Molecular Microbiology (2003) **47**(4), 879–889

MicroReview

Microarray expression profiling: capturing a genome-wide portrait of the transcriptome

Tyrrell Conway[1]* and Gary K. Schoolnik[2]
[1]*Advanced Center for Genome Technology, The University of Oklahoma, Norman, OK 73019-0245, USA.*
[2]*Departments of Medicine (Infectious Diseases and Geographic Medicine) and Microbiology and Immunology, Stanford Medical School, Stanford, CA 94304, USA.*

Summary

The bacterial transcriptome is a dynamic entity that reflects the organism's immediate, ongoing and genome-wide response to its environment. Microarray expression profiling provides a comprehensive portrait of the transcriptional world enabling us to view the organism as a 'system' that is more than the sum of its parts. The vigilance of microorganisms to environmental change, the alacrity of the transcriptional response, the short half-life of bacterial mRNA and the genome-scale nature of the investigation collectively explain the power of this method. These same features pose the most significant experimental design and execution issues which, unless surmounted, predictably generate a distorted image of the transcriptome. Conversely, the expression profile of a properly conceived and conducted microarray experiment can be used for hypothesis testing: disclosure of the metabolic and biosynthetic pathways that underlie adaptation of the organism to changing conditions of growth; the identification of coordinately regulated genes; the regulatory circuits and signal transduction systems that mediate the adaptive response; and temporal features of developmental programmes. The study of bacterial pathogenesis by microarray expression profiling poses special challenges and opportunities. Although the technical hurdles are many, obtaining expression profiles of an organism growing in tissue will probably reveal strategies for growth and survival in the host's microenvironment. Identifying these colonization strategies and their cognate expression patterns involves a 'deconstruction' process that combines bioinformatics analysis and *in vitro* DNA array experimentation.

Accepted 24 October, 2002. *For correspondence. E-mail tconway@ou.edu; Tel. (+1) 405 325 1683; Fax (+1) 405 325 7619.

Introduction

For many, the jury is still out, and debate over the value of DNA arrays for microbiology research continues. To us, the verdict is in, and the evidence is overwhelming – DNA arrays are providing novel insights into cellular processes. A growing number of high-quality publications demonstrate the power of arrays for microbial investigations. We are convinced that whole-genome expression profiling will shape our thinking about bacterial systems for years to come.

The studies reviewed below provide compelling evidence that DNA arrays are able to resolve the changes in gene expression that accompany adjustments to cellular physiology. Accordingly, researchers are using this technology to identify genes that are differentially expressed in response to changes in environmental parameters, to define developmental programmes and to evaluate mutations in regulatory and metabolic pathways. For many, the ultimate goal is to capture the transcriptome of bacteria growing within infected tissues and thus to disclose the far larger universe of host-adapted transcriptional responses. However, to use this method, experiments must be designed to minimize undesirable artifacts. Where possible, experimental conditions should be chosen to highlight the genetic response of interest without impacting unrelated systems. These features highlight the importance of microarray experimental design. It is our contention that, when properly managed, DNA arrays can provide unparalleled snapshots of condition-specific cellular physiology and moving pictures of programmed genetic events.

Below, we discuss the physiological reasons that underlie the need for exquisite attention to experimental design and execution. We begin by discussing the properties of bacteria that govern mRNA levels. Based on this understanding, we then discuss what DNA arrays can and can-

not do. Lastly, we review several papers that illustrate effective strategies for gene expression profiling of bacteria both in the laboratory and in pathogenesis models.

The bacterial transcriptional response and its implications for microarray experimental design

Underlying mechanisms of bacterial adaptability

The hallmark of bacteria is their adaptability. In many environments, those that adapt most rapidly are most successful. The speed with which bacteria adjust to changing environmental conditions creates significant technical challenges for obtaining high-quality array data from bacterial cells. These rapid responses are complicated further by overlapping global regulatory networks and adaptive genetic programmes. Unless great care is taken to control the numerous growth parameters that can quickly reverberate through the transcriptome, researchers might inadvertently investigate the microbial responses to uncontrolled variables and therefore miss the specific, variable-dependent responses they sought to study. Good array data depend as much upon good microbial physiology technique as they do on good DNA array technique.

