U.S. Department of Veterans Affairs Staff Publications

U.S. Department of Veterans Affairs

2002

New Technologies, Human-Microbe Interactions, and the Search for Previously Unrecognized Pathogens

David A. Relman

Stanford University School of Medicine, relman@stanford.edu

New Technologies, Human-Microbe Interactions, and the Search for Previously Unrecognized Pathogens

David A. Relman

Departments of Microbiology and Immunology and of Medicine, Stanford University, Stanford, and Department of Veterans Affairs Palo Alto Health Care System, Palo Alto, California

Evidence suggests that a significant number of clinically important microbial pathogens remain unrecognized. Observations from the natural world, from patterns of disease in human populations, from the bedside, and from the clinical laboratory all contribute to this body of evidence. A variety of acute and chronic neurologic syndromes illustrate this point; despite features of infection, most cases of aseptic meningitis, encephalitis, and cerebral vasculitis cannot be assigned a microbiologic diagnosis. The development and clinical application of molecular methods have led to the discovery of novel members of the endogenous normal flora as well as putative disease agents. Current challenges include the establishment of criteria for disease causation and further characterization of the human microbiome during states of health. These challenges and the goal of understanding microbial contributions to inflammatory disease may be addressed effectively through the thoughtful integration of modern technologies and clinical insight.

Microbial Diversity and the Limitations of Cultivation Methods

About 25 years ago, explorations of the natural microbial world turned toward extreme environments and exploited the use of newly described molecular approaches for phylogenetic analysis and classification. Recovery of sequence-based signatures of life directly from these environments confirmed revolutionary proposals for three aboriginal lines of descent [1] and led to the realization that nearly all microbial life is resistant to cultivation in the laboratory. With increasing reliance on molecular methods in environmental microbiology, a picture of microbial diversity emerged that currently includes as many as 40 major divisions of bacteria, a broad and cosmopolitan domain of life known as the *Archaea,* and an intertwined early history of endosymbiotic prokaryotes, eukaryotic protists, and lateral gene transfer events [2, 3].

The inadequacies of available cultivation techniques are reflected by the fact that at least 90% of all known cultivated bacterial species lie within just 4 of the 40 divisions, even though many of the other divisions are equally diverse and well populated [3]. That 65% of all published microbiologic research over a 6-year period was related to just 8 bacterial genera dramatically illustrates our strong bias toward bacteria that are amenable to cultivation [4]. Given this recent history, it is of

concern that clinical microbiology continues to rely heavily upon cultivation-based methods, but perhaps not surprising. The internal environmental conditions of the human body are seemingly more familiar to us than those that are external, the internal niches have been the subjects of frequent study, and there is certainly no dearth of known microbial pathogens. But we should not be so complacent.

When traditional diagnostic methods are rigorously applied to syndromes of suspected infectious etiology, such as pneumonia, encephalitis, lymphocyte-predominant meningitis, pericarditis, acute diarrhea, and sepsis, only a minority of cases can be explained microbiologically. In addition, a long list of chronic inflammatory diseases with features of infection remains poorly understood. Thus, it seems fair to speculate that the distribution of known pathogens in only 7 bacterial divisions and the absence of any known pathogens within the domain *Archaea* may represent an imperfect understanding of the true diversity of microbes capable of causing human disease.

Molecular Approaches for Microbial Pathogen Detection and Identification: Seeking Signatures

In an effort to avoid reliance on cultivation and to establish alternative and complementary approaches, one might view the goal in microbial detection and pathogen discovery as a quest to identify molecular signatures of infection. These signatures must be reliable for identifying a microorganism and for establishing the relationships of a previously uncharacterized organism with those previously characterized. Molecular signatures can be based directly on the features of the microbe itself or upon the features of the host response to a pathogen. There are a variety of methods and techniques with which to acquire each of these two types of signatures [5] (see figure 1).

Pathogen as source of signature

● **broad range PCR**

● **microbial/viral survey "phyloarray"**

● **subtractive/comparative methods**

　　• **representational difference analysis**

　　• **differential display**

● **expression or phage display library screening**

　　(using host antisera or T-cells)

● **small molecule or protein detection (e.g. with mass**

　　spectroscopy)

Host as source of signature

● **host genome-wide transcript profiling (e.g. using**

　　microarrays)

● **host protein profiling (e.g. using microarrays, or mass**

　　spectroscopy)

Figure 1.　Pathogen discovery by seeking molecular signatures. PCR, polymerase chain reaction. Modified from figure 4 in [6].

