; *ordinate* number of strains analyzed; *GNA* genome-derived *Neisseria* antigens. *Line 0* indicates the sequence of the reference strain MC58. *Blue lines above the 0* represent amino acid sequence differences within the 22 strains of serogroup B *N. meningitidis*. *Red lines below the 0* indicate amino acid sequence variability within the 9 *N. meningitidis* strains from serogroups A, C, Y, X, Z, and W135. *Bars* below GNA2001 and GNA2132 denote regions that are missing from some strains. This figure is adapted from Science (Pizza et al. 2000)

Probiotics: ecological approaches to infectious diseases

From birth to death, we share a benign coexistence with a vast, complex, and dynamic consortium of microbes. Most of our microbial commensals reside in our gastrointestinal (GI) tract (Hooper and Gordon 2001). The GI tract harbors a rich flora of more than 500 different bacterial species. Some of these microparasites have important health functions. These include stimulating the immune system, protecting the host from invading bacteria and viruses, and aiding digestion. The gut microflora, which is essential for human homeostasis, is established rapidly after birth and remains relatively stable throughout life (Alvarez-Olmos and Oberhelman 2001). The GI mucosa provides a protective interface between the internal environment and the constant external challenge from food-derived antigens and microbes.

Several environmental factors may cause an alteration in the composition and effect of the normal microflora. These include use of antibiotics, immunosuppressive therapy, irradiation, hygiene, and imbalance of nutrition. As a result of all the mentioned factors there has been a decline in the incidence of microbial stimulation that may reduce host defense and predispose us to infectious diseases (Alvarez-Olmos and Oberhelman 2001). Therefore, the introduction of beneficial live bacteria into the GI tract (probiotics) may be a very attractive rationale for providing a microbial stimulus to the host immune system against pathogens (Isolauri 2001). Multiple mechanisms of probiotic therapy have been postulated, including production of antimicrobial agents, competition for space or nutrients, and immunomodulation. The microbes frequently used as probiotic agents include *Lactobacillum* and *Bifidobacterium*. For example, *Lactobacillum* spp. is able to attenuate colitis in IL-10-deficient mice; probiotic agents containing *Lactobacillum*, *Bifidobacterium*, and *Streptococcus* spp. are effective in treatment of chronic “pouchitis”, a complication subsequent to surgical therapy for ulcerative colitis (Hooper and Gordon 2001). Recent studies suggest that the use of lactic acid bacteria (LAB) as live vectors is a promising approach for delivering drugs, antimicrobial agents, and vaccines to defined