When *Escherichia coli* grows at 37°C, the frequency of transcription initiation at any given promoter varies from once per second to once per generation, depending on promoter strength and genetic regulation (Record *et al*., 1996). Transcript elongation proceeds at an approximate rate of 40–50 nucleotides per second with up to 30 ribosomes following closely behind the leading wave of transcription (Richardson and Greenblatt, 1996). At this pace, even for a very large gene such as *lacZ*, the first molecule of beta-galactosidase is synthesized just 1 min after the signal for gene induction is received. Moreover, under optimal conditions, tens of transcripts and hundreds of gene products will be completed in the following minute. Thus, the cellular response to changing signal inputs, and hence transcription rates, is very rapid – on the order of minutes.

A second aspect of RNA metabolism, mRNA turnover, has a large impact on transcript levels and hence on cellular physiology. The average half-life of *E. coli* mRNA, as measured by isotopically labelled RNA decay, is 1 min (Baracchini and Bremer, 1987), with a range of 40 s to 20 min for individual transcripts (Pedersen and Reeh, 1978; Kushner, 1996). Global array analysis of mRNA decay indicates that 80% of *E. coli* transcripts have half-lives ranging from 3 to 8 min (Bernstein *et al*., 2002). Bacterial RNases are responsible for this rapid turnover of mRNA (Kushner, 1996; 2002). This makes it essential that the cell harvest protocol and RNA sampling method include, as early as possible, a step for RNase inactivation. Failure in the experimental design to account for rapid mRNA turnover will invariably lead to overall loss of mRNA in the sample, a decreased ratio of message to stable RNA and bias in transcript representation because of differential turnover rates.

What arrays can and cannot do

DNA arrays are usually printed in miniature with polymerase chain reaction (PCR) products or oligonucleotides on glass microscope slides, or in larger formats on membranes. A number of excellent reviews of DNA array technology are available (Winzeler *et al*., 1999; Cummings and Relman, 2000; Rhodius *et al*., 2002), including comparisons of array types (Baldwin *et al*., 1999; Richmond *et al*., 1999). Each of these technologies has been used successfully for genome-wide expression profiling. The point we would like to make is that the strategies for experimental design and data analysis are essentially the same, regardless of the DNA array technology chosen.

To a first approximation, transcriptome data reflect the relative distribution of RNA polymerase on promoters across the genome (Wei *et al*., 2001; Rhodius *et al*., 2002). Unfortunately, for arrays to be fully quantitative, a control containing known amounts of all mRNA species is required. Of course, this is not practical, but a surrogate correction factor, generated by labelling and hybridization with genomic DNA, has proved useful for obtaining measures of relative transcript abundance (Wei *et al*., 2001; Rhodius *et al*., 2002). In practice, most researchers choose to use an RNA control from cells in a well-defined physiological state (e.g. mid-logarithmic phase), so that conditions appropriately reflect the 'ground state' for their experimental variable(s). This strategy has served well. However, the problem with using mid-log phase RNA as a reference is the lack of expression of non-mid-log phase-specific genes. Accordingly, their absence in the control sample will lead to infinite induction ratios should they be expressed in the experimental sample. Others have chosen to use a mixture of reference RNAs obtained from several sampling conditions, thus avoiding the inherent concerns with a single RNA reference (Laub *et al*., 2000). Recently, a calibrated reference for oligonucleotide arrays was designed using labelled oligonucleotides complementary to every probe on the array (Dudley *et al*., 2002); this strategy offers considerable promise. Once a reference is selected, many researchers favour using it to compare all experimental expression results. Differences in gene expression between two particular experimental conditions are then deduced *in silico* from the difference in each experimental result from the common reference, in part because some microarray-specific statistical packages are more amenable to the use of a common reference. The bottom line is that the choice of an appropriate

control or reference is a critical aspect of experimental design and could significantly affect results.

Even with replication and appropriate controls, DNA arrays only provide a semi-quantitative measure of the relative transcript levels in an RNA sample. On the other hand, neither can the absolute measure of transcript abundance be ascertained by Northern hybridization; what is usually measured is an induction ratio. The semi-quantitative and relative nature of expression results leads to the following question: what is the correspondence between mRNA induction ratios measured with microarