Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Clinical Microbiology Newsletter

Vol. 26, No. 20 October 15, 2004

Microarrays and Gene Expression Profiling in Microbiology and Infectious Diseases: A Clinician's Perspective

Joseph B. Domachowske, M.D., Department of Pediatrics, SUNY Upstate Medical University, Syracuse, NY

Abstract

Advances in molecular biology, bioinformatics, and robotics have allowed microarray technology to be used for in-depth, basic science studies in all fields of microbiology. Recently, translation of these basic science applications to clinical microbiology and infectious diseases has also progressed. From a clinical infectious disease perspective, genome-based organism identification, pathogen discovery, and antimicrobial susceptibility testing of problematic organisms offer the potential to yield diagnostic information that may not otherwise become available. Moreover, microarray-based studies have the ability to provide "signatures" of host cell transcriptional responses for individual pathogens and/or groups of pathogens. This type of information has the potential to confirm difficult diagnoses, to monitor responses to therapeutic intervention, or even to predict prognosis and sequelae following an infectious disease. Examples are presented to illustrate ways in which microarray technology has already impacted these areas of clinical microbiology.

Introduction

In the past decade, we have moved from a time when entire research papers were based on the sequencing of a single gene or a single bacterial operon to single manuscripts that describe entire genomes. The availability of this level of genetic information has spawned the terms functional genomics, transcriptomics, proteomics, and even metabonomics. These fields of study describe the large-scale applications of global gene expression profiling, protein analysis, and metabolite analysis. In each case, powerful bioinformatics tools with links to the comprehensive databases are essential to drive the process. This review focuses on the role that gene expression profiling has begun to play in clinical microbiology and discusses potentials for pathogen identification, pathogen discovery, and novel therapeutics discovery. The availability of complete microbial genomes continues to grow rapidly. Some of the more than 200 different complete pathogen genomes and the dates they became available are listed in Table 1. Links to these and other complete microbial genomes, partial sequences, and ongoing pathogen-related genomic projects are available on the web site for The Institute for Genomic Research at http://www.tigr.org.

With the explosion of available microbial genome sequence data, molecular biology approaches for gene expression studies have advanced quickly. While differential display, gene macroarrays, and serial analysis of gene expression have each provided great insights into disease pathogenesis, microarrays are more powerful because of their high throughput potential. Several examples of how this technology can be and is being applied to address translation research questions are presented. Excellent comprehensive reviews discuss the advantages, nuances, and technologic hurdles involved in using these techniques in the context of clinical microbiology (1-5).

Mailing address: Joseph B. Domachowske, M.D., SUNY Upstate Medical University, Department of Pediatrics, 750 E. Adams St., Syracuse, NY 13210. Tel.: 315-464-6331. Fax: 315-464-7564. E-mail: domachoj@upstate.edu

The Technologies

Differential display

Differential display, first described in 1992 (6), is a labor-intensive, semi-quantitative reverse transcriptase PCR (RT-PCR)-based technique that is used to compare mRNAs derived from dif-

  0196-4399/00 (see frontmatter)

ferent experimental conditions. Total RNA is extracted from the control and experimental conditions of interest. Arbitrary primer pairs are used in an RT-PCR-based reaction for each condition, generating a differential display of signal intensities seen side by side. Signals of interest, such as an amplicon band that appears induced or upregulated, are cut from the gel, eluted, and re-amplified. To confirm differential expression, the amplicon is labeled and used as a probe on a Northern blot. Once confirmed, the amplicon is sequenced. While this sounds very inefficient given today's technology, it does offer the unique potential for novel gene discovery. In fact, in 2001, Van den Hoogen et al. (7) identified the first human metapneumovirus genetic sequences using this technology. They studied RNA extracts from cells that appeared to exhibit respiratory syncytial virus-like cytopathic effects but that could not be confirmed to contain any previously known human virus. Human metapneumovirus is now a well-recognized cause of acute respiratory tract infections.

