2006

Early days: genomics and human responses to infection

Minghsun Liu
Stanford University School of Medicine, lmh@ucla.edu

Stephen J. Popper
Stanford University School of Medicine

Kathleen H. Rubins
Stanford University School of Medicine, rubins@wi.mit.edu

David A. Relman
Stanford University School of Medicine, relman@stanford.edu

ELSEVIER

Early days: genomics and human responses to infection

Minghsun Liu[1], Stephen J Popper[2], Kathleen H Rubins[2] and David A Relman[1,2,3]

DNA microarray-based gene transcript-profiling of the responses of primates to infection has begun to yield new insights into host–pathogen interactions; this approach, however, remains plagued by challenges and complexities that have yet to be adequately addressed. The rapidly changing nature over time of acute infectious diseases in a host, and the genetic diversity of microbial pathogens present unique problems for the design and interpretation of functional-genomic studies in this field. In addition, there are the more common problems related to heterogeneity within clinical samples, the complex, non-standardized confounding variables associated with human subjects and the complexities posed by the analysis and validation of highly parallel data. Whereas various approaches have been developed to address each of these issues, there are significant limitations that remain to be overcome. The resolution of these problems should lead to a better understanding of the dialogue between the host and pathogen.

Addresses
[1] Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University School of Medicine, 300 Pasteur Drive, Grant S-169, Stanford, CA 94305, USA
[2] Department of Microbiology and Immunology, Stanford University School of Medicine, 279 Campus Drive, Beckman B403, Stanford, CA 94305, USA
[3] Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304, USA

Corresponding author: Relman, David A (relman@stanford.edu)

Current Opinion in Microbiology 2006, **9**:312–319

This review comes from a themed issue on
Techniques
Edited by Barry L Wanner and Guido Grandi

Available online 6th May 2006

1369-5274/$ – see front matter

DOI 10.1016/j.mib.2006.04.006

Introduction

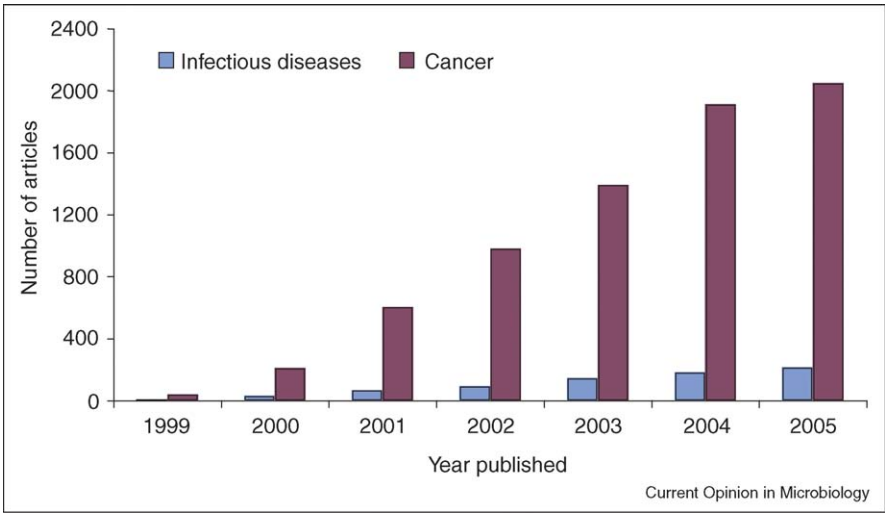
Since the last review in this journal of DNA microarray-based transcriptome analysis for studying host responses to infection [1], the adoption of this approach by research groups around the world has continued unabated. This is not surprising, given the rewards that have so far been reaped from this genomic technology, especially in understanding cancer and in the development of new practical tools for classifying cancer patients on the basis of disease outcome or predicted treatment-response. Yet, as shown in Figure 1, the number of published papers on DNA microarray-based host response profiling in the setting of infectious diseases still lags significantly behind the number of papers on the use of this approach in the study of cancer. There are many probable reasons for this discrepancy.

In this review, we focus on studies published from 2004 to present that rely on the use of DNA microarrays to study the host-response to pathogens using human clinical samples or non-human primate models. The use of clinical samples allows the study of host–pathogen interactions in a clinically and physiologically relevant setting by exploring the relationships between patterns of human transcript abundance, human physiological parameters, the natural history of the disease process, and pathogen characteristics. The resulting complexity presents challenges beyond those posed by *in vitro* experimental systems or small animal models. In this review, we address some of the methodological issues raised by this approach in this setting. An overview of some of the factors to consider in designing DNA microarray-based studies of infectious diseases is provided in Figure 2. Readers interested in a broad overview of DNA microarray technology and microbial pathogenesis are referred to two other recent reviews [2•,3].

