

Functional and comparative genomics of pathogenic bacteria

Gary K Schoolnik

Microarray expression profiling and the development of data-mining tools and new statistical instruments affords an unprecedented opportunity for the genome-scale study of bacterial pathogenicity. Expression profiles obtained from bacteria grown in media simulating host microenvironments yield a portrait of interacting metabolic pathways and multistage developmental programs and disclose regulatory networks. The analysis of closely related strains and species by microarray-based comparative genomics provides a measure of genetic variability within natural populations and identifies crucial differences between pathogen and commensal. In the near future, the combined use of bacterial and host microarrays to study the same infected tissue will reveal the host–pathogen dialogue in a gene-by-gene and site- and time-specific manner. This review discusses the use of microarray-based expression profiling to identify genes of pathogenic bacteria that are differentially regulated in response to host-specific signals. Additionally, the review describes the application of microarray methods to disclose differences in gene content between taxonomically related strains that vary with respect to pathogenic phenotype.

Addresses

Beckman Center, Room 241, Stanford Medical School, Stanford, California 94305, USA; e-mail: schoolni@cmgm.stanford.edu

Current Opinion in Microbiology 2002, 5:20–26

1369-5274/02/\$ – see front matter

© 2002 Elsevier Science Ltd. All rights reserved.

Abbreviations

BCG	bacille Calmette–Guerin
CSP	competence-stimulating peptide
INH	isoniazid
ORF	open reading frame
SDS	sodium dodecyl sulfate

Introduction

A microarray is a device that provides a surface containing representations of all (or most) of the identified open reading frames (ORFs) of a sequenced and annotated genome. Whether fabricated as a glass-spotted microarray [1–3], high-density oligonucleotide array [4,5] or nylon membrane macroarray, this simple experimental system provides the basis for two quite distinctive experimental paradigms. Functional genomics uses expression profiling of mRNA to provide a condition-specific and time-specific genome-scale snapshot of the transcriptome [6,7]. Comparative genomics contrasts two or more genomes at the ORF-content level of resolution [8,9]. Both applications entail the use of clustering algorithms [10,11] and pathway databases [12,13] to identify co-regulated genes that perform common metabolic and biosynthetic functions. However, microarray expression profiling data poses special analytical challenges that have required the development of new statistical instruments and the recognition that multiple biological replicates of the same experiment are

needed to identify significantly regulated genes [14•,15•,16•]. Although not the topic of this review, some of the most comprehensive microarray studies to date concern the non-pathogenic model systems *Escherichia coli* K-12 [17–23], *Bacillus subtilis* [24] and *Caulobacter crescentus* [25].

This review focuses on two aspects of bacterial pathogenicity, gleaned from an emerging literature that describes the use of genome microarrays to study the biology of infectious agents. Included here are expression profiling studies of bacteria cultured *in vitro* under conditions intended to induce *in vivo* transcriptional programs, to define regulon membership, and to illuminate crucial biosynthetic pathways. Also discussed are the capacity of comparative genomic studies by microarray methods to characterize the genetic variability of natural populations, and to identify differences between pathogen and commensal, and between related pathogens that occupy different host niches or vary in some other subtle manner. The review concludes with a brief summary of expression studies of infected host cells and the promise that this holds for capturing the dialogue between host and pathogen.

Microarray expression profiling of pathogenic bacteria

The ultimate goal of whole genome expression studies of pathogenic bacteria is the identification of bacterial genes that are differentially regulated in the host. Within this class of genes are those that adapt the microbe to host-specific microenvironments or encode virulence determinants. Ideally, studies of this kind would compare expression profiles of bacteria within infected tissues with profiles from bacteria cultured under standardized *in vitro* conditions of growth. Unfortunately, this technically formidable goal has not been achieved using non-amplification methods because the number of organisms within infected tissues is often small, RNA from host cells is vastly more abundant than bacterial RNA, and no efficient method to differentially extract stabilized bacterial RNA from tissues has been described. Because of these considerations, all published microarray reports of pathogenic bacteria have studied organisms grown *in vitro*. Inherent in this experimental design is the quite reasonable assumption that the expression of genes encoding virulence determinants can often be induced *in vitro* by simple modifications of laboratory media [26].