Genomic sequence is the most frequently used "currency" in the identification of microbial signatures, and broad-range (or consensus) polymerase chain reaction (PCR) is the most practical tool for generating this currency [5]. Furthermore, rDNA is among the most useful genome sequences from which organismal ancestry and interrelationships can be reliably inferred and pathogen discovery approaches designed [7, 8]. With recent improvements in the speed at which primary genome sequence can be acquired and analyzed, other detection or screening formats may become widely available and additional regions of microbial (and viral) genomes more commonly targeted. For example, high-density microarrays of oligonucleotides or amplified DNA products can be designed to screen complex pools of microbial nucleic acid for specific agents in a massively parallel and efficient manner [9]. This technical platform obviates the need to clone and sequence large numbers of variant microbial molecules. This is particularly relevant to the analysis of clinical specimens with a significant burden of "background" microorganisms (see further discussion below) and facilitates more sophisticated uses of pattern recognition analysis as a tool for microbial signature identification.

There are at least two alternative kinds of approaches for detecting diagnostic signatures of microbial origin that incorporate features to help discriminate between signal and noise. The first relies on differential analysis of microbial sequences in specimens from host sites that are involved and uninvolved

in disease. Differential display is a screen for differences in sequence diversity and abundance between involved and uninvolved sites; representational difference analysis (RDA) in essence selects for sequences of differential abundance by use of PCR [10, 11]. Both share the disadvantage that the sequences of differential abundance that are revealed by these methods may not be useful markers for microbial identification. A study of Crohn's disease illustrates this problem [12]. The investigators applied RDA to specimens from involved and uninvolved segments of intestinal tissue from a patient with Crohn's disease and obtained random fragments of bacterial genome. Although the encoded molecules may be relevant to the disease process [13], the identity of the source organism(s) remains unknown. The second kind of alternative approach relies upon the host immune response to identify sequences that may originate from a putative pathogen. These techniques include screening of expression or phage display peptide libraries with patient antisera or reactive T cells [14].

Despite a preponderance of efforts to discover and detect microbes by targeting them directly, the nature of the host response to infection offers attractive features for pathogen detection and classification that are unique and complementary. In theory, the host response can serve as a source of microbial signatures that are by definition clinically relevant—that is, it may provide signatures that distinguish between infection and disease. The logic behind this speculation lies in the fact that regulation of gene expression is one mechanism at the basis of host pathology. Gene expression patterns should reflect the success or failure of microbial virulence strategies and of host defenses. This type of signature is intimately connected to clinical outcome and may therefore provide prognostic value. Finally, signatures based on host response do not require the presence of the putative pathogen in a clinical specimen.

Exploring the Human Microbiome in Health and Disease

Molecular technique surveys of the microbial communities associated with humans during states of health have belatedly been initiated. These surveys are important for a number of reasons. In addition to the numerous but poorly characterized beneficial effects of the endogenous microflora on human health, a proper understanding of community membership, relative abundance, and variations therein will be critical for recognizing potential pathogens and patterns that are predictive of disease.

The subgingival crevice in the mouth is one of the more intensively studied and better-understood colonized sites of the human body. Molecular surveys that used broad-range rDNA PCR suggest that 50%–60% of the bacteria present at this site are distinct from all of those previously described at the taxonomic level of species—albeit, a term that is loosely defined [15, 16]. Some of these bacteria are not assigned to the dominant

four divisions, Actinobacteria, Firmicutes, Bacteroidetes, and Proteobacteria, and belong to divisions such as TM7 and OP11 that have not been previously discussed by clinicians or clinical microbiologists, probably because they contain no known cultivated members.

Molecular surveys of the endogenous flora of the human intestinal tract have only just begun. Although the clinical significance of the newly discovered community membership is not established, it is clear that human endogenous microflora play an important role in a variety of important disease states involving the skin and mucosal tissues. The relatively recent discovery of *Helicobacter pylori* as a common persistent human colonizer and a cause of peptic ulcer disease and gastric adenocarcinoma in certain subsets of hosts illustrates some of the complexities of host-microbe interactions at the mucosal boundary [17, 18]. What is not clear, but widely speculated about, is the possible role of the endogenous flora in either provoking or propagating disease at distant sites, including the central nervous system. The proposed associations between viral respiratory infections and subsequent flares of multiple sclerosis and between *Campylobacter jejuni* enterocolitis and Guillain-Barré syndrome [19] are just two examples.