Where to 'listen': target tissue and cell type

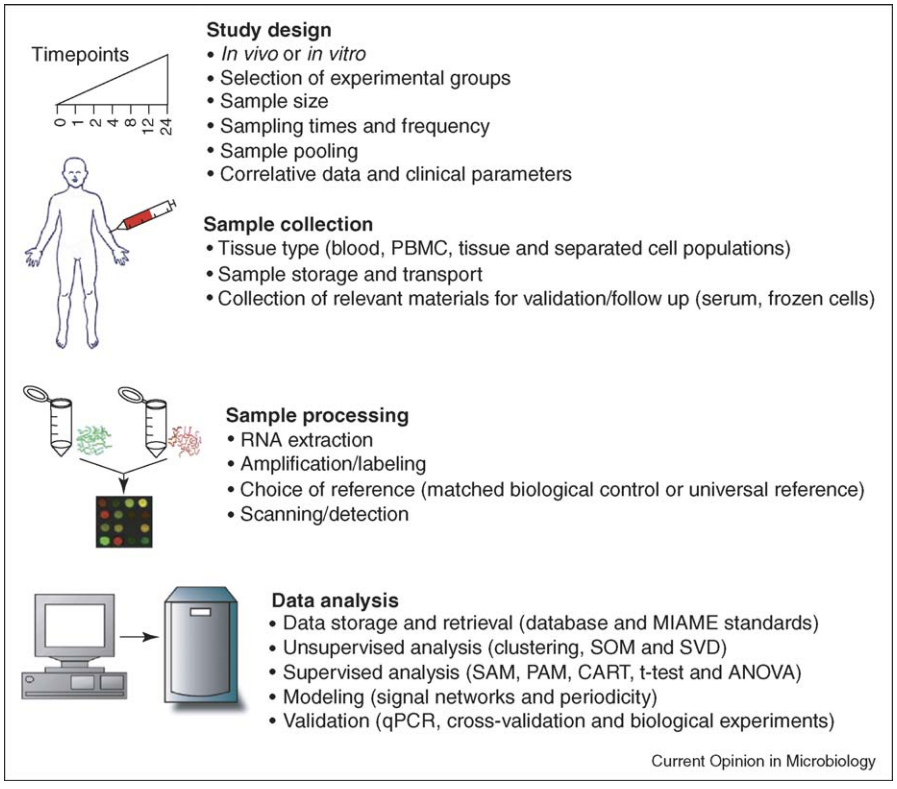
In the study of naturally-occurring human disease the restricted availability of tissues and cells of various types often dictates the kinds of studies that are feasible. As reflected in the current literature, two different approaches dominate in the selection of human clinical samples. The first approach focuses on peripheral blood: peripheral blood is a relatively accessible source of human cells and RNA, and is a natural choice for analyzing host transcript-based responses in systemic infection. This approach has been used in the study of HIV infection [4,5], sepsis [6], severe acute respiratory syndrome (SARS) [7,8], malaria [9•] and febrile illness as a result of upper respiratory tract infections [10]. Most often, RNA is harvested from either peripheral blood mononuclear cells (PBMC) or from whole blood; whole blood analysis involves less manipulation of the specimen than do most PBMC collection methods [11] and for this reason might be particularly suitable for clinical field studies. However, whole blood (which includes erythrocytes, granulocytes and platelets in addition to the PBMC) is obviously a more complex mixture of cell types, creating additional confounding variables, such as the relative abundance levels of these different cell types (see below). In addition, the large

Figure 1



Number of published articles on gene expression profiling per year, by topic. The following search strategy was used in PubMed (http:/www.pubmed.com) to identify articles focused on host-pathogen interactions: (''Gene Expression Profiling'' [MeSH] OR ''Oligonucleotide Array Sequence Analysis'' [MESH]) AND (''Parasitic Diseases'' [MeSH] OR ''Bacterial Infections and Mycoses'' [MeSH] OR ''Virus Diseases'' [MeSH]). To identify similar articles discussing cancer-related gene expression patterns, the pathogen MeSH terms were replaced with ''Neoplasms'' [MeSH]. Abbreviation: MeSH, medical subject heading.

Figure 2



Factors to consider in microarray-based studies of the host response to infection. An overview of variables and points to consider when designing and performing microarray experiments. Abbreviations: ANOVA, analysis of variance; CART, classification and regression tree; MIAME, minimum information about a microarray experiment; PAM, predictive analysis of microarrays; qPCR, quantitative PCR; SAM, significance analysis of microarrays; SOM, self organizing map; SVD, singular value decomposition.

amount of haemoglobin gene transcripts present in RNA preparations from whole blood samples might interfere with microarray signals on some platforms (http://www.affymetrix.com/support/technical/technotes/blood2_technote.pdf), although we have not found any evidence of this problem using cDNA arrays in our laboratory (J Yen, unpublished data).