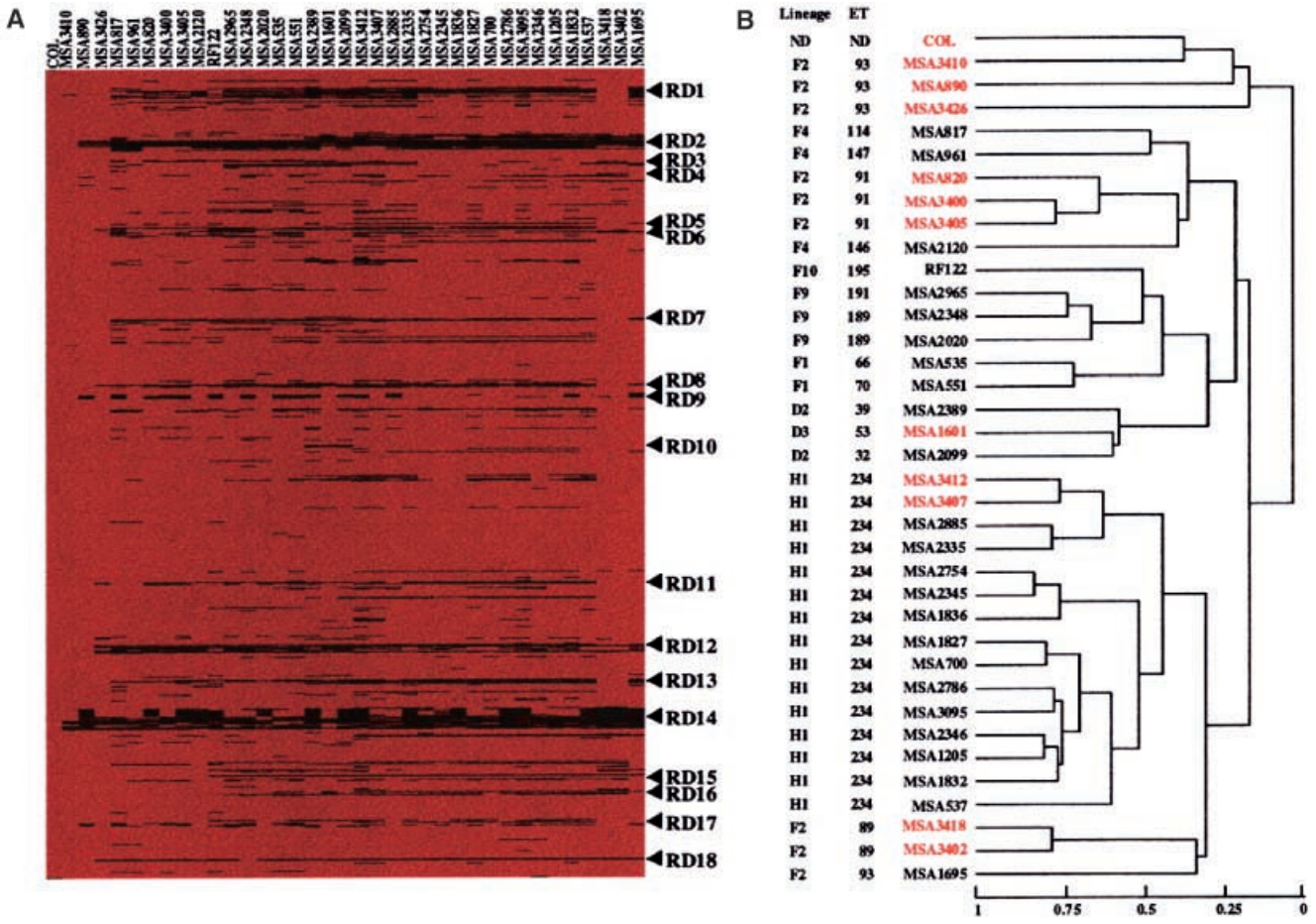

Fig. 4 A DNA microarray showing *Staphylococcus aureus* genetic diversity. **A** Genome-wide distribution of missing ORFs among the 36 *S. aureus* strains investigated. *Red* and *black* represent presence or absence of ORFs, respectively. Regions of difference (*RD*) are marked by *black arrows*. **B** Dendrogram indicating genomic relationships of the 36 strains constructed by complete linkage hierarchical cluster analysis. Strains were grouped with the CLUSTER program on the basis of the presence or absence of ORFs with RDs treated as single loci. The computer program TREEVIEW was used to display the output. Phylogenetic lineage and electrophoretic type (*ET*) of each strain are shown. *Red text* denotes methicillin-resistant *S. aureus* (MRSA) strains. The *scale* indicates the value of the Pearson correlation coefficient at each node. For a pairwise comparison, a coefficient of 1 represents absolute identity, and zero indicates complete independence [reprinted with permission from *Proceedings of the National Academy of Science USA* (Fitzgerald et al. 2001)]

host niches, due to their safety, ability to persist within the indigenous microflora, adjuvant properties, and low intrinsic antigenicity (Alvarez-Olmos and Oberhelman 2001). Dissecting the role of probiotics as modulators of the host defense system will be challenging and may require simultaneous monitoring of host and microbial gene expression profiles during the course of colonization. Therefore, global studies on the collective genome of our normal microflora (microbiome) may be important for biomedical sciences (Hooper and Gordon 2001).