Iron limitation

Paustian *et al.* [27•] studied the transcriptional response of the animal pathogen *Pasteurella multocida* growing under iron-limited conditions *in vitro*. Available free iron is limited in host tissues and iron deprivation is known to induce a complex iron-scavenging response. To characterize the *P. multocida* low-iron response, the organism was grown either in an iron-replete medium or in the same medium

containing an iron chelator. The expression profile, obtained with a DNA microarray containing 96% of the organism's 2014 ORFs, comprised 135 genes (~7% of the genome), including many known in other organisms to be iron-regulated. However, the differential regulation of many other genes showed that the transcriptional response to a change in the concentration of only one metal is complex and pleiotrophic, and points to the interdependency of multiple metabolic pathways. This result also raises the following question, typical of many whole genome expression responses: which genes in the overall expression profile compose the primary or direct response to iron limitation, and which are secondary, downstream consequences of an intracellular deficiency in iron that might affect multiple pathways? Direct effects are best identified by the use of short time courses and experimental conditions that do not cause differences in growth rate between the experimental and control cultures.

Adaptation to acidic microenvironments

The pH of some host microenvironments, like the concentration of available iron, challenges the capacity of bacteria to adapt, especially in strongly acidic organs. The most extreme of these, the stomach, normally reaches pH values as low as 2.0, and virtually all bacteria that ultimately compose the rich and varied microflora of the colon and distal small intestine have successfully transited the 'gastric acidity barrier' before coming to reside in the more alkaline environment of the intestine. Some pathogenic species, particularly *Shigella spp.* and the enterohemorrhagic and enteroinvasive *E. coli* biotypes are particularly acid-resistant [28], explaining why infections with these organisms can be initiated by fewer than 30 bacteria.

Growth of *E. coli* with acetate, which lowers the intracellular pH, significantly increases the acid resistance of the *E. coli*. To learn more about the mechanism of acetate-induced acid tolerance, Arnold *et al.* [29•] incubated enterohemorrhagic *E. coli* serotype O157:H7 in a pH 7.0 medium, with or without 100 mM acetate, and then obtained an expression profile using an *E. coli* K-12 genome membrane macroarray lacking the 1387 ORFs that are present in the *E. coli* O157:H7 genome but absent in the genome of *E. coli* K-12 [30]. Despite this limitation, among the 26 upregulated genes were six genes previously known to specify functions that defend the intracellular pH during acidic conditions of growth. Also induced by acetate were: *cfa*, whose product cyclopropanates unsaturated fatty acids in the inner membrane, possibly decreasing its permeability to protons; *hdeA*, which specifies a periplasmic chaperon hypothesized to prevent acid-induced denaturation of periplasmic proteins; and three oxidative-stress genes, *dps*, *katE* and *grxB*, indicating that exposure to acetate generates reactive oxygen species.

Helicobacter pylori is the most remarkable of the acid tolerant group. It chronically colonizes gastric mucous membranes and is, thus, normally exposed, often for decades, to an

acidic environment. To identify the transcriptional response of *H. pylori* to acidic conditions of growth, Ang *et al.* [31] grew a recent gastric isolate on Columbia agar titrated either to pH 7.2 or to pH 5.5. An expression profile obtained 48 hours thereafter with a membrane macroarray containing 96% of the 1534 predicted ORFs of *H. pylori* strain 26695 identified 80 acid-upregulated ORFs. Sixteen ORFs were previously known to be acid-induced either in *H. pylori* or in other species, whereas an additional 43 functionally annotated ORFs code for proteins not previously known to be acid-regulated. Thus, only a minority of the identified acid-regulated genes could be directly associated with acid tolerance by a plausible physiological mechanism or by reference to prior work. In part, this could reflect the complex and relatively unexplored nature of the acid-resistant phenotype in this species and the large number of its ORFs that code for proteins of unknown function. However, in part, this may also come from this study's use of a solid medium, which could result in a heterogeneous growth environment because of the generation of metabolic gradients in the surrounding agar. It may also result from the protracted 48-hour time course, at the end of which the transcriptome likely reflects a new steady state rather than the adaptive process itself.