The second approach has been to examine tissues or cells that are prominently involved in infectious disease-associated pathology: synovial fluid cells in septic knees [12], explanted hearts in cardiomyopathy as a result of Chagas disease [13], liver biopsies in hepatitis C-related fibrosis, or in response to interferon therapy for this disease [14–18], and gastric biopsies in *Helicobacter pylori* infection [19,20]. In addition, several studies have examined HIV-related dementia and encephalitis using brain tissues obtained during autopsy [21–23]. The degree to which host responses to infection are compartmentalized and distinct between different compartments remains poorly described. For example, the correlation between transcript abundance patterns in the peripheral blood and those within local cells of various types at the site of a local infection is unclear. It is not known how well peripheral blood cells might discriminate between local infections caused by different agents at different sites, or how well they might reveal basic mechanisms of local immune response. Clearly, the boundaries between local and systemic compartments are not always distinct, because there are molecular markers in the systemic circulation that can be used as indicators of local infection. For example, serum levels of creatine kinase-MB isoenzyme and cardiac troponin I can be used to monitor myocardial involvement in viral myocarditis; and serum hepatic transaminase levels are used to measure the extent of hepatocellular injury in viral hepatitis. Furthermore, microbial disease agents might traffic through the peripheral blood to a greater degree than previously suspected (e.g. gastroenteritis-associated rotavirus).

Human clinical samples other than blood, and cells and tissue from the primary site of infection associated pathology might be useful for DNA microarray-based analysis of infectious diseases. It remains to be seen how robust and informative the RNA-based signatures are from saliva [24] and other body fluids in this context; despite the obvious challenges, the ease of access to some of these sample types justifies further investigation. Various body fluids have already been pursued as sources of protein-based signatures of infectious disease.

The challenges of minimizing biases and controlling confounding variables

Transcript abundance analysis of the host-response to infection in model systems provides the opportunity to control for, or eliminate, many of the variables present during infection, including differences in environmental factors, host genetic background, pathogen strain, infectious dose, route and timing. Controlled studies of global gene expression patterns in human subjects following infection have not yet been published, although several studies have reported on the consequences of administering microbial components [25•,26]. Instead, all of the microarray-based studies cited above were observational studies. As with any observational study, findings take the form of associations between the measured variable (gene transcript abundance) and other characteristics of the process being studied (e.g. associated microbial agent, survival and response to treatment); these findings are subject to potential bias and confounding effects. An unidentified or unspecified factor, or characteristic, might lead to an incorrect interpretation — an unwarranted inference of causality — or might limit the ability to infer generalizations from the findings. This issue is prominent in the typical microarray experiment, where thousands, or tens of thousands of genes, reflecting diverse physiological processes are examined all at once.

The potential for selection bias can be seen in one report aimed at identifying genes whose expression levels might be used as pretreatment predictors of responsiveness to interferon therapy in patients with chronic hepatitis C [17]. Patients infected with genotype 1 hepatitis C virus have a significantly lower rate of response to interferon therapy [27]. Not surprisingly, in these studies, patients with genotype 1 infection were relatively over-represented in non-responder groups. Chen *et al.* [17] addressed the possibility that the non-response transcriptional profile was simply a genotype 1 profile, by determining that the same set of genes distinguished responders and non-responders among those with genotype 1 infections. However, they did not evaluate their findings in an independent set of samples, and because all non-responders had genotype 1 virus, it is possible that the identified genes only are associated with treatment response in genotype 1 infection.

The collection of comprehensive data for a variety of potentially-relevant clinical parameters provides an important means of evaluating the role of potential confounders, and should be carefully considered during study design. Studies in rats have shown that the expression of some liver genes is influenced by fasting and circadian rhythms [28]. However, none of the reports evaluating liver gene expression in hepatitis C indicate whether diet and time of day were recorded during sample collection, or even considered in the data analysis [14–18,29]. The issue is not unique to hepatic gene expression; transcript abundance profiles in blood are also known to vary according to gender, blood cell subset abundance, time of day and age [11,30]. One approach, used in a study of *H. pylori*-associated patterns of expression in gastric tissue [19], is to match cases and controls on the basis of potential confounding variables — in this study, samples

were matched for age and gender. A second approach is to eliminate from further analysis genes known to be significantly associated with certain clinical parameters [31•]. Whereas both of these approaches if executed carefully reduce the possibility of arriving at spurious conclusions about the relationship between gene expression and an outcome of interest, they also limit the set of possible subjects from which to draw samples for analysis, and limit the possibility of discovering novel, interesting associations between gene transcript abundance and the matching variables. A third approach is to evaluate the role of the clinical parameters during analysis. Chen *et al.* [17] used multivariate analysis to evaluate the association of specific genes with clinical parameters, and Griffiths *et al.* [9•] evaluated the strength of association of gene expression patterns with each clinical parameter, and identified gene subsets that were associated with each variable. This latter study also demonstrated that certain clinical parameters can help highlight for further study specific aspects of the host transcript-based response to infection, and help to elucidate biological mechanisms.