As one broad approach to the problem of a poorly understood endogenous flora, one might consider a second human genome project [20]. Such a project would entail a comprehensive inventory of microbial genes and genomes at the 4 major sites of microbial colonization in the human body: mouth, gut, vagina, and skin. It would be approached through random shotgun sequencing procedures, targeted large-insert clone sequencing, and assessments of intra- and interindividual variation by using high-density microarrays. With increasing degrees of population sampling in well-characterized settings and with the integration of host genome-wide expression analysis [20, 21], major insights into the role of the endogenous flora in health and disease will be gained.

Microbial Signatures: Complications

Important lessons have been learned in recent years about the use and limitations of molecular methods for microbial pathogen detection and signature analysis. First, despite the expectation that sequences identified as universally conserved within a group of organisms are in fact found in all members of the group, this assumption is not always justified. As previously unrecognized members of a group are revealed, small additional degrees of sequence variation are sometimes discovered. The small subunit rDNA sequences that were originally described as universal are now known to be conserved in only a subset of cellular life [22]; revised sequence sites have taken their place. Second, PCR can exhibit bias and favor certain members of a mixed starting pool of molecules. The use of multiple broad-range primer pairs or reaction cosolvents may avoid a skewed perspective. Third, conserved sequences for use

in broad-range PCR have not yet been identified and validated for all groups of viruses. This limitation most certainly contributed to the sizeable number of cases that remained unexplained after investigation by the Unexplained Deaths and Critical Illnesses Working Group within the CDC's Emerging Infections Program [23].

One issue of particular importance concerns the complexity and widespread distribution of microbial sequence "background" or "noise" observed in the analysis of human clinical specimens (both experimental and biological). The distribution and nature of this sequence background still is not well characterized. Findings of bacterial rDNA in association with blood samples from healthy humans threatens to expand the extent of this problem into anatomic compartments that have been traditionally viewed as usually sterile [24]. A different perspective on this same apparent problem was provided in an analysis of expressed sequence tag libraries from human tissues [25, 26]. Some of these transcripts that were originally assumed to derive from the human genome appear on closer inspection to be of microbial origin. Whether some of these molecules were intrinsic to the original specimen or introduced later remains unclear, but some are easily attributed to agents that are common, persistent, or dormant infectious agents found within these human tissues.

The increasing availability of molecular pathogen discovery methods and the ease with which molecular signatures are generated create a pressing problem of a different kind. How can one build a convincing body of evidence for a causative role of the putative pathogen in a disease process when the pathogen is identified with molecular signatures and has not been isolated or purified? The issues surrounding this problem are familiar to epidemiologists and have been addressed during the past half century. One can adapt the same concepts to the kinds of data and techniques generated by modern approaches to pathogen detection and discovery [27].

Among a variety of important steps in building an argument for causation, the ability to connect a signature physically to the sites of pathology where one most expects to find the putative disease agent is one helpful evidentiary component. Fluorescent in situ hybridization allows correlation of a specific sequence with areas of pathology and tissue-based microbial structures [28, 29]. This approach also examines signature "dosage" effects. Alternatively, anatomic sites of interest can be targeted specifically for signature detection by using laser capture microdissection [30, 31].

Finally, the problems associated with clinical specimens can be substantial and potentially prohibitive as highlighted by results of the Unexplained Deaths Project [23, 32]. Clinical specimens from cases of suspected but unproven infectious etiology are often obtained late in the disease course, potentially when the putative agent is no longer present. The site from which the specimen is obtained may not coincide optimally with the expected anatomic distribution of the agent. The quantity of

specimen may be insufficient for the expected concentration of the agent and a reasonable probability of its presence (as a single particle or genome-equivalent) in the specimen. And, in the real world of clinical medicine, specimen handling and storage may introduce exogenous contamination, spurious signals, or target degradation.

Recognition and Classification of Microbial Disease Based on Host Gene Expression Patterns

The limitations of methods for analysis of microbial signatures and the emergence of technology platforms for rapid highly parallel gene expression measurements have facilitated a potentially important independent approach for identification of microbial disease. The basic question raised is, can one recognize and classify clinical (and preclinical) states of infection by examining host gene response patterns [9, 33]? This approach offers several advantages. First, changes in gene transcript abundance occur within minutes of a new exogenous stimulus. Second, the complexity and diversity of signal transduction mechanisms that impact on human gene expression and the complexity of the output (at a genome-wide level) are extensive; therefore, discrimination between numerous diverse stimuli (e.g., different classes of pathogens) may be discernable. Third, a clinical specimen need not contain the exogenous stimulus (i.e., the infectious agent). Fourth, the intrinsic nature of the host response may be directly informative about clinical relevance of the stimulus (host-microbe interaction) and the clinical outcome. However, at present, the answer to this basic question is not available.