Cell-density-dependent gene regulation

Streptococcus pneumoniae, a common cause of pneumonia and meningitis, ordinarily colonizes pharyngeal mucous membranes, a site where the development of genetic diversity in this species is favored by transformation with DNA from the other strains and species that inhabit this ecosystem. Competence — the capacity to bind and take up extraneous DNA — is a transient, temporally programmed physiological state induced by competence-stimulating peptide (CSP). Rimini *et al.* [32••] studied the kinetics of CSP-induced gene expression using a *S. pneumoniae* membrane macroarray. After addition of CSP, results from expression profiling corroborated the kinetics of the previously documented early and late competence transcriptional patterns. However, the identification of 23 other upregulated genes not previously recognized to be associated with competence was more important, as was the unexpected discovery of seven downregulated CSP genes. This study demonstrates how microarray expression profiling can illuminate a programmed physiological process by relating the transcriptional state of genes to a time course that marks its initiation, manifestation and decline. It also shows that a genome-wide study will nearly always reveal new information, even about a well-studied phenomenon. Other examples of this kind with non-pathogenic bacteria include studies of the cell cycle in *C. crescentus* and sporulation in *B. subtilis* [24,25].

De Saizieu *et al.* [33••] used a high-density oligonucleotide Affymetrix *S. pneumoniae* array to study a bacteriocin-like peptide (BlpC) two-component quorum-sensing system that is remarkably similar with respect to its regulation, processing, export and signal transduction to the competence

system described above. The expression profile that was induced by the addition of BlpC to an exponentially growing culture included 16 genes that are clustered on the chromosome near *blpHR*, which encodes the cognate two-component system. These results and the demonstration by Throup *et al.* [34] that a *blpHR* mutant exhibits attenuated virulence in a murine model of pneumococcal pneumonia shows that microarray expression analysis was able to identify a quorum-sensing system that is expressed *in vivo* and contributes to pathogenicity.

Low-oxygen gene regulation and induction of dormancy

Like *S. pneumoniae*, *Mycobacterium tuberculosis* mainly infects the lung. However, unlike patients with pneumococcal pneumonia, an acute infectious process, most people infected with *M. tuberculosis* have a latent form of the disease and are non-infectious and asymptomatic. *In vitro*, a state of non-replicating persistence can be produced by allowing a non-stirred culture to generate a low oxygen gradient as the respiring bacteria settle to the bottom of a culture tube [35]. Non-replicating cultures of this kind, viable for months and perhaps years, can be resuscitated to the replicating state by re-introduction of oxygen. To identify *M. tuberculosis* genes differentially regulated by hypoxia, Sherman *et al.* [36**] shifted an early exponential-phase culture from growth in air (~20% O₂) to growth in a hypoxic atmosphere (0.2% O₂, 99.8% N₂). The genes induced by hypoxia were identified using a DNA microarray containing >97% of the 3924 identified ORFs. Forty-seven ORFs were upregulated and, although approximately two-thirds of the induced genes are of unknown function, several with annotated functions are plausibly involved with adaptation to hypoxia. Among these genes is *acr*, which encodes α -crystallin, a 14 kDa protein with chaperone activity that was previously shown to accumulate during non-replicating persistence [37]. Several genes coding for putative transcription factors were also identified within the induced gene set, including two within a three-gene operon. One of these two is predicted to encode a membrane-bound sensor histidine kinase and the other a two-component response regulator. Mutational analysis of this operon showed that disruption of the response regulator, but not the adjacent sensor histidine kinase, prevented hypoxic induction of *acr*, whose expression was thought to reflect the transcriptional state of other genes within the low-oxygen stimulon. Compared to the wild-type parent strain, this mutant survived less well in late stationary phase. The identification of the cognate response regulator of this low-oxygen response exemplifies how microarray expression analysis can be used to identify transcription factors and dissect regulatory networks.

σ factor regulator cascades

M. tuberculosis dormancy may be considered to be a kind of developmental program that is characterized by temporally ordered transcriptional events governed by a hierarchy of transcription factors, including alternative σ factors. The extracytoplasmic function (ECF) subset of alternative σ

factors are of particular interest because, in other species, some have been found to confer adaptive responses to environmental factors and stress or to be required for virulence. Manganeli *et al.* [38**] used a combination of mutational and microarray methods to define the function and characterize the regulon of the *M. tuberculosis* ECF σ factor, σ^E . Disruption of *sigE* yielded a strain that was more sensitive than the wild-type parent to heat shock, the ionic detergent sodium dodecyl sulfate (SDS) and to oxidants, and that exhibited impaired growth in macrophage cell lines. To identify σ^E -regulated genes, expression profiles were obtained from mid-exponential-phase cultures of a σ^E mutant and the wild-type parent that had been exposed to 0.05% SDS, a treatment that likely perturbs cell envelope lipids. Twenty-three genes were identified whose SDS-induced expression required σ^E . Among these were *sigB*, *aceA* (which encodes isocitrate lyase of the glyoxalate shunt, an activity required for full virulence in a murine model of tuberculosis) [39] and *fadB2* (which encodes 3-hydroxyacyl CoA dehydrogenase, a component of the fatty-acid β -oxidation pathway that may enable assimilation of host fatty acids). These data and results from the preceding study demonstrate that microarray expression analysis can, in principle, identify gene sets whose regulation requires, directly or indirectly, each of the annotated alternative σ factors or two-component regulators in the genome. Moreover, because some of the regulated genes will be other transcription factors, it should be possible to reconstruct hierarchies of regulatory networks from these data.