Table 1. Examples of available complete genomes of human pathogens

Year published	Pathogen
1998	*Haemophilus influenzae*
	Mycoplasma pneumoniae M129
	Escherichia coli
	Helicobacter pylori
	Mycobacterium tuberculosis H37Rv
	Treponema pallidum
	Chlamydia trachomatis
1999	*Chlamydia pneumoniae*
2000	*Neisseria meningitidis*
	Campylobacter jejuni
	Vibrio cholerae
2001	*Escherichia coli* O157:H7
	Pasteurella multocida
	Mycobacterium leprae
	Streptococcus pyogenes
	Staphylococcus aureus
	Streptococcus pneumoniae
	Listeria monocytogenes
2002	*Clostridium perfringens*
	Streptococcus agalactiae
	Shigella flexneri
2003	*Coxiella burnetii*
	Bordetella pertussis
2004	*Mycobacterium avium*
	Bacillus anthracis
	Corynebacterium diphtheriae

Macroarrays

The labor-intensive technology of differential display was quickly joined by the more efficient gene macroarray technology. Macroarrays are membranes imprinted with hundreds of cDNA probes, usually consisting of a thematic element, such as genes associated with inflammation. The technique used is familiar, and the number of genes evaluated is targeted, but somewhat limited. Analysis and interpretation of macroarray results is accomplished using standard protocols. This technique offers no potential for novel gene discovery because the probes included on the membrane are of known sequences. The concept of depositing multiple DNA spots representing different genes onto a membrane surface is not new, especially in microbiology. More than 10 years ago, Chuang et al. (8) investigated *Escherichia coli* gene expression on macroarray membranes, and commercial macroarrays have become widely available since their report.

DNA microarrays

A giant leap forward in the assessment of transcription on the genomic scale has been realized with the development and availability of DNA microarrays. These arrays are glass slides or chips containing ordered mosaics of entire genomes as collections of either oligonucleotides (oligonucleotide microarrays) or cDNAs representing

individual genes (cDNA microarrays). Classic Northern and Southern blotting approaches for the detection of specific DNA and mRNA species provide the technical basis for microarray hybridization with fluorescently labeled cDNA, while the application of robotics to achieve high spotting densities on glass slides facilitates the construction of microarrays containing up to 50,000 genes on a single slide or chip.

Like differential display, the process of obtaining gene expression data using microarray techniques starts with RNA extraction from the biologic conditions of interest. An in vitro transcription reaction allows the incorporation of biotinylated nucleotides into the now labeled transcript pool; the transcripts are fragmented by heat and hybridized to the microarray gene chip. The array is washed, stained, and scanned into a database. The scanned images are compared to replicates of data generated from other samples or to samples derived from different experimental conditions of interest. Following rigorous validation, statistical algorithms are applied to determine which genes are expressed. Normalization of expression data per gene and/or per chip allows relative gene expression intensities to be compared across the different experimental conditions. Data reduction software and bioinformatics databases lend enormous power to these kinds of comparisons.

Critics and skeptics of microarray technology argue that genome scale research is largely non-hypothesis-driven science. However, advocates view the technology as positive, because it has already revealed the functions of genes that may have been missed by other conventional approaches and will likely continue to do so. The need for genome-wide approaches to gene expression experiments becomes apparent when one considers that the genome sequence of *E. coli* has been in hand for more than 6 years, but the microbe still contains over 1,000 genes of unknown function. The role of such genes will not be discovered without application of functional genomic technologies combined with creative experiments. Functional genomics allows researchers to make new and unexpected links between the functions of unrelated and hitherto uncharacterized genes and to put forth hypotheses that can be subsequently tested by more traditional methods.

The clinical potential of microarray technology is enormous. Global transcriptional analyses have been used to identify pathogens, to discover new pathogens, to predict antimicrobial susceptibility, and to determine a host's "gene expression signature" in attempts to define markers of disease severity and/or prognosis. While an in-depth review of each of these applications is well beyond the scope of this perspective, an example of how this technology applies to each of these areas is offered.

Applications of Microarray Technology

How can microarray technology be applied to pathogen identification?

In a pioneering study, Behr and colleagues (9) used microarray technology to genetically characterize different variants of *Mycobacterium bovis* bacillus Calmette-Guérin (BCG) strains (part of the *Mycobacterium tuberculosis* complex). For their purposes, they generated a PCR-based array that represented over 99% of the open reading frames from *M. tuberculosis*. Their array was used as a framework for the genomic analysis of BCG strains by hybridizing with labeled genomic DNA. The study documented 16 regions that were deleted in BCG strains compared to *M. tuberculosis* strains, ranging from ~1,900 bases to almost 13,000 bases. Of particular interest were nine regions that represented 61 open reading frames that were present in *M. tuberculosis*, but consistently absent from *M. bovis* strains, including BCG. The authors provided a hypothesis that the lack of a cluster of phospholipase C genes in *M. bovis* genomes might be responsible for the decreased ability of this organism to spread from person to person compared to *M. tuberculosis*.