Sample size, pooling, replication, and frequency

Determining the minimum and/or optimal number of samples (the sample size) for microarray-based studies remains an active research area (reviewed in [32•]). None of the studies reviewed for this paper made sample-size determination explicit. At this time, sample size remains largely determined by resource and sample availability. Yet, it is disconcerting to find published microarray-based *in vivo* studies that include only one sample for each experimental group [33,34]. In addition to the obvious benefit of achieving statistical significance, sample size estimations facilitate collection of samples in sufficient numbers so that a 'tester set' is available for cross-validation (see further discussion below).

Allison *et al.* [32•] suggested that pooling of biological samples can reduce variability but obscure the ability to compare variables across individual samples. Pooling of human clinical samples masks the clinical data from each individual sample. The subsequent ability to correlate gene expression patterns with specific clinical parameters is then lost. Once a technical platform is proven reliable in a given laboratory, biological replicates provide much more information and value than technical replicates (i.e. the same biological sample processed and hybridized more than once).

Determining appropriate sampling frequency and time-points is another important aspect of experimental design. Unlike oncological or rheumatological diseases, where the process evolves over months and years, the host response during acute infectious disease is a much more dynamic process, and can produce gene expression changes over time scales ranging from minutes to hours. Calvano *et al.* [25•] showed that transcript responses to one dose of intravenous endotoxin occur by the second hour and resolve within 24 hours. In a previous report from our lab, Rubins *et al.* [35••] sampled blood from monkeys with smallpox infection over multiple days and found that an interferon-associated transcript abundance response had begun by the first day or two after infection, increased during the first four days and then plateaued around day six. Other data from this study showed that transcript abundance changes can precede the development of clinical signs [35••]. Taken together, these studies suggest that dense sampling over the time course of infection, beginning prior to the appearance of clinical signs, might be crucial for a complete picture of the gene expression program during acute infectious disease.

Heterogeneity of cell populations in samples

Most human tissues and samples collected for microarray analysis are a mixture of distinct cell types. A mixed cell population, such as whole blood or PBMC, gives a comprehensive picture of gene expression during the systemic host-response to infection, and reflects the overall interactions within a complex system. However, the complex dynamics of a heterogeneous sample poses two different challenges for *in vivo* experiments. An important gene expression pattern might be undetectable amidst a 'noisy' environment, particularly a pattern generated by a rare cell population. In addition, increases or decreases in the relative abundance of a cell type alters the overall proportion of unique transcripts from that cell type in the total pool of RNA from a given sample. The resulting gene transcript data might simply be a reflection of the sum of unique, constitutive transcript profiles from each cell type, rather than an actual increase or decrease in transcript levels within a given cell type in response to the infection or stimulus.

Two studies have examined the different contributions of individual cell types in a complex *in vivo* mixed population. Mueller *et al.* [36••] used laser micro-dissection to harvest the three major epithelial cell types from murine stomach tissue and showed that a *H. pylori*-specific transcriptional profile is induced only in the mucus-producing pit cells. McLaren *et al.* [34] compared gene expression differences between antigen-stimulated PBMC and $CD4^+$ and $CD8^+$ T-lymphocyte subpopulations. Their limited analysis suggested that the transcript patterns of lymphocyte subpopulations are distinct and differ from the PBMC population as a whole.

Some *in vitro* experiments have explored gene transcript patterns of specific cell subpopulations, sometimes in response to specific stimuli [30,37] (Waddell *et al.*, manuscript in preparation). These 'reference studies' have produced gene lists that can potentially be used as signatures for determining the contributions of different cell subsets to the profile of a mixed population without

having to separate each subset *ex vivo*. Several *in vivo* studies have recognized the contribution of cell population dynamics to gene expression patterns and have performed correlation analysis to identify cell subset abundance effects [9•,11,35••]. However, the utility of these secondary data analysis techniques, for example, correlation and signature gene lists, for identifying infection responses in a complex cell mixture remains to be determined.

Whether cell sorting and separation is necessary in a study will depend upon the specific experimental aims. For studies with a focus on diagnosis, prediction or classification, it might be acceptable to include patterns reflective of both fluctuating cell populations and gene regulation. However, for studies that are designed to elucidate pathogenesis or mechanism, the difficulties in sorting out contributions from each of these factors could prove too problematic. At a minimum, studies that sample complex tissues and mixtures of cells should collect data on the abundance of various cell types during the timeframe of the disease process. Reduction of sample complexity through purification of individual cell populations might be a necessary and important approach in the future.