Inhibition of biosynthetic pathways and the identification of new drug targets

M. tuberculosis is not only a formidable human pathogen, but is also increasingly difficult to treat because of emerging resistance to one or more antitubercular drugs, including isoniazid (INH). It has long been known that INH blocks the biosynthesis of mycolic acids, an essential component of the mycobacterial cell envelope, and recent biochemical studies indicate that it does so by inhibiting the type II fatty-acid synthase (FAS-II) complex that is required for the production of the full-length meromycolate chain, either by binding NADH within the active site of enoyl-acyl carrier protein reductase (InhA) [40] or by forming a ternary complex with β -ketoacyl-ACP-synthase (KasA) and the acyl carrier protein, AcpM [41].

Wilson *et al.* [42] obtained expression profiles from INH-treated mid-log phase cultures of *M. tuberculosis*; only 14 differentially regulated ORFs were identified out of the 3834 whose expression state had been modified. Amongst these genes was the induction by INH of an operon-like cluster that encodes components of the FAS-II complex, including *AcpM* and *KasA*. This result, evident after only 40 minutes, indicates that an expression profile can provide useful information about a compound's mode of action and demonstrates that inhibition of a biosynthetic pathway is sensed and responded to at the transcriptional level within minutes. Also induced was *fbpC*, which

encodes trehalose-dimycolyl transferase, an activity at the end of the mycolate biosynthetic pathway that esterifies mycolic acids with cell-wall carbohydrates. This result shows that an expression result can illuminate components of a multicomponent pathway that are remote from its direct site of action — newly identified pathway components could contribute to the drug discovery process by disclosing novel drug targets.

Microarray-based comparative genomics

Genetic variability and natural selection yield strains and species adapted to particular microenvironments of the host and result in phenotypic differences between non-pathogenic commensals and virulent biotypes. Accordingly, genomic comparisons between pathogenic and non-pathogenic strains of the same species can be particularly informative because genes exclusively present in the former may be required for infectivity, virulence or adaptation to a particular host niche. Microarray-based comparisons between a fully sequenced genome and an unsequenced, but related, genome can provide valuable information about the diversity and evolution of pathogens and symbionts [8,9]. Comparisons of this kind employ a microarray containing representations of all the ORFs of the sequenced, reference strain and labeled DNA from the unsequenced, experimental strain. The resulting hybridized array will disclose genes common to both strains and genes that are present in the reference strain but absent in the experimental strain. This method, however, cannot detect genes present in the experimental strain but absent in the reference strain: point mutations, including frame-shift mutations; small deletions and deletions in homologous repetitive elements; rearrangements of the genome that have not resulted in deletion of a gene; and differences in the number of multicopy genes [8,9,43**]. Additionally, in contrast to high-density oligonucleotide arrays, DNA-spotted microarrays and membrane macroarrays do not ordinarily include representations of intergenic regions of the genome and, thus, cannot detect deletions within these non-coding segments, even though these specify promoter elements and small, non-translated RNAs and thus could be functionally important [44]. Despite these limitations, the few published studies of this kind have been quite informative, in part because events leading to gene acquisition and gene loss are a major source of diversity in bacterial pathogens [8] and many changes of this kind are readily detected by microarray methods.