While these discoveries have obvious implications for microbial pathogenesis and virulence, how do they relate to pathogen identification? The collection of 61 open reading frames that are absent from *M. bovis* isolates is likely to prove valuable diagnostically, because people who are infected with *M. tuberculosis* could theoretically be distinguished from people who have been vaccinated with BCG. This is exactly the type of clinical situation where specific, direct, and reliable pathogen identification could influence long-term therapeutic decisions.

How can microarrays be used to discover new pathogens?

Oligonucleotide microarrays have already been shown to be useful in the discovery of novel agents. Obvious advantages include the possible detection of the cause of an epidemic or the cause of a biological terrorism event. This technique, from a paper published in the fall of 2002 (10) that described the methodology, was immediately applied to the identification of the agent of severe acute respiratory syndrome (SARS). Several months later, the technique was used in the evaluation of patients with SARS. Although electron microscopy showed the presence of a virus, it was RNA from infected cells that was hybridized to an oligonucleotide microarray from J. DiRisi that showed that the agent had a genetic signature most consistent with a novel coronavirus.

On the pan-virus microarray, consistent positive signals at seven loci were identified. Four of the "matches" were similar probes against the 3' loop stem of two different astroviruses, ovine astrovirus and avian nephritis virus, and two coronaviruses, avian infectious bronchitis virus and bovine coronavirus. There were also three positive signals identified that related to the polymerases of three different coronaviruses, including an avian, a bovine, and a human coronavirus. Taken together with the findings of electron microscopy, these results immediately identified the most likely SARS agent as a novel human coronavirus. Within several weeks, the causative agent of SARS was known, and at least 16 complete SARS coronavirus sequences were available in GenBank. Koch's postulates were fulfilled when Fouchier et al. (11) infected primates with the newly discovered virus and they developed a respiratory illness in response to the virus inoculum. Treatment and prevention experiments are under way.

What about antimicrobial susceptibility testing using microarray approaches?

Rapid results using microarray approaches have the potential to simplify and facilitate susceptibility testing,

0196-4399/00 (see frontmatter)

particularly for troublesome pathogens. Diagnostic microarrays have been explored for plasmid- and transposon-mediated resistance in *Staphylococcus aureus, E. coli*, and *Salmonella* spp. They have been used to test for quinolone resistance in *Neisseria gonorrhoeae* and for multidrug resistance in *M. tuberculosis*. They have also been used to identify susceptibility profiles for antiretroviral drugs used against human immunodeficiency virus. Antifungal testing has also been explored in some detail, including the effects of antifungal agents on gene expression in *Candida albicans*.

In an interesting twist on the more straightforward applications of antimicrobial resistance testing using microarrays, Wilson et al. (12) used whole genome arrays to investigate the effect of the antimycobacterial drug isoniazid on gene expression patterns of *M. tuberculosis*. The known mechanism of action of isoniazid is a selective interruption of mycolic acid synthesis. This study served as a model experiment for looking at inhibitors of an essential metabolic pathway by expression profiling. Several genes encoding mycolic acid biosynthetic enzymes were upregulated following isoniazid challenge, both upstream and downstream of the known site of action of the drug. It has been suggested that accumulation of upstream metabolites and depletion of downstream metabolites serve as signals for induction of gene expression. In addition, several genes not known to be involved in mycolic acid biosynthesis were induced, potentially identifying linked pathways that previously had not been appreciated to be interrelated.

What are the potential advantages of using host gene expression signatures to identify new targets for therapeutic intervention, to determine diagnostic markers of infection, or to reveal prognostic markers of outcome?

The potential for using host-cell expression signatures as diagnostic or prognostic markers of infection are profound. First, the technique might permit early detection of exposure to pathogens that are novel or not yet cultivatable. Second, variations in host cell signatures could be used to infer time since exposure. Third, because host cell responses may continue even after the pathogen is eradicated, the method could be used to evaluate infectious disease sequelae. Finally, as already demonstrated in the microarray-based detection of the SARS coronavirus agent, even a single clinical sample has the potential to be used to diagnose infection with a novel agent.

It is becoming well appreciated that innate immune responses to pathogens are patterned and stereotyped. With the advent of DNA microarrays, researchers are now in a position to examine the host-pathogen relationship systematically in much greater detail than had been possible previously. For example, it is possible to document how a cell, tissue, or organ "sees" a pathogen from the viewpoint of gene expression responses. The temporal features of the interactions between the host and the pathogen can be characterized, prognostic markers of outcome can be studied, and novel therapeutic interventions can be identified and tested.

