

# Biological Threat Detection via Host Gene Expression Profiling

BAOCHUAN LIN,<sup>1\*</sup> MARYANNE T. VAHEY,<sup>2</sup> DZUNG THACH,<sup>1</sup> DAVID A. STENGER,<sup>1</sup> and JOSEPH J. PANCAZIO<sup>1</sup>

**With the increased threat posed by biological weapons, detection techniques for biothreat pathogens are critically needed to monitor and assess the severity of the illness once exposure has occurred. Current approaches for detecting biological threats are either time-consuming or highly specific but provide little information regarding pathogenicity. Genotyping of pathogens by PCR provides a fast and definitive means for identifying pathogens, but reliance on pathogen genotypic endpoints has several limitations. Current progress in DNA microarrays technology provides an alternative way to address the issues faced by traditional detection systems through host gene expression profiles of peripheral blood cells. We discuss the advantages and critical issues facing the use of host gene expression profiling for biological threat detection.**

© 2003 American Association for Clinical Chemistry

The task of detecting biological threats is far more complicated than monitoring for other weapons of mass destruction. Measuring and quantifying radiation and chromatographic profiles indicative of nuclear and chemical threats, respectively, are relatively straightforward compared with the detection of biological weapon threats amid the tremendous biological background that comprises our environment. Most of the current biological warfare and environmental agent detection systems in field use or under prototype development rely on structural recognition approaches to identify anticipated agents or rely on “gold standard” culture conditions. Such assays, based primarily on antibodies, are highly specific for anticipated targets but provide only partial information concerning the pathogenicity of the threat. Current

systems the military has implemented for threat identification have performed with limited success. In fact, the Pentagon Inspector General has criticized the “Joint Biological Point Detection System” for erratic limited performance (1). In addition, it is unclear how such an approach is applicable to the civilian population, where mass deployment of point detectors is not a feasible solution.

Even with ideal detectors, proximity to a biological threat does not guarantee infection or illness. For threats such as *Bacillus anthracis*, the ability to preemptively treat individuals who were near dissemination sources with antibiotics is an option, if the dissemination source can be identified. Of further concern, a report has been published of a *B. anthracis* vaccine strain that had been engineered by Russian scientists to resist tetracycline antibiotics (2), raising the possibility that a quinolone-resistant *B. anthracis* could emerge as a biological weapon threat. Thus, not all biological threats can be treated as readily as during recent events, and there is a critical need to be able to quickly discern those individuals who are ill as a result of a bioagent etiology and to yield insight to the severity of illness.

## Current Biological Threat Detection Systems

The archaic boundaries demarking detector and diagnostic domains may have been suitable in the past when the signatures of chemical and nuclear weapons of mass destruction were readily distinguishable over the environmental background. The reality of biological threat detection is that the best (and most definitive) monitors for exposure are currently the soldiers in the field or the civilian population. Given this reality, there are several critical technology gaps in our ability to detect illness before the onset of clinical symptoms and to perform quantitative predictions of disease outcomes.

In general, genotypes are considered to provide definitive means for identification and characterization of pathogens. In fact, PCR assays such as the R.A.P.I.D. (3) and the Cepheid SmartCycler<sup>®</sup> (4) have gained increasing attention from the Department of Defense as a basis for pathogen detection. Still, reliance on pathogen genotyp-

<sup>1</sup> Center for Bio/Molecular Science & Engineering, Code 6900, Naval Research Laboratory, Washington, DC 20375.

<sup>2</sup> Division of Retrovirology, Walter Reed Army Institute of Research, 1600 East Gude Drive, Rockville, MD 20850.

\*Author for correspondence. Fax 202-767-9594; e-mail blin@cbmse.nrl.navy.mil.