M. tuberculosis, *M. bovis* and BCG vaccine strains

Behr *et al.* [45], in perhaps the first example of a study of this kind, used a DNA microarray to compare the genome composition of the sequenced *M. tuberculosis* laboratory strain H37Rv with the closely related pathogenic species, *M. bovis*, and with several strains of the bacille Calmette–Guerin (BCG) vaccine variant that was produced by serial *in vitro* passage of *M. bovis* between 1908 and 1921. Compared to *M. tuberculosis*, 11 regions containing 91 ORFs were found to have been deleted from one or more

of the tested pathogenic *M. bovis* strains. Furthermore, compared to pathogenic *M. bovis* strains, five additional regions containing 38 ORFs were deleted from one or more of the tested BCG strains. Analysis of the regions deleted from BCG, but present in the sequenced *M. tuberculosis* strain, showed that genes classified as transcriptional regulators were lost disproportionately and may, thus, control the expression of genes required for virulence. When analyzed within a historical context, these results show that microarray-based comparative genomics can be used to reconstruct the genealogy of related strains at the genomic level of resolution. In a second study by the same group, Kato-Maeda *et al.* [46*] used a high-density oligonucleotide Affymetrix array to compare 19 recent clinical isolates of *M. tuberculosis*. Compared to the sequenced reference strain, each unique clinical isolate was found to have lost, on average, ~17 ORFs corresponding to ~0.3% of the H37Rv genome. In all, 25 deleted sequences were detected, including 22 intergenic segments and all or part of 93 ORFs. On the basis of their functional annotations, several of the deleted genes could conceivably affect virulence, including three encoding phospholipase-C, one encoding a polyketide synthase and three encoding putative transcriptional regulators. Remarkably, strains that had sustained the most deletions were less likely to have been isolated from patients with pulmonary cavitation. Cavity formation is a hallmark of tuberculosis and essential for the efficient transmission of the organism to susceptible hosts. Thus, degradation of the genome may be associated in this species with a trend to decreased infectivity.

H. pylori strain diversity

H. pylori infection of the upper gastrointestinal tract causes a spectrum of conditions ranging from asymptomatic infection to gastritis, gastric and duodenal peptic ulcer disease and gastric cancer. Comparison of two complete *H. pylori* sequences revealed that ~6% of each genome was not present in the other genome and that recombinations, insertions and deletions, changes in repetitive elements and single-nucleotide substitutions had created considerable diversity [47]. To further explore the genomic diversity of this species, Salama *et al.* [43**] used a microarray representing 98.6% of the ORFs of both sequenced species to examine the genomic content of 15 *H. pylori* strains. They identified 1281 ORFs that were common to all the tested strains; these represent the ‘functional core’ of this species’ genome. Among them were genes coding for metabolic and biosynthetic pathways and for cellular and regulatory functions. By contrast, 362 ORFs, comprising 22% of the genome, were absent from one or more of the tested strains; these comprise strain-specific genes and were hypothesized to encode functions that adapt the organism to a particular host niche. An intriguing aspect of this study was the use of a clustering algorithm for the analysis of strain-specific genes and the identification of several genes that may have been co-inherited with genes in the pathogenicity island and may therefore also encode virulence determinants. In a separate study, Israel *et al.* [48] used the

same *H. pylori* microarray to compare the genomic content of two clinical strains that produce significantly different levels of gastritis, cellular proliferation and apoptosis in the gerbil gastritis model. The microarray results showed that the less proinflammatory strain had sustained a large deletion of the *cag* pathogenicity island, providing a genetic explanation for its relative attenuation.

Profiling the dialogue between pathogen and host

Pathogenesis entails not only the differential expression of bacterial genes, but also responses by the host. In principle, then, microarray expression analysis of bacterially infected cells and tissues can identify, simultaneously and in the same sample, host and pathogen genes that are regulated during the infectious process. Although not within the scope of this review, the host contribution to this process has been the focus of five published microarray studies [49*,50**,51*,52*,53**] characterizing the host-cell response to attached or invading bacteria. Each of these studies used transformed cell lines, relatively large multiplicities-of-infection (MOIs), and short time courses, so the results reflect early events. Three of these studies employed macrophage cell lines [49*,50**,51*] and therefore explored aspects of innate rather than acquired immunity. Four of these studies [49*,50**,52*,53**] also obtained expression profiles of cell lines exposed to purified bacterial products, including pro-inflammatory cell-wall constituents and toxins, or to bacteria with mutations in genes coding for virulence determinants. This successful strategy identified microbial molecules responsible for the induction of some components of the host-cell transcriptional response.

The referenced studies illustrate that experiments using transformed cell lines are informative, but results from the use of primary cells freshly isolated from tissues, for example, primary bone-marrow-derived macrophages, may better simulate *in vivo* conditions. However, studies using single cell types, whether primary or transformed, lack the context of a multicellular milieu where signaling between cells of different lineages modulates the response of individual cells to an infectious agent. This kind of complexity is difficult to simulate *in vitro* and will require technical innovations that permit expression profiling of individual cell types within infected tissues.