To date the vast majority of work in this area has focused on the response of host cells to microbial stimuli in vitro (for examples, see [34, 35]). A large body of work predates this more recent focus on microbial stimuli and addresses the nature of the expression patterns associated with various forms of cancer [36–38]. These studies revealed gene expression signatures that help distinguish subsets of patients that were not previously apparent and that have different responses to treatment and different outcomes. As many might have predicted, findings from examination of host-microbe encounters in vitro indicate the predominance of shared gene expression patterns, suggesting a stereotyped temporally controlled response to microbes in human cells [34]. Gene expression responses exhibit microbial dose dependence; yet universal, shared dose-equivalence relationships are not apparent. From these early experiments, it appears that identification of discriminatory (diagnostic) signatures may be possible. Furthermore, active virulence-associated mechanisms may provide the basis for specific pathogen class recognition.

The transition to an analysis of humans with and without known infectious diseases ex vivo is accompanied by a number of interesting but complex questions. What is the most useful and practical type of clinical specimen from which to record genome-wide expression patterns and discern meaningful information about infection? Blood cells are attractive given that they circulate, make contact with a wide variety of microenvironments and other cell types, and are easy to obtain. But it is unclear how well they might reflect a localized infectious process (e.g., in the brain). How much variability occurs within and between persons during various states of health and during noninfectious stimuli? Must each person serve as his or her own control for proper interpretation of infection-associated responses? What kinds of host-specific genetic information and proclivities are embedded in expression data? These questions are currently being explored but will require extensive sampling before they can be answered in a comprehensive fashion.

One of the most intriguing questions is, on what basis do humans classify noxious stimuli and, in particular, microbial causes of disease? Among the most likely uses and practical outcomes of these investigations is the identification of patterns that predict disease outcome [36]. Furthermore, expression analysis can be used to identify predicted membrane-associated and secreted proteins [39]. With this approach, diagnostic and prognostic transcript abundance patterns can be converted to sets of easily measured proteins in body fluids.

Summary and Future Directions

A significant number of human-associated microorganisms remain unrecognized or poorly characterized. Some may be important triggers or promoting factors in unexplained disease. Molecular strategies and genomic approaches will enhance our ability to recognize these organisms, discern pathophysiologic mechanisms, and develop new interventions. Complex biologic signatures and pattern recognition are relevant not only to the use of genome-wide expression responses for classification and characterization of infectious diseases but also to the analysis of the endogenous microbial flora, secreted or exhaled volatile small molecules [40, 41], and spectral properties of human cells and tissues. Work of this type will require an intimate collaboration of experts in multiple disciplines plus shrewd clinical insight. Technology cannot substitute for a holistic understanding of biologic systems but exciting clinical investigation will be greatly accelerated by new and emerging technologies.

References

1. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 1977;74:5088–90.
2. Pace NR. A molecular view of microbial diversity and the biosphere. Science 1997;276:734–40.
3. Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 1998;180:4765–74.
4. Hugenholtz P. Exploring prokaryotic diversity in the genomic era. Genome Biol 2002;3:REVIEWS0003.
5. Relman DA. The search for unrecognized pathogens. Science 1999;284: 1308–10.

6. Relman DA. Sequence-based methods for pathogen discovery: the complex associations of microbes, microbial sequences, and host. In: Scheld WM, Craig WA, Hughes JM, eds. Emerging infections 4. Washington, DC: American Society for Microbiology Press, 2000:69–81.

7. Relman DA, Loutit JS, Schmidt TM, Falkow S, Tompkins LS. The agent of bacillary angiomatosis. An approach to the identification of uncultured pathogens. N Engl J Med 1990;323:1573–80.

8. Relman DA, Schmidt TM, MacDermott RP, Falkow S. Identification of the uncultured bacillus of Whipple's disease. N Engl J Med 1992;327:293–301.

9. Cummings CA, Relman DA. Using DNA microarrays to study host-microbe interactions. Emerg Infect Dis 2000;6:513–25.

10 Chang Y, Cesarman E, Pessin MS, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266: 1865–9.

11. Gao SJ, Moore PS. Molecular approaches to the identification of unculturable infectious agents. Emerg Infect Dis 1996;2:159–67.