Mining microbial genomes for antimicrobial targets

The development and use of antibiotics to control infectious diseases has been one of the greatest achievements in modern medicine. However, over the last few decades the search for new antibiotics has been difficult. This is due to the limitation of the screening approach that is largely restricted to well-known compound classes active against a standard set of drug targets (Loferer 2000). The limited chemical variability is unable to prevent a serious escalation in drug resistance. Antibiotic therapy also faces major challenges in other aspects. There has been a rapid worldwide increase in the evolution of multi-drug-resistant human pathogens in recent years that correlates with overuse of antibiotics in humans and livestock. The increased percentage of older and immunocompromised patients predisposes humans to opportunistic infections (Fussenegger 2001). These problems will certainly exacerbate the current crisis of antibiotic resistance.

Recent advances in genomics and proteomics have provided a tremendous opportunity to dissolve the present crisis of antibiotic resistance and expand the range of potential antimicrobial targets. Global approaches based on microbial genomes have also facilitated a fundamental shift from direct antimicrobial screening programs toward rational and genomewide target-based strategies (Emilien et al. 2000; Ohlstein et al. 2000; Rosamond and Allsop 2000). Whole-genome DNA microarray analysis

has been used to compare the genomes of variants of the tuberculosis vaccine strain (*B. bacillus* Calmette-Guérin), *M. tuberculosis* H37Rv, *H. pylori*, and methicillin-resistant *Staphylococcus aureus* (MRSA; Behr et al. 1999; Salama et al. 2000; Fitzgerald et al. 2001). These data provided new information about the evolution of these human pathogens and, more importantly, suggested rational approaches to the design of improved diagnostics and antimicrobial agents. For example, MRSA has been a significant problem in clinics because it causes hospital-acquired infections that can be difficult to combat, as most MRSA strains are susceptible only to vancomycin. The origin of MRSA strains has been a controversial issue. The data obtained from DNA microarray analyses demonstrated that the MRSA strains have arisen independently multiple times by lateral transfer of the *mec* element into methicillin-susceptible precursors (Fig. 4). Therefore, this finding unambiguously resolves a longstanding controversy in the *S. aureus* field (Fitzgerald et al. 2001). Meanwhile, some putative virulence factors or proteins mediating antibiotic resistance have been identified. The mining of microarray data for these pathogens will certainly identify many potential antimicrobial targets. Bioinformatic tools have also been used to pick up highly conserved microbial genes lacking a close human homologue. Through comparative analyses of genomes with the powerful bioinformatic tools, eighteen essential *E. coli* genes have been identified as the molecular targets of both the quinolone and macrolide antibiotics (Rosamond and Allsop 2000). In future, combination of the strengths and advantages of both the genome-wide and conventional screening strategies will greatly revolutionize drug discovery.

Conclusion

Global approaches based on genomic sequences have been facilitating a fundamental shift from reductionist approaches toward holistic strategies. The ultimate goal of infectomics is to provide a more global understanding and to integrate the dynamic interactions between microbial pathogens and their hosts during the development of infectious diseases. Combination of the strengths and advantages of both genome-based global approaches and the conventional biological tools will revolutionize the ways to approach infectious diseases. These include: (1) global detection and integrative dissection of microbial and host infectomes for microbial pathogenesis, and diagnosis and treatment of infectious diseases; (2) mining microbial and human genomes for discovery and development of new antimicrobials and vaccines, and dissolving the present crisis of antibiotics; (3) holistic detection and utilization of pharmacomes for the optimal drug therapy; and (4) full exploitation of probiotics as ecological approaches to infectious diseases. The prevention and treatment of diseases including microbial infections will eventually be entering an era when holistic solutions to health problem can be efficiently individualized.