Received February 20, 2003; accepted April 24, 2003.

ing faces many problems. One problem is that PCR approaches are susceptible to contamination such that bacterial DNA can be readily detected in clinical specimens from sterile sites of healthy individuals (5). A second problem is that PCR assays do not address the viability issue. The bacterial or viral DNA could be amplified even after a loss of viability to induce infection and disease. The third problem is that some pathogens induce disease via toxins such that clinical samples may not necessarily harbor genotypic information related to the causative pathogens; examples include botulinum toxin and superantigen toxins. The fourth problem is that genetically engineered biological threats could easily confound health surveillance based on genotypes. For example, modified mousepox virus with an interleukin-4 gene was shown to induce unexpected lethality in mice usually protected from mousepox via vaccination (6). In this respect, monitoring of host gene expression response may be the only useful and relevant means of detection, especially because interleukin-4 is not considered to be a virulence factor that one might consider in a genotypic analysis. The fifth problem is that bacterial populations are subject to genome diversification that involves the transmission and acquisition of exogenous genetic material, producing varying abilities to cause infection and disease (7). The sixth problem is that the kinetics for pathogen clearance and disease symptoms may not manifest concurrently. For example, invasion in the central nervous system by Venezuelan equine encephalitis virus follows clearance from the periphery in a short time span. Thus, the time window where clinical genomic markers and signs are present is very short (8, 9).

#### Use of DNA Microarrays for Gene Expression Profiles

The substantial progress in the field of molecular biology, including automated gene sequencers, industrial use of the PCR, and results from the Human Genome Project, has yielded microfabricated arrays of specific cDNA or specific oligonucleotide sequences. A single microarray contains a vast amount of DNA information. Emphasizing the specificity of this technology, Andreadis et al. (10) showed that even exposure to different  $\gamma$ -aminobutyric acid A receptor antagonists induced gene expression patterns in cultured rat cortical cells that could be differentiated with the gene microarray. With the use of these arrays, gene expression can be analyzed using thousands of probe DNA sequences where detection can be accomplished. This approach provides a very powerful tool to conduct a vast number of assays simultaneously.

The recent literature shows that gene array technology has been used to gain a clearer understanding of cancer (11–13), Huntington disease (14), and pulmonary fibrosis (15). Gene expression profiles resulting from exposure to cholera toxin (16) have been reproducibly identified in a human colon cell line in vitro. With respect to pathogens, gene arrays have been used to monitor changes in fibroblast gene expression subsequent to infection with cyto-

megalovirus (17) and *Toxoplasma gondii* (18). Gene expression changes resulting from exposure to rotavirus (19) and a laboratory strain of HIV-1 (20, 21) have been reported in a human intestinal cell line and in a CD4+-transformed cell line, respectively. More recently, Chaussabel et al. (22) reported that macrophages and monocyte-derived dendritic cells exposed to *Mycobacterium tuberculosis* and to phylogenetically distinct protozoan (*Leishmania major*, *L. donovani*, *T. gondii*) and helminth (*Brugia malay*) exhibited discrete gene expression signatures associated with each infectious agent.

For diagnostic purposes using microarrays, peripheral blood mononuclear cells (PBMCs) have been suggested as a cellular vehicle for assessing host responses. Boldrick et al. (23) recently elucidated gene expression profiles in PBMCs resulting from exposure to bacteria and bacterial toxins, showing qualitative and quantitative differences among *Bordetella pertussis*, *Escherichia coli*, and *Staphylococcus aureus*. Nau et al. (24) reported that pathogen-specific responses were also apparent in the macrophage expression profiles. In addition, gene expression patterns resulting from in vitro exposure of human PBMCs to HIV (25) and measles virus (26) have been reported. It has been suggested that with gene arrays, biological agent-induced changes in cellular gene expression can be identified, perhaps yielding a profile of biomarkers that may distinguish one agent from another or degrees of pathogenicity of classes of agents (5). The positive findings from these in vitro studies suggest that it may be possible to design an immune-cell-based biosensor relying on changes in RNA transcript concentrations or protein expression as quantitative endpoints for detection. Recent work with animal exposure is also promising, which suggests that a select set of genes expressed in PBMCs can yield potentially distinct profiles useful in distinguishing exposure to a biological threat (27).

#### Advantages and Disadvantages of Host Gene Expression Profiling as a Biothreat Detection System

It is apparent that our technical inability to rapidly detect illness and make predictions of disease outcomes may be addressed through a comprehensive examination of host gene expression profiles of PBMCs or from whole blood in response to biological threats. In this approach, health surveillance would be achieved based on how the host or individual experiences a pathogen. Although PBMCs do not necessarily constitute physiologic endpoints for pathogen interaction, these cells have the advantages of widely interacting with many of the target systems of pathogens and constituting accessible clinical samples.