12. Sutton CL, Kim J, Yamane A, et al. Identification of a novel bacterial sequence associated with Crohn's disease. Gastroenterology 2000;119: 23–31.

13. Dalwadi H, Wei B, Kronenberg M, Sutton CL, Braun J. The Crohn's disease–associated bacterial protein I2 is a novel enteric T cell superantigen. Immunity 2001;15:149–58.

14. Hemmer B, Gran B, Zhao Y, et al. Identification of candidate T-cell epitopes and molecular mimics in chronic Lyme disease. Nat Med 1999;5:1375–82.

15. Kroes I, Lepp PW, Relman DA. Bacterial diversity within the human subgingival crevice. Proc Natl Acad Sci USA 1999;96:14547–52.

16. Paster BJ, Boches SK, Galvin JL, et al. Bacterial diversity in human subgingival plaque. J Bacteriol 2001;183:3770–83.

17. Blaser MJ. Hypothesis: the changing relationships of Helicobacter pylori and humans: implications for health and disease. J Infect Dis 1999;179: 1523–30.

18. Blaser MJ, Berg DE. Helicobacter pylori genetic diversity and risk of human disease. J Clin Invest 2001;107:767–73.

19. Ang CW, Laman JD, Willison HJ, et al. Structure of Campylobacter jejuni lipopolysaccharides determines antiganglioside specificity and clinical features of Guillain-Barré and Miller Fisher patients. Infect Immun 2002; 70:1202–8.

20. Relman DA, Falkow S. The meaning and impact of the human genome sequence for microbiology. Trends Microbiol 2001;9:206–8.

21. Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI. Molecular analysis of commensal host-microbial relationships in the intestine. Science 2001;291:881–4.

22. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 1985;82:6955–9.

23. Hajjeh RA, Relman D, Cieslak PR, et al. Surveillance for unexplained deaths and critical illnesses due to possibly infectious causes, United States, 1995–1998. Emerg Infect Dis 2002;8:145–53.

24. Nikkari S, McLaughlin IJ, Bi W, Dodge DE, Relman DA. Does blood of healthy subjects contain bacterial ribosomal DNA? J Clin Microbiol 2001;39:1956–9.

25. Weber G, Shendure J, Tanenbaum DM, Church GM, Meyerson M. Identification of foreign gene sequences by transcript filtering against the human genome. Nat Genet 2002;30:141–2.

26. Relman DA. The human body as microbial observatory. Nat Genet 2002; 30:131–3.

27. Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol Rev 1996;9:18–33.

28. Fredricks DN, Relman DA. Localization of Tropheryma whippelii rRNA in tissues from patients with Whipple's disease. J Infect Dis 2001;183: 1229–37.

29. Fredricks DN, Jolley JA, Lepp PW, Kosek JC, Relman DA. Rhinosporidium seeberi: a human pathogen from a novel group of aquatic protistan parasites. Emerg Infect Dis 2000;6:273–82.

30. Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. Science 1996;274:998–1001.

31. Becich MJ. The role of the pathologist as tissue refiner and data miner: the impact of functional genomics on the modern pathology laboratory and the critical roles of pathology informatics and bioinformatics. Mol Diagn 2000;5:287–99.

32. Nikkari S, Lopez FA, Lepp PW, et al. Broad-range bacterial detection and the analysis of unexplained death and critical illness. Emerg Infect Dis 2002;8:188–94.

33. Diehn M, Alizadeh AA, Brown PO. Examining the living genome in health and disease with DNA microarrays. JAMA 2000;283:2298–9.

34. Boldrick JC, Alizadeh AA, Diehn M, et al. Stereotyped and specific gene expression programs in human innate immune responses to bacteria. Proc Natl Acad Sci USA 2002;99:972–7.

35. Nau GJ, Richmond JF, Schlesinger A, Jennings EG, Lander ES, Young RA. Human macrophage activation programs induced by bacterial pathogens. Proc Natl Acad Sci USA 2002;99:1503–8.

36. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403: 503–11.

37. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000;24:227–35.

38. Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 1999;96:9212–7.

39. Diehn M, Eisen MB, Botstein D, Brown PO. Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. Nat Genet 2000;25:58–62.

40. Haab BB, Dunham MJ, Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol 2001;2:RESEARCH0004.

41. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.