Acknowledgements The authors are grateful to Kelly Ojdana for assistance with analysis of microarray data, Dr. Jonathan Buckley for providing the software Genetrix established by him, Dr. Monique Stins for providing HBMEC, and Dr. Steven Chen for doing DNA microarray experiments. This work was supported by the National American Heart Association Grants-in-Aid program AHA 995046 N (S.H.H.), AHA 0150094 N (A.Y.J.), Public Health Service grants R29-AI40635 (S.H.H.), and the National Cancer Institute UO1 CA88199 (T.J.T.).

References

Allen NL, Hilton AC, Betts R, Penn CW (2001) Use of representational difference analysis to identify *Escherichia coli* O157-specific DNA sequences. FEMS Microbiol Lett 197:195–201

Alvarez-Olmos MI, Oberhelman RA (2001) Probiotic agents and infectious diseases: a modern perspective on a traditional therapy. Clin Infect Dis32:1567–1576

Ang S, Lee CZ, Peck K, Sindici M, Matrubutham U, Gleeson MA, Wang JT (2001) Acid-induced gene expression in *Helicobacter pylori*: study in genomic scale by microarray. Infect Immun 69:1679–1686

Aravind L, Dixit VM, Koonin EV (2001) Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. Science. 291:1279–1284

Bedell MA, Jenkins NA, Copeland NG (1997) Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. Genes Dev 11:1–10

Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284:1520–1523

Ben Mamoun C, Gluzman IY, Hott C, MacMillan SK, Amarakone AS, Anderson DL, Carlton JM, Dame JB, Chakrabarti D, Martin RK, Brownstein BH, Goldberg DE (2001) Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. Mol Microbiol 39:26–36

Bigger CB, Brasky KM, Lanford RE (2001) DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. J Virol 75:7059–7066

Blader IJ, Manger ID, Boothroyd JC (2001) Microarray analysis reveals previously unknown changes in *Toxoplasma gondii*-infected human cells. J Biol Chem 276:24223–24231

Bunnell BA, Morgan RA (1998) Gene therapy for infectious diseases. Clin Microbiol Rev 11:42–56

Casadevall A, Pirofski LA (2000) Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. Infect Immun 68:6511–6518

Chakravarti DN, Fiske MJ, Fletcher LD, Zagursky RJ (2000) Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. Vaccine 19:601–612

Chiang SL, Mekalanos JJ, Holden DW (1999) In vivo genetic analysis of bacterial virulence. Annu Rev Microbiol 53:129–154

Cohen P, Bouaboula M, Bellis M, Baron V, Jbilo O, Poinot-Chazel C, Galiegue S, Hadibi EH, Casellas P (2000) Monitoring cellular responses to *Listeria* monocytogenes with oligonucleotide arrays. J Biol Chem 275:11181–11190

Cotter PA, Miller JF (1998) In vivo and ex vivo regulation of bacterial virulence gene expression. Curr Opin Microbiol 1:17–26

Cummings CA, Relman DA (2000) Using DNA microarrays to study host-microbe interactions. Emerg Infect Dis 6:513–525

Dobrindt U, Blum-Oehler G, Hartsch T, Gottschalk G, Ron EZ, Funfstuck R, Hacker J (2001) S-Fimbria-encoding determinant sfa(I) is located on pathogenicity island III (536) of uropathogenic *Escherichia coli* strain 536. Infect Immun 69:4248–4256

Dubensky TW Jr, Liu MA, Ulmer JB (2000) Delivery systems for gene-based vaccines. Mol Med 6:723–732

Emilien G, Ponchon M, Caldas C, Isacson O, Maloteaux J-M (2000) Impact of genomics on drug discovery and clinical medicine. Q J Med 93:391–423

Evertsson U, Monstein HJ, Johansson AG (2000) Detection and identification of fungi in blood using broad-range 28S rDNA PCR amplification and species-specific hybridisation. APMIS 108:385–392

Fahrer AM, Bazan JF, Papathanasiou P, Nelms KA, Goodnow CC (2001) A genomic view of immunology. Nature 409:836–838

Fauci AS (2001) Infectious diseases: considerations for the 21st century. Clin Infect Dis 32:675–685