Although the generation of an extensive database containing expression profiles can be accomplished, the question of how to mine these datasets becomes paramount. At this time, no clear consensus on analysis techniques has emerged for making use of host gene expression profiles. Major roles for bioinformatics include

the identification of patterns associated with responses to pathogens and the capacity to perform profile quantitative matching of novel records to a library of definitive response profiles.

Efforts in functional genomics related to cancer research have yielded major successes in the pursuit of gene expression signatures. Approaches for gene expression analysis, such as time-series analysis, pattern discovery, clustering, and class prediction, have been recently reviewed (28). Expression-based criteria or class predictors have been defined based on neighborhood analysis (29), a supervised method based on a subset of genes whose expression strongly correlates with specific classes, as well as Bayesian regression models (30) and artificial neural networks (31). These predictors were successfully used to classify novel samples in a manner consistent with clinical assessments. In fact, classifications based on gene expression alone or class discovery have also been demonstrated, suggesting that gene expression profiling has the capacity to identify subtypes that have not been previously defined (29). Although these results are promising, one should note that many of the previously conducted cancer line gene expression analyses are one-dimensional; in contrast, a host expression profile evoked by pathogen exposure would be expected to be temporal and may also exhibit dose dependence. Comprehensive sets of gene expression profiles that explore temporal and dose ranges for pathogen exposure must be produced to map the continuum of gene expression changes.

One major question concerns the ability to distinguish early gene response to a large dose of a biothreat agent vs a late response to a small dose. With well-conceived and structured studies, it is possible that valuable information can be gained by assessing gene expression profiles over two or more time points. Such derivative information may allow the differentiation of an early response to a large dose vs a late response to a smaller dose.

When using human gene expression profiles for the purpose of detection and diagnosis, a good understanding of normal variation from a diverse healthy control population is required. In fact, Whitney et al. (32) reported that gene expression profiles in the blood of healthy individuals showed high consistency in the patterns. The variation in the gene expression patterns was associated with the difference in the cellular composition of the blood sample and with gender, age, and time of day. The ability to recognize systematic variation in human blood provides the feasibility of using gene expression profiles in peripheral blood as a basis for biothreat detection. Furthermore, the variances among gene arrays also need to be taken into consideration. Using the same human-derived sample, our group has examined the variation in signals among gene arrays (33). For both radioactively labeled membrane arrays and, more recently, single-color high-density microarrays, the mean CV for gene signals was ~25%. Although this seems noisy, the strength of the approach is based on the large

numbers of genes used to perform classification. The chief source of variation does not appear to be differences among chips, but rather from the operators and reagent activity. In fact, it has been known among gene chip laboratories that clustering techniques applied to consistent types of samples from different laboratories can identify operators/laboratories. For this reason, automation of sample processing is critical for gaining data uniformity.

One attractive alternative is to examine technologic advances in pattern recognition from other fields where the processing and interpretation of hundreds of thousands of multiparametric data records have been successfully implemented. These areas include face recognition (34, 35) and radar signal discrimination (36).

With a comprehensive database, it will be interesting to assess the degree of differentiation by gene expression analysis for particular pathogens. In fact, recent work by Chaussabel et al. (22) showed unique gene expression profiles associated with phylogenetically distinct pathogens in human macrophages and dendritic cells. Nevertheless, one should remain cautious. We speculate that gene expression profiles will impact the classification of pathogens but that specific assays such as those that provide pathogen genomic sequence information will provide a definite basis for differentiation.

Although use of host gene expression profiling as a biological threat detection system may provide more definite information regarding pathogens, the major limitations facing the use of host gene expression profiling as a diagnostic tool involve sample collection and the time required to process a microarray assay. Indeed, with the emergence of clinical blood collection tubes that allow for the preservation of RNA samples derived from blood (37, 38), the systematic acquisition of meaningful gene expression values can be achieved. In practice, the blood sample required to provide sufficient RNA analysis can be achieved with as little as 2.5–10 mL of whole blood based on our experience and recent work by Whitney et al. (32). Still, most gene microarray hybridization steps require hours rather than minutes, an undesirable and perhaps intolerable delay for evaluation of exposures to potentially lethal biological agents. It is necessary that once libraries of gene expression profiles have been generated with use of whole-genome microarrays, subsets of meaningful genes would be used as alternative means of detection. Alternatives include bead array technologies (39, 40) and microelectronic array devices (41, 42), which yield information in seconds or minutes rather than hours or days, that allow for rapid estimation of gene expression. For example, by applying a positive charge at electrode contacts, microelectronic arrays make use of the fact that DNA molecules have a net negative charge to concentrate the molecule to the desired location and accelerate hybridization, which is considered to be the rate-limiting step.