Most microarray-based gene expression studies in humans have searched for genes that are differentially expressed in various pathologic states. How can such a paradigm be applied to host-pathogen relationships? It has been suggested that every pathogen elicits a unique transcriptional response in the host, in part because of its own expression of unique virulence factors. The host responds through a series of events, including secretion of mediators, cell-to-cell interactions, and a cascade of inflammatory events in the infected cells. The recruited inflammatory cells also undergo changes in gene expression signatures, again characteristic for an offending pathogen. By measuring the aggregate gene expression patterns in infected cells or tissues, it is likely that signatures of gene expression patterns will emerge that are not only diagnostic of a specific pathogen or category of pathogens but also predictive of prognosis.

To study the temporal patterns of gene expression during severe pneumovirus infection, we applied microarray technology to RNA extracted from pneumovirus-infected lung tissue (13). Here, using a mouse model of severe pneumovirus infection, we identified proinflammatory markers that were highly upregulated during infection with a virulent virus strain but not during infection with a less virulent strain. One of the receptor-ligand pairs that we found to be highly upregulated was macrophage inflammatory protein 1α (MIP-1α) and its primary receptor, chemokine receptor 1 (CCR1). Using this information, we went on to study the impact of interrupting this pathway on pneumovirus-associated inflammation, demonstrating the absolute requirement for MIP-1α signaling in pneumovirus-associated inflammation (14,15). Clinical studies in infants infected with the human pneumovirus, respiratory syncytial virus, have confirmed a direct correlation between bronchoalveolar lavage concentrations of MIP-1α and clinical disease severity in humans (16,17). Future studies in humans will determine whether this therapeutic intervention, the concept of which was initially derived from microarray analysis, is effective in the treatment of pneumovirus-associated inflammatory lung disease.

Summary

Microarray technology continues to gain attention as an important and productive basic science research tool in the study of infectious disease pathogenesis. Translating the advantages of this growing field into daily clinical practice is becoming reality as the practical niches for the technique emerge.

References

1. Shaw, K.J. 2002. Pathogen genomics: impact on human health. Humana Press, Totowa, NJ.
2. Ye, R.W. et al. 2001. Applications of DNA microarrays in microbial systems. J. Microbiol. Methods 47:257-272.
3. Kato-Maeda, M. et al. 2001. Microarray analysis of pathogens and their interaction with hosts. Cell. Microbiol. 3:713-719.
4. Lucchini, S. et al. 2001. Microarrays for microbiologists. Microbiology 147:1403-1414.
5. Ivanov, I. et al. 2000. DNA microarray technology and antimicrobial drug discovery. Pharmacogenomics 1:169-178.
6. Liang, P. and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967-971.
7. Van den Hoogen, B.G. et al. 2001. A newly discovered human pneumovirus isolated from young children with

respiratory tract disease. Nat. Med. 7:719-724.

8. Chuang, S.E. et al. 1993. Global regulation of gene expression in *Escherichia coli*. J. Bacteriol. 175:2026-2036.
9. Behr, M.A. et al. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284:1520-1523.
10. Wang, D. et al. 2002. Microarray-based detection and genotyping of viral pathogens. Proc. Natl. Acad. Sci. USA 99:15687-15692.
11. Fouchier, R.A. et al. 2003. Aetiology: Koch's postulates fulfilled for SARS virus. Nature 423:240.
12. Wilson, M. et al. 1999. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc. Natl. Acad. Sci. USA 96:12833-12838.
13. Domachowske, J.B. et al. 2002. Differential expression of pro-inflammatory cytokine genes in vivo in response to pathogenic and nonpathogenic pneumovirus infection. J. Infect. Dis. 186:8-14.
14. Bonville, C.A. et al. 2003. Altered pathogenesis of severe pneumovirus infection in response to combined antiviral and specific immunomodulatory agents. J. Virol. 77:1237-1244.
15. Bonville, C.A. et al. 2004. Functional antagonism of the chemokine receptor CCR1 reduces mortality in acute pneumovirus infection in vivo. J. Virol. 78:7984-7989.
16. Garofalo, R.P. et al. 2001. Macrophage inflammatory protein-1α (not T helper type 2 cytokines) is associated with severe forms of respiratory syncytial virus bronchiolitis. J. Infect. Dis. 184:393-399.
17. Harrison, M.A., 1999. Respiratory syncytial virus-induced chemokine expression in the lower airways: eosinophil recruitment and degranulation. Am. J. Respir. Crit. Care Med. 159:1918-1924.

  0196-4399/00 (see frontmatter)