Conclusions

The studies reviewed here show that microarray expression profiling is a powerful method to identify genes differentially regulated by biochemical signatures of host microenvironments, genes that are controlled, directly or indirectly, by transcription factors and genes that code for components of multistep metabolic and biosynthetic pathways. Equally illuminating are microarray-based comparative studies to assess the extent and nature of genetic variability within natural populations of related species and strains and to delineate differences, at the ORF level of resolution, between pathogen and commensal. Complementary

expression studies of infected host cells have been conducted as well and hold considerable promise for capturing the intricate sequence of measure and counter-measure between pathogen and host. Thus far, however, all such studies have been carried out *in vitro*. Technical innovations will be necessary before microarray-based bacterial expression studies of infected tissues can be conducted. The availability of improved methods and more powerful bioinformatic tools will provide whole-genome portraits of the transcriptomes of pathogen and host in a time-, tissue- and cell-specific manner.

Acknowledgements

I thank past and current members of my laboratory for their contributions to the art and science of microarray expression work in my laboratory, which was supported by grants from the National Institutes of Health (AI 39521, AI 43422 and AI44826) and the Glaxo Group Research Action TB Program.

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. Eisen MB, Brown PO: **DNA arrays for analysis of gene expression.** *Methods Enzymol* 1999, **303**:179-205.
 2. De Risi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.
 3. Schoolnik GK, Voskuil MI, Schnappinger D, Yildiz FH, Meibom K, Dolganov NA, Wilson MA, Chong KH: **Whole genome DNA microarray expression analysis of biofilm development by *Vibrio cholerae* O1 El Tor.** *Methods Enzymol* 2001, **336**:3-18.
 4. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays.** *Nat Genet* 1999, **21**(Suppl 1):20-24.
 5. Lockhart DJ, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring by hybridization to high-density oligonucleotide arrays.** *Nat Biotechnol* 1996, **14**:1675-1680.
 6. Ferea TL, Brown PO: **Observing the living genome.** *Curr Opin Gen Dev* 1999, **9**:715-722.
 7. Harrington CA, Rosenow C, Retief J: **Monitoring gene expression using DNA microarrays.** *Curr Opin Microbiol* 2000, **3**:285-291.
 8. Ochman H, and Jones IB: **Evolutionary dynamics of full genome content in *Escherichia coli*.** *EMBO* 2000, **19**:6637-6643.
 9. Ochman H, Moran NA: **Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis.** *Science* 2001, **292**:1096-1098.
 10. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
 11. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.** *Proc Natl Acad Sci USA* 1999, **96**:2907-2912.
 12. Karp PD, Krummenacker M, Paley S, Wagg J: **Integrated pathway-genome databases and their role in drug discovery.** *Trends Biotechnol* 1999, **17**:275-281.
 13. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A: **The EcoCyc and MetaCyc databases.** *Nucleic Acids Res* 2000, **28**:56-59.
 14. Arfin SM, Long AD, Ito ET, Tolleri L, Riehle MM, Paegle ES, Hatfield GW: **Global gene expression profiling in *Escherichia coli* K-12.** *J Biol Chem* 2000, **275**:29672-29684.

This is a seminal study that proposes a new statistical method for the analysis of microarray expression data on the basis of a linear analysis of variance (ANOVA) model. Use of this instrument to analyze an *E. coli* K-12 membrane microarray study of the effect of integration host factor showed

that arbitrarily selected fold-difference criteria cannot reliably identify genes whose expression levels are significantly different between two growth conditions.

15. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.

An alternative statistical method is described, designated significance analysis of microarrays (SAM), for the analysis of microarray expression data. An attractive feature of this program is the use of an adjustable threshold allowing an investigator-selected false discovery rate. This method and the ANOVA-based method in [14**] are available via the Internet.

16. Lee M-L, Kuo FC, Whitmore GA, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.** *Proc Natl Acad Sci USA* 2000, **97**:9834-9839.

The inherent variability of microarray expression datasets is demonstrated, leading to the recommended use of at least three experimental replicates.

17. Richmond CS, Glasner JD, Mau R, Jin H, Blattner FR: **Genome-wide expression profiling in *Escherichia coli* K-12.** *Nucleic Acids Res* 1999, **27**:3821-3835.

18. Zheng M, Wang X, Templeton LJ, Smulski DR, LaRossa RA, Storz G: **DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide.** *J Bacteriol* 2001, **183**:4562-4570.