Reporting and availability of microarray data

One of the most powerful and as yet, under-utilized approaches to microarray data analysis is to combine primary gene transcript abundance data from different studies in a 'meta-analysis'. Jenner and Young [2•] employed this approach with a re-analysis of data from 32 studies of host–pathogen interactions. This study demonstrated both a common transcriptional response to pathogens across cell types and studies, as well as specific responses involving Toll-like receptors and subsets of pathogens. However, the usefulness of this approach is limited by the quality and availability of the underlying data. Many transcript profiling papers simply list ''induced'' or ''repressed'' genes, without making available the underlying raw data, or even the values of the clustered and filtered data upon which the lists were based. This type of data reporting structure does not allow for rigorous peer-review of the data, and prohibits future types of combinatorial analysis. In addition, many investigators simply create rank lists of fold changes in transcript abundance as proof for differential expression and proceed to make inferential statements and conclusions based on the highly ranked genes. This approach assumes that the variance in transcript levels for every gene is the same — this is not true in most experiments, leading to unfounded conclusions [32•].

In 2001, a set of standards was proposed for the review and publication of microarray data [38] (http://www.mged.org/Workgroups/MIAME/miame.html). Submission of microarray data to a public database such as ArrayExpress [39], Gene Expression Omnibus (GEO) [40], or the Center for Information Biology Gene Expression Database (CIBEX) [41], should be a basic requirement by all journals for publication of manuscripts that present these data, in the same way that submission of gene or protein sequence data to public databases is currently required for publication. The latter requirement has proved invaluable for researchers worldwide. The wealth of information generated with microarray experiments can only be fully exploited if the data are made accessible to the public.

Validation

'Validation' is a term that has been used loosely in the gene expression literature. Generally, validation of results from microarray profiling of host response refers to the corroboration of changes in gene transcript levels using an independent experimental method, such as quantitative real-time PCR (RT-PCR). There is no current consensus on the suitability or need for RT-PCR as an appropriate and/or indicated method for confirmation of microarray data [32•]. Need might be, in part, based on the importance of obtaining precise quantitative data for specific transcripts in a particular study. At the same time, when a group of co-varying transcripts is found, all of which are known to be associated with the same biological system or signaling pathway, this is strong evidence that this system or pathway is affected by the process under study; independent measurements of specific transcripts might not be necessary. Some recent microarray designs incorporate extensive internal controls, further reducing the need for RT-PCR: one example is the HEEBO (human exonic evidence-based oligonucleotide) and MEEBO (mouse exonic evidence-based oligonucleotide) arrays designed by Alizadeh *et al.* (http://alizadehlab.stanford.edu). As an alternative, measurement of predicted protein products might be more appropriate in some settings [19]. Rubins *et al.* [35••] pursued leads generated by microarray experiments with measurements of interferon protein — following up on the increased abundance of known interferon-regulated gene transcripts, and with detection of apoptotic T-cells by TUNEL (terminal deoxynucleotidyl transferase biotin–dUTP nick end labelling) assays — following up on the decreased abundance of certain T-lymphocyte-associated transcripts. Obviously, validation of biological leads using *in vitro* models of host–pathogen interaction remains far easier than validation of such findings in humans [42•,43].

One of the major advantages of an agnostic method like genome-wide transcript profiling is that it can highlight potentially important genes without being biased or limited by prior assumptions about an experimental system. Researchers tend to ''only look where the streetlight is'', in focusing on genes with familiar names. Pre-existing reagents and tools encourage this behavior, but the increasing availability of comprehensive, genome-wide

gene 'knock-down' tools [44] should significantly improve the ability to study unknown genes identified in microarray experiments.

Appropriate cross-validation should be performed in studies that generate classifier or predictor sets of genes based on microarray data; Tibshirani *et al.* [45] provides a brief discussion on the challenges of classification in this setting. In several studies, sets of differentially expressed genes were used to assign samples or cases to clinical categories [6,9•,10,14,17]. However, in at least one study, no effort was made to deal with generalization error [6], thus severely weakening the strength of the findings.

Conclusion: forming the big picture

The promise that genomics would improve our understanding of host–pathogen interactions has yet to be fully realized. Yet there are clear indications that progress is accelerating and new insights are accruing. The popularity of microarray-based technologies for examining these interactions will continue to expand as costs diminish, standards are propagated and techniques improve. The degree to which the associated explosion in the quantity of available data is accompanied by comparable gains in biological insight will depend on thoughtful study design, careful data analysis, appropriate follow-up experiments, and public availability of well-annotated, raw, standardized microarray data.

Currently, it is also clear that we are not extracting all available higher-order information from these data. Several gene class annotation tools have been developed to identify and organize biological themes among differentially expressed genes [46,47]. Bioinformatics tools, such as those employed by Calvano *et al.* [25•] and Koller's group [48,49•], will further enhance the interpretation of microarray data. In particular, Koller *et al.* have developed creative computational modeling approaches for deducing biological processes and regulatory networks, based on integration of multiple kinds of highly-parallel genome-wide data, such as primary sequence, gene transcript abundance and protein–DNA association data [48].