In conclusion, to achieve the vision of accomplishing predictions of health status based on host gene expression, several goals must be realized. It is important to understand that gene expression experiments conventionally performed in academic laboratories represent snapshots of a limited set of uniformly sampled time points after exposure and particular "doses" of pathogen. Comprehensive sets of gene expression profiles that explore temporal and dose ranges for pathogen exposure must be produced to map the continuum of gene expression changes. Highly coordinated research activity across many laboratories will be required to generate these profiles with standard operating protocols, which should include a common platform that has the capacity to eventually become an approved diagnostic tool. The collected gene expression data profiles could then form the basis of a massive library that could be mined with tools capable of quantitatively seeking similarity with novel data sets. These tools could be implemented for use on web-accessible, secure interfaces that can expand to accommodate tens of thousands of data records.

### References

1. Early-warning bio-attack systems remain unreliable. Taipei Times <http://www.taipetimes.com/news/2001/10/31/story/0000109588> (Accessed February 2003).
2. Stepanov AV, Marinin LI, Pomerantsev AP, Staritsin NA. Development of novel vaccines against anthrax in man. *J Biotechnol* 1996;44:155–60.
3. Idaho Technology, Inc. R.A.P.I.D. Ruggedized Advanced Pathogen Identification System. <http://www.idahotech.com/rapid/> (Accessed February 2003).
4. Cepheid. Smart Cycler® II System. <http://www.smartcycler.com/> (Accessed February 2003).
5. Relman DA. The search for unrecognized pathogens. *Science* 1999;284:1308–10.
6. Jackson RJ, Ramsay AJ, Christensen CD, Beaton S, Hall DF, Ramshaw IA. Expression of mouse interleukin-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox. *J Virol* 2001;75:1205–10.
7. Joyce EA, Chan K, Salama NR, Falkow S. Redefining bacterial populations: a post-genomic reformation. *Nat Rev Genet* 2002;3:462–73.
8. Grieder FB, Davis NL, Aronson JF, Charles PC, Sellon DC, Suzuki K, et al. Specific restrictions in the progression of Venezuelan equine encephalitis virus-induced disease resulting from single amino-acid changes in the glycoproteins. *Virology* 1995;206:994–1006.
9. Charles PC, Trgovcich J, Davis NL, Johnston RE. Immunopathogenesis and immune modulation of Venezuelan equine encephalitis virus-induced disease in the mouse. *Virology* 2001;284:190–202.
10. Andreadis JD, Mann TT, Russell AC, Stenger DA, Pancrazio JJ. Identification of differential gene expression profiles in rat cortical cells exposed to the neuroactive agents trimethylolpropane phosphate and bicuculline. *Biosens Bioelectron* 2001;16:593–601.
11. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1999;96:6745–50.
12. Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 2000;406:532–5.
13. Iacobuzio-Donahue CA, Argani P, Hempfen PM, Jones J, Kern SE. The desmoplastic response to infiltrating breast carcinoma: gene expression at the site of primary invasion and implications for comparisons between tumor types. *Cancer Res* 2002;62:5351–7.
14. Luthi-Carter R, Strand A, Peters NL, Solano SM, Hollingsworth ZR, Menon AS, et al. Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Hum Mol Genet* 2000;9:1259–71.
15. Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJ, Morris D, et al. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci U S A* 2000;97:1778–83.
16. Lin HJ, Charles PT, Andreadis JD, Churilla AM, Stenger DA, Pancrazio JJ. Cholera toxin-induced modulation of gene expression: elucidation via cDNA microarray for rational cell-based sensor design. *Anal Chim Acta* 2002;457:97–108.
17. Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1998;95:14470–5.
18. Blader IJ, Manger ID, Boothroyd JC. Microarray analysis reveals previously unknown changes in *Toxoplasma gondii*-infected human cells. *J Biol Chem* 2001;276:24223–31.
19. Cuadras MA, Feigelstock DA, An S, Greenberg HB. Gene expression pattern in Caco-2 cells following rotavirus infection. *J Virol* 2002;76:4467–82.
20. Geiss GK, Bumgarner RE, An MC, Agy MB, van't Wout AB, Hammersmark E, et al. Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. *Virology* 2000;266:8–16.
21. Corbeil J, Sheeter D, Genini D, Rought S, Leoni L, Du P, et al. Temporal gene regulation during HIV-1 infection of human CD4+ T cells. *Genome Res* 2001;11:1198–204.
22. Chaussabel D, Tolouei Semnani R, McDowell MA, Sacks D, Sher A, Nutman TB. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood* 2003;Mar 27 (epub ahead of print).
23. Boldrick JC, Alizadeh AA, Diehn M, Dudoit S, Liu CL, Belcher CE, et al. Stereotypes and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci U S A* 2002;99:972–7.
24. Nau GJ, Richmond JF, Schlesinger A, Jennings EG, Lander ES, Young RA. Human macrophage activation programs induced by bacterial pathogens. *Proc Natl Acad Sci U S A* 2002;99:1503–8.
25. Vahey MT, Nau ME, Jagodzinski LL, Yalley-Ogunro J, Taubman M, Michael NL, et al. Impact of viral infection on the gene expression profiles of proliferating normal human peripheral blood mononuclear cells infected with HIV-1. *AIDS Res Hum Retroviruses* 2002;18:179–92.
26. Bolt G, Berg K, Blixenkrone-Moller M. Measles virus-induced modulation of host-cell gene expression. *J Gen Virol* 2002;83(Pt 5):1157–65.
27. Bi SG, Das R, Mani S, Van Gessel Y, Neill R, Jett M. Correlation of the kinetics of global gene expression patterns with progression of lethal shock in response to staphylococcal enterotoxin A or B in piglets [Abstract]. *FASEB J* 2002;16:A162.
28. Solnink DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002;32(Suppl):502–8.
29. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.

30. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A* 2001; 98:11462–7.
31. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001; 7:673–9.
32. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression pattern in human blood. *Proc Natl Acad Sci U S A* 2002;100:1896–901.
33. Dodson JM, Charles PT, Stenger DA, Pancrazio JJ. Quantitative assessment of filter-based cDNA microarrays: gene expression profiles of human T-lymphoma cell lines. *Bioinformatics* 2002;18: 1–8.
34. Gutta S, Wechsler K. Face recognition using hybrid classifiers. *Pattern Recognit* 1997;30:539–53.
35. Khuwaja GA. An adaptive combined classifier system for invariant face recognition. *Digit Signal Process* 2002;12:21–46.
36. Sciortino JC. Autonomous ESM systems. *Naval Engineers Journal* 1997;109:73–84.
37. Qiagen, PAXgene blood RNA system (PreAnalytiX). [http://www.qiagen.com/catalog/auto/cget.asp?p=PAXgene\\_blood\\_RNA\\_system](http://www.qiagen.com/catalog/auto/cget.asp?p=PAXgene_blood_RNA_system) (Accessed February 2003).
38. Becton, Dickinson and Company. Vacutainer™ CPT™ tube. <http://www.bd.com/vacutainer/products/molecular/> (Accessed February 2003).
39. Ferguson JA, Boles TC, Adams CP, Walt DR. A fiber-optic DNA biosensor microarray for the analysis of gene expression. *Nat Biotechnol* 1996;14:1681–4.
40. Yang L, Tran DK, Wang X. BADGE, beads array for the detection of gene expression, a high-throughput diagnostic bioassay. *Genome Res* 2001;11:1888–98.
41. Cheng J, Sheldon EL, Wu L, Uribe A, Gerrue LO, Carrino J, et al. Preparation and hybridization analysis of DNA/RNA from *E. coli* on microfabricated bioelectronic chips. *Nat Biotechnol* 1998;16: 541–6.
42. Gurtner C, Tu E, Jamshidi N, Haigis RW, Onofrey TJ, Edman CF, et al. Microelectronic array devices and techniques for electric field enhanced DNA hybridization in low-conductance buffers. *Electrophoresis* 2002;23:1543–50.