Fischer-Romero C, Luthy-Hottenstein J, Altwegg M (2000) Development and evaluation of a broad-range PCR-ELISA assay with *Borrelia burgdorferi* and *Streptococcus pneumoniae* as model organisms for reactive arthritis and bacterial meningitis. J Microbiol Methods 40:79–88

Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM (2001) Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc Natl Acad Sci USA 98:8821–8826

Fussenegger M (2001) The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies. Biotechnol Prog 17:1–51

Geiss GK, Bumgarner RE, An MC, Agy MB, van 't Wout AB, Hammersmark E, Carter VS, Upchurch D, Mullins JI, Katze MG (2000) Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. Virology 266:8–16

Gomez M, Johnson S, Gennaro ML (2000) Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. Infect Immun 68:2323–2327

Greenbaum D, Luscombe NM, Jansen R, Qian J, Gerstein M (2001) Interrelating different types of genomic data, from proteome to secretome: 'oming' in on function. Genome Res 11:1463–1468

Hecker M, Engelmann S (2000) Proteomics, DNA arrays and the analysis of still unknown regulons and unknown proteins of *Bacillus subtilis* and pathogenic gram-positive bacteria. Int J Med Microbiol 290:123–134

Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RA (1999) Phylogenetic perspectives in innate immunity. Science 284:1313–1318

Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. Science 292:1115–1118

Huang SH, Jong A (2001) Cellular mechanisms of microbial proteins contributing to invasion of the blood-brain barrier. Cell Microbiol 3:277–287

Huang SH, Chen YH, Kong GY, Chen SHM, Borodovsky M, Besemer J, Jong AY (2001) A novel genetic island of meningitic *Escherichia coli* K1 containing the *ibeA* invasion gene (GimA): functional annotation and carbon source-regulated invasion of human endothelial cells. Funct Integr Genomics 1:312–322

Ichikawa JK, Norris A, Bangera MG, Geiss GK, van 't Wout AB, Bumgarner RE, Lory S (2000) Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. Proc Natl Acad Sci USA 97:9659–9664

Isolauri E (2001) Probiotics in human disease. Am J Clin Nutr 73:1142S-1146S

Johnston DA, Blaxter ML, Degrave WM, Foster J, Ivens AC, Melville SE (1999) Genomics and the biology of parasites. Bioessays 21:131–147

Jungblut PR, Zimny-Arndt U, Zeindl-Eberhart E, Stulik J, Koupilova K, Pleissner KP, Otto A, Muller EC, Sokolowska-Kohler W, Grabher G, Stoffler G (1999) Proteomics in human disease: cancer, heart and infectious diseases. Electrophoresis 20:2100–2110

Jungblut PR, Bumann D, Haas G, Zimny-Arndt U, Holland P, Lamer S, Siejak F, Aebischer A, Meyer TF (2000) Comparative proteome analysis of *Helicobacter pylori*. Mol Microbiol 36:710–725

Jungblut PR, Muller EC, Mattow J, Kaufmann SH (2001) Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. Infect Immun 69:5905–5907

Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, Small PM (2001) Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res 11:547–554

Klee SR, Nassif X, Kusecek B, Merker P, Beretti JL, Achtman M, Tinsley CR (2000) Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. Infect Immun 68:2082–2095

Klinman DM, Verthelyi D, Takeshita F, Ishii KJ (1999) Immune recognition of foreign DNA: a cure for bioterrorism? Immunity 11:123–129

Livesey FJ, Hunt SP (2000) Functional genomics: approaches and methodologies. In: Hunt SP, Livesey FJ (eds) Functional genomics. Oxford University Press, Oxford, pp 1–7

Loferer H (2000) Mining bacterial genomes for antimicrobial targets. Mol Med Today 6:470–474

Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A (1998) "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc Natl Acad Sci USA 95:3943–3948

McDaniel TK, Valdivia RH (2000) Promising new tools for virulence gene discovery. In: Cossart P, Boquet P, Normark S, Rappuoli R (eds) Cellular microbiology. ASM Press, Washington, D.C. pp 333–345

McLeod HL, Evans WE (2001) Pharmacogenomics: unlocking the human genome for better drug therapy. Annu Rev Pharmacol Toxicol 41:101–121

McNicholl JM, Downer MV, Udhayakumar V, Alper CA, Swerdlow DL (2000) Host-pathogen interactions in emerging and re-emerging infectious diseases: a genomic perspective of tuberculosis, malaria, human immunodeficiency virus infection, hepatitis B, and cholera. Annu Rev Public Health 21:15–46

Medzhitov R, Janeway CA (1998) Self-defense: the fruit fly style. Proc Natl Acad Sci USA 95:429–430

Medzhitov R, Preston-Hurlburt P, Janeway CA (1997) A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. Nature 388:323–324

Ochman H, Moran NA (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292: 1096–1099

Ohlstein EH, Ruffolo RR Jr, Elliott JD (2000) Drug discovery in the next millennium. Annu Rev Pharmacol Toxicol 40:177–191

Paigen K (1995) A miracle enough: the power of mice. Nat Med 1:215–220

Peltonen L, McKusick VA (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291:1224–1229

Perna NT, Plunkett G III, Burland V, et al (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* 0157:H7. Nature 409:529–533

Pizza M, Scarlato V, Masignani V, et al (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820

Qureshi ST, Skamene E, Malo D (1999) Comparative genomics and host resistance against infectious diseases. Emerg Infect Dis 5:36–47

Relman DA (1999) The search for unrecognized pathogens. Science 292:1308–1310

Rosamond J, Allsop A (2000) Harnessing the power of the genome in the search for new antibiotics. Science 287:1973–1976

Rosenberger CM, Scott MG, Gold MR, Hancock RE, Finlay BB (2000) *Salmonella typhimurium* infection and lipopolysaccharide stimulation induce similar changes in macrophage gene expression. J Immunol164:5894–5904

Ross BC, Czajkowski L, Hocking D, et al (2001) Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. Vaccine 19:4135–4142

Rubin GM, Yandell MD, Wortman JR, et al (2000) Comparative genomics of the eukaryotes. Science 287:2204–2215

Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S (2000) A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc Natl Acad Sci USA 97:14668–14673

Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, Struve C, Schembri MA, Hasty DL (1998) Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. Proc Natl Acad Sci USA 95:8922–8926

Springfeld C, Bugert JJ, Schnitzler P, Tobiasch E, Kehm R, Darai G (2000) TT virus as a human pathogen: significance and problems. Virus Genes 20:35–45

Strauss EJ, Falkow S (1997) Microbial pathogenesis: genomics and beyond. Science 276:707–712

Taylor LA, Carthy CM, Yang D, Saad K, Wong D, Schreiner G, Stanton LW, McManus BM (2000) Host gene regulation during coxsackievirus B3 infection in mice: assessment by microarrays. Circ Res 87:328–334

Triche TJ, Schofield D, Buckley J (2001) DNA microarrays in pediatric cancer. Cancer J 7:2–15

Vahey M, Nau ME, Barrick S, Cooley JD, Sawyer R, Sleeker AA, Vickerman P, Bloor S, Larder B, Michael NL, Wegner SA (1999) Performance of the Affymetrix GeneChip HIV PRT 440 platform for antiretroviral drug resistance genotyping of human immunodeficiency virus type 1 clades and viral isolates with length polymorphisms. J Clin Microbiol 37:2533–2537

Venter JC, Adams MD, Myers EW, et al (2001) The sequence of the human genome. Science 291:1304–1351

Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Brown PO, Schoolnik GK (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc Natl Acad Sci USA 96:12833–12838

Wizemann TM, Heinrichs JH, Adamou JE, et al (2001) Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. Infect Immun 69:1593–1598

Zagursky RJ, Russell D (2001) Bioinformatics: use in bacterial vaccine discovery. Biotechniques 31:636–659

Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T (1998) Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. Proc Natl Acad Sci USA 95:14470–14475