From the genomic perspective, host–pathogen interactions involve the interplay of two (or more) distinct genomes. A more comprehensive study of these interactions will entail simultaneous monitoring of all of the associated genes and gene products [50,51]. We believe that one of the most exciting prospects for this field lies in the integration of genome-wide data from multiple sources. Large-scale random shotgun sequencing efforts using samples from complex microbial communities provide an opportunity to discover and monitor the expression of genes directly from samples of these communities, whether they are host-associated or environment-associated. Community-wide patterns of gene transcript abundance and their variation with respect to time, space and perturbation will reveal fundamentally important aspects of microbial community ecology and physiology, and provide potentially useful diagnostic and prognostic information about the hosts in which communities play critical roles in health and disease. As we construct a high resolution and dynamic picture of the conversation between humans and microbes, we will not only enhance our understanding about fundamental aspects of human biology, but improve the clinical management of infectious diseases.

Acknowledgements

We would like to thank Judy Yen (Stanford University) for providing helpful comments and outstanding contributions to the experimental work of our group. We also thank Pat O Brown (Stanford University) for his immense insight, valuable guidance, and unflagging dedication to a better understanding of human health and disease. Our work is supported by grants from the National Institutes of Health (AI54922, AI51259, AI57188), Wellcome Trust (059141/Z/99/B), Ellison Medical Foundation (ID-SS-0103-01) and the Horn Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Walduck A, Rudel T, Meyer TF: **Proteomic and gene profiling approaches to study host responses to bacterial infection**. *Curr Opin Microbiol* 2004, **7**:33-38.

2. • Jenner RG, Young RA: **Insights into host responses against pathogens from transcriptional profiling**. *Nat Rev Microbiol* 2005, **3**:281-294.
This review provides a meta-analysis of many previous data sets, and reveals the difficulty and challenges at the present time with this potentially valuable approach.

3. Coombes BK, Hardwidge PR, Finlay BB: **Interpreting the host–pathogen dialogue through microarrays**. *Adv Appl Microbiol* 2004, **54**:291-331.

4. Motomura K, Toyoda N, Oishi K, Sato H, Nagai S, Hashimoto S, Tugume SB, Enzama R, Mugewa R, Mutuluuza CK *et al.*: **Identification of a host gene subset related to disease prognosis of HIV-1 infected individuals**. *Int Immunopharmacol* 2004, **4**:1829-1836.

5. Pulliam L, Sun B, Rempel H: **Invasive chronic inflammatory monocyte phenotype in subjects with high HIV-1 viral load**. *J Neuroimmunol* 2004, **157**:93-98.

6. Prucha M, Ruryk A, Boriss H, Moller E, Zazula R, Herold I, Claus RA, Reinhart KA, Deigner P, Russwurm S: **Expression profiling: toward an application in sepsis diagnostics**. *Shock* 2004, **22**:29-33.

7. Lee YS, Chen CH, Chao A, Chen ES, Wei ML, Chen LK, Yang KD, Lin MC, Wang YH, Liu JW *et al.*: **Molecular signature of clinical severity in patients infected with severe acute respiratory syndrome coronavirus (SARS-CoV)**. *BMC Genomics* 2005, **6**:132.

8. Yu SY, Hu YW, Liu XY, Xiong W, Zhou ZT, Yuan ZH: **Gene expression profiles in peripheral blood mononuclear cells of SARS patients**. *World J Gastroenterol* 2005, **11**:5037-5043.

9. • Griffiths MJ, Shafi MJ, Popper SJ, Hemingway CA, Kortok MM, Wathen A, Rockett KA, Mott R, Levin M, Newton CR *et al.*: **Genomewide analysis of the host response to malaria in Kenyan children**. *J Infect Dis* 2005, **191**:1599-1611.
This study explores the complex clinical parameters associated with patterns of gene transcript abundance.

10. Thach DC, Agan BK, Olsen C, Diao J, Lin B, Gomez J, Jesse M, Jenkins M, Rowley R, Hanson E *et al.*: **Surveillance of**

transcriptomes in basic military trainees with normal, febrile respiratory illness, and convalescent phenotypes. *Genes Immun* 2005, **6**:588-595.

11. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO: **Individuality and variation in gene expression patterns in human blood**. *Proc Natl Acad Sci USA* 2003, **100**:1896-1901.

12. Deirmengian C, Lonner JH, Booth RE Jr: **The Mark Coventry Award: white blood cell gene expression: a new approach toward the study and diagnosis of infection**. *Clin Orthop Relat Res* 2005, **440**:38-44.

13. Cunha-Neto E, Dzau VJ, Allen PD, Stamatiou D, Benvenutti L, Higuchi ML, Koyama NS, Silva JS, Kalil J, Liew CC: **Cardiac gene expression profiling provides evidence for cytokinopathy as a molecular mechanism in Chagas' disease cardiomyopathy**. *Am J Pathol* 2005, **167**:305-313.