19. Wei Y, Lee J-M, Richmond C, Blattner FR, Rafalski JA, LaRossa RA: **High-density microarray-mediated gene expression profiling of *Escherichia coli*.** *J Bacteriol* 2001, **183**:545-556.

20. Tao H, Bausch C, Richmond C, Blattner FR, Conway T: **Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media.** *J Bacteriol* 1999, **181**:6425-6440.

21. Zimmer DP, Soupene E, Lee HL, Wendisch VF, Khodursky AB, Peter BJ, Bender RA, Kustu S: **Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation.** *Proc Natl Acad Sci USA* 2000, **97**:14674-14679.

22. Khodursky AB, Peter BJ, Cozzarelli NR, Botstein D, Brown PO, Yanofsky C: **DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*.** *Proc Natl Acad Sci USA* 2000, **97**:12170-12175.

23. Barbosa TM and Levy SB: **Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA.** *J Bacteriol* 2000, **182**:3467-3474.

24. Fawcett P, Eichenberger P, Losick R, Youngman P: **The transcription profile of early to middle sporulation in *Bacillus subtilis*.** *Proc Natl Acad Sci USA* 2000, **97**:8063-8068.

25. Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L: **Global analysis of the genetic network controlling a bacterial cell cycle.** *Science* 2000, **290**:2144-2148.

26. Mekalonos JJ: **Environmental signals controlling the expression of virulence determinants in bacteria.** *J Bacteriol* 1992, **174**:1-7.

27. Paustian ML, May BJ, Kapur V: ***Pasteurella multocida* gene expression in response to iron limitation.** *Infect Immun* 2001, **69**:4109-4115.

This is the first published genome-scale microarray expression study of iron deprivation. The results reveal unanticipated complexity, suggesting a system-wide adaptation that involves interdependent metabolic pathways.

28. Gordon J, Small PL: **Acid resistance in enteric bacteria.** *Infect Immun* 1993, **61**:364-367.

29. Arnold CA, McElhanon J, Lee A, Leonhart R, Siegele DA: **Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response.** *J Bacteriol* 2001, **183**:2178-2186.

Adaptation to growth in acetate by the acid-resistant enterohemorrhagic *E. coli* biotype is shown to involve multiple homeostatic adjustments, some not previously identified by other methods of investigation.

30. Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA *et al.*: **Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.** *Nature* 2001, **409**:529-533.

31. Ang S, Lee C-Z, Peck K, Sindici M, Matrubutham U, Gleeson MA, Wang JT: **Acid-induced gene expression in *Helicobacter pylori*:**

study in genomic scale by microarray. *Infect Immun* 2001, **69**:1679-1686.

32. Rimini R, Jansson B, Feger G, Roberts TC, de Francesco M, Gozzi A, Faggioni F, Domenici E, Wallace DM, Frandsen N, Polissi A: **Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays.** *Mol Microbiol* 2000, **36**:1279-1292.

See annotation to [33**].

33. De Saizieu A, Gardes C, Flint N, Wagner C, Kamber M, Mitchell TJ, Keck W, Amrein KE, Lange R: **Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide.** *J Bacteriol* 2000, **182**:4696-4703.

Papers [32**,33**] used synthetic peptide pheromones to initiate quorum-sensing developmental programs. The results demonstrate how microarray analysis can provide a genome-wide portrait of a temporally orchestrated, multistep process.

34. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, Marra A, Wallis NG, Brown JR, Holmes DJ *et al.*: **A genomic analysis of two-component signal-transduction in *Streptococcus pneumoniae*.** *Mol Microbiol* 2000, **35**:566-576.

35. Wayne LG: **Dynamics of submerged growth of *Mycobacterium tuberculosis* under aerobic and microaerophilic conditions.** *Am Rev Respir Dis* 1976, **114**:807-811.

36. Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK: **Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding α -crystallin.** *Proc Natl Acad Sci USA* 2001, **98**:7534-7539.

Microarray analysis of adaptation to hypoxia discloses the cognate response regulator of the low oxygen regulon, thus demonstrating the capacity of expression profiling to identify regulators of crucial physiological responses.

37. Yuan Y, Crane D, Simpson RM, Zhu Y, Hickey MJ, Sherman DR, Barry CE III: **The 16-kDa α -crystallin (Acr) protein of *Mycobacterium tuberculosis* is required for growth in macrophages.** *Proc Natl Acad Sci USA* 1998, **95**:9578-9583.