14. Hayashida K, Daiba A, Sakai A, Tanaka T, Kaji K, Inaba N, Ando S, Kajiyama N, Terasaki H, Abe A *et al.*: **Pretreatment prediction of interferon-alfa efficacy in chronic hepatitis C patients**. *Clin Gastroenterol Hepatol* 2005, **3**:1253-1259.

15. Smith MW, Walters KA, Korth MJ, Fitzgibbon M, Proll S, Thompson JC, Yeh MM, Shuhart MC, Furlong JC, Cox PP *et al.*: **Gene expression patterns that correlate with hepatitis C and early progression to fibrosis in liver transplant recipients**. *Gastroenterology* 2006, **130**:179-187.

16. Lau DT, Luxon BA, Xiao SY, Beard MR, Lemon SM: **Intrahepatic gene expression profiles and alpha-smooth muscle actin patterns in hepatitis C virus induced fibrosis**. *Hepatology* 2005, **42**:273-281.

17. Chen L, Borozan I, Feld J, Sun J, Tannis LL, Coltescu C, Heathcote J, Edwards AM, McGilvray ID: **Hepatic gene expression discriminates responders and nonresponders in treatment of chronic hepatitis C viral infection**. *Gastroenterology* 2005, **128**:1437-1444.

18. Shao RX, Hoshida Y, Otsuka M, Kato N, Tateishi R, Teratani T, Shiina S, Taniguchi H, Moriyama M, Kawabe T *et al.*: **Hepatic gene expression profiles associated with fibrosis progression and hepatocarcinogenesis in hepatitis C patients**. *World J Gastroenterol* 2005, **11**:1995-1999.

19. Mannick EE, Schurr JR, Zapata A, Lentz JJ, Gastanaduy M, Cote RL, Delgado A, Correa P, Correa H: **Gene expression in gastric biopsies from patients infected with *Helicobacter pylori***. *Scand J Gastroenterol* 2004, **39**:1192-1200.

20. Wen S, Felley CP, Bouzourene H, Reimers M, Michetti P, Pan-Hammarstrom Q: **Inflammatory gene profiles in gastric mucosa during *Helicobacter pylori* infection in humans**. *J Immunol* 2004, **172**:2595-2606.

21. Everall I, Salaria S, Roberts E, Corbeil J, Sasik R, Fox H, Grant I, Masliah E: **Methamphetamine stimulates interferon inducible genes in HIV infected brain**. *J Neuroimmunol* 2005, **170**:158-171.

22. Masliah E, Roberts ES, Langford D, Everall I, Crews L, Adame A, Rockenstein E, Fox HS: **Patterns of gene dysregulation in the frontal cortex of patients with HIV encephalitis**. *J Neuroimmunol* 2004, **157**:163-175.

23. Gelman BB, Soukup VM, Schuenke KW, Keherly MJ, Holzer C III, Richey FJ, Lahart CJ: **Acquired neuronal channelopathies in HIV-associated dementia**. *J Neuroimmunol* 2004, **157**:111-119.

24. Li Y, Zhou X, St John MA, Wong DT: **RNA profiling of cell-free saliva using microarray technology**. *J Dent Res* 2004, **83**:199-203.

25. • Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK *et al.*: **A network-based analysis of systemic inflammation in humans**. *Nature* 2005, **437**:1032-1037.

The authors build a knowledge-based network analysis tool for identifying functional groups of genes that are regulated in response to a specific stimulus and apply it to a human model of endotoxin exposure.

26. Prabhakar U, Conway TM, Murdock P, Mooney JL, Clark S, Hedge P, Bond BC, Jazwinska EC, Barnes MR, Tobin F *et al.*: **Correlation of protein and gene expression profiles of inflammatory proteins after endotoxin challenge in human subjects**. *DNA Cell Biol* 2005, **24**:410-431.

27. National Institutes of Health: **NIH Consensus Statement on Management of Hepatitis C: 2002**. *NIH Consens State Sci Statements* 2002, **19**:1-46.

28. Morgan KT, Jayyosi Z, Hower MA, Pino MV, Connolly TM, Kotlenga K, Lin J, Wang M, Schmidts HL, Bonnefoi MS *et al.*: **The hepatic transcriptome as a window on whole-body physiology and pathophysiology**. *Toxicol Pathol* 2005, **33**:136-145.

29. Bieche I, Asselah T, Laurendeau I, Vidaud D, Degot C, Paradis V, Bedossa P, Valla DC, Marcellin P, Vidaud M: **Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection**. *Virology* 2005, **332**:130-144.

30. Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL *et al.*: **Application of genome-wide expression analysis to human health and disease**. *Proc Natl Acad Sci USA* 2005, **102**:4801-4806.