38. Manganello R, Voskuil MI, Schoolnik GK, Smith I: **The *Mycobacterium tuberculosis* ECF sigma factor σ^E : role in global gene expression and survival in macrophages.** *Mol Microbiol* 2001, **41**:423-437.

A proof of principle is provided by this study of a *M. tuberculosis* ECF alternative σ factor, in which it was shown how the use of σ -factor mutants and expression profiling can define regulons.

39. McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jacobs WR Jr, Russell DG: **Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase.** *Nature* 2000, **406**:735-738.

40. Rozwarski DA, Grant GA, Barton DH, Jacobs WR Jr, Sacchettini JC: **Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*.** *Science* 1998, **279**:98-102.

41. Mdluli K, Slayden RA, Zhu Y, Ramaswamy S, Pan X, Mead D, Crane DD, Musser JM, Barry CE III: **1998 Inhibition of a *Mycobacterium tuberculosis* beta-ketoacyl ACP synthase by isoniazid.** *Science* 1998, **280**:1607-1610.

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

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

A comprehensive comparative genome database of multiple *H. pylori* strains and the use of a clustering algorithm identified genes of unknown function likely to have been co-inherited with the *cag* pathogenicity island and, thus, possibly encoding new virulence determinants.

44. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S: **Identification of novel small RNAs using comparative genomic microarrays.** *Genes Dev* 2001, **15**:1637-1651.

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

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

This paper and [45] demonstrate how microarray-based comparative genomics can measure variability within natural populations of the same species and between closely related species, and lead to the identification of deletions that may explain differences in infectivity, virulence and immunogenicity.

47. Alm RA, Ling L-S, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL *et al.*: **Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori***. *Nature* 1999, **397**:176-180.

48. Israel DA, Salama N, Arnold CN, Moss SF, Ando T, Wirth HP, Tham KT, Camorlinga M, Blaser MJ, Falkow S, Peek RM Jr: ***Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses**. *J Clin Invest* 2000, **107**:611-620.

49. Rosenberger CM, Scott MG, Gold MR, Hancock RE, Finlay BB:
 • ***Salmonella typhimurium* infection and lipopolysaccharide stimulation induce similar changes in macrophage gene expression**. *J Immun* 2000, **164**:5894-5904.

A murine array containing 588 genes and expressed sequence tags (ESTs) identified macrophage genes induced by *Salmonella typhimurium* and purified lipopolysaccharide. Both induced previously reported components of innate immunity.

50. Detweiler C, Cunanan DB, Falkow S: **Host microarray analysis •• reveals a role for the *Salmonella* response regulator *phoP* in human macrophage cell death**. *Proc Natl Acad Sci USA* 2001, **98**:5850-5855.

An array containing 22 571 human genes and ESTs was used to study the macrophage response to *S. typhimurium* and to an isogenic *phoP* mutant. PhoP was found to be required for the induction of macrophage genes associated with programmed cell death.

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

A high-density Affymetrix human array containing 6800 genes showed that *Listeria monocytogenes* induces not only a generalized innate immune response in a macrophage cell line, but also pathogen-specific genes coding for actin-binding protein and tubulin, reflecting recruitment and reorganization of cytoskeletal proteins by *Listeria* during the process of invasion.

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

Pseudomonas aeruginosa cells attached to a type II pneumocyte cell line induced a variety of genes, including the gene for interferon regulatory factor 1 (IRF-1), a previously described positive regulator of inducible nitric oxide synthase (iNOS). This result provides a possible explanation for the presence of nitric oxide in the infected sputa of cystic fibrosis patients.

53. Belcher CE, Drenkow J, Kehoe B, Gingeras TR, McNamara N,
 •• Lemjabbar H, Basbaum C, Relman DA: **The transcriptional response of respiratory epithelial cells to *Bordetella pertussis* reveal host defensive and pathogen counter-defensive strategies**. *Proc Natl Acad Sci USA* 2000, **97**:13847-13852.

This paper reports a comprehensive, high-density, oligonucleotide, Affymetrix human array study of the interaction of *Bordetella pertussis* with a transformed bronchial epithelial cell line. The study showed that, in addition to a generalized NFκB-dependent, pro-inflammatory response, the following genes were indicated: genes specifically activated by pertussis toxin; genes coding for respiratory tract mucins; and genes that specifically encode anti-apoptosis factors. The resulting expression profile provides a transcriptional portrait that is congruent with the histopathology of whooping cough in man.