31. • O'Toole M, Janszen DB, Slonim DK, Reddy PS, Ellis DK, Legault HM, Hill AA, Whitley MZ, Mounts WM, Zuberek K *et al.*: **Risk factors associated with beta-amyloid(1-42) immunotherapy in preimmunization gene expression patterns of blood cells**. *Arch Neurol* 2005, **62**:1531-1536.

This is an interesting study that suggests the possibility of prognostic biomarkers for evaluation of vaccines.

32. • Allison DB, Cui X, Page GP, Sabripour M: **Microarray data analysis: from disarray to consolidation and consensus**. *Nat Rev Genet* 2006, **7**:55-65.

This is a particularly useful review of a rapidly emerging discipline.

33. Baskin CR, Garcia-Sastre A, Tumpey TM, Bielefeldt-Ohmann H, Carter VS, Nistal-Villan E, Katze MG: **Integration of clinical data, pathology, and cDNA microarrays in influenza virus-infected pigtailed macaques (Macaca nemestrina)**. *J Virol* 2004, **78**:10420-10432.

34. McLaren PJ, Mayne M, Rosser S, Moffatt T, Becker KG, Plummer FA, Fowke KR: **Antigen-specific gene expression profiles of peripheral blood mononuclear cells do not reflect those of T-lymphocyte subsets**. *Clin Diagn Lab Immunol* 2004, **11**:977-982.

35. •• Rubins KH, Hensley LE, Jahrling PB, Whitney AR, Geisbert TW, Huggins JW, Owen A, Leduc JW, Brown PO, Relman DA: **The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model**. *Proc Natl Acad Sci USA* 2004, **101**:15190-15195.

The authors evaluate multiple sources of variation in gene transcript abundance patterns, and provide new insight into an infectious disease of immense historical importance.

36. •• Mueller A, Merrell DS, Grimm J, Falkow S: **Profiling of microdissected gastric epithelial cells reveals a cell type-specific response to *Helicobacter pylori* infection**. *Gastroenterology* 2004, **127**:1446-1462.

This study provides an important example of the benefits of measuring transcript abundance patterns from single cell types.

37. Palmer C, Diehn M, Alizadeh AA, Brown PO: **Cell-type specific gene expression profiles of leukocytes in human peripheral blood.** *BMC Genomics* 2006, in press.

38. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC *et al.*: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data**. *Nat Genet* 2001, **29**:365-371.

39. Rocca-Serra P, Brazma A, Parkinson H, Sarkans U, Shojatalab M, Contrino S, Vilo J, Abeygunawardena N, Mukherjee G, Hollowayx E *et al.*: **ArrayExpress: a public database of gene expression data at EBI**. *C R Biol* 2003, **326**:1075-1078.

40. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository**. *Nucleic Acids Res* 2002, **30**:207-210.

41. Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y: **CIBEX: center for information biology gene expression database**. *C R Biol* 2003, **326**:1079-1082.

42. • Pathan N, Hemingway CA, Alizadeh AA, Stephens AC, Boldrick JC, Oragui EE, McCabe C, Welch SB, Whitney A, O'Gara P *et al.*: **Role of interleukin 6 in myocardial dysfunction of meningococcal septic shock**. *Lancet* 2004, **363**:203-209.
The authors pursue leads generated by microarray data with a creative mix of follow-up and functional assays.

43. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, Ochoa MT, Schauber J, Wu K, Meinken C *et al.*: **Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response**. *Science* 2006, **311**:1770-1773.

44. Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O'Shaughnessy A, Gnoj L, Scobie K *et al.*: **A resource for large-scale RNA-interference-based screens in mammals**. *Nature* 2004, **428**:427-431.

45. Tibshirani R, Hastie T, Narasimhan B, Chu G: **Diagnosis of multiple cancer types by shrunken centroids of gene expression**. *Proc Natl Acad Sci USA* 2002, **99**:6567-6572.

46. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S *et al.*: **GoMiner: a resource for biological interpretation of genomic and proteomic data**. *Genome Biol* 2003, **4**:R28.

47. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR: **MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data**. *Genome Biol* 2003, **4**:R7.1-R7.12.

48. Battle A, Segal E, Koller D: **Probabilistic discovery of overlapping cellular processes and their regulation**. *J Comput Biol* 2005, **12**:909-927.

49. • Segal E, Friedman N, Koller D, Regev A: **A module map showing conditional activity of expression modules in cancer**. *Nat Genet* 2004, **36**:1090-1098.
The authors develop a modeling approach for analyzing microarray expression data that reveals functional gene modules and regulatory relationships.

50. Huang SH, Triche T, Jong AY: **Infectomics: genomics and proteomics of microbial infections**. *Funct Integr Genomics* 2002, **1**:331-344.

51. Motley ST, Morrow BJ, Liu X, Dodge IL, Vitiello A, Ward CK, Shaw KJ: **Simultaneous analysis of host and pathogen interactions during an *in vivo* infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment**. *Cell Microbiol* 2004, **6**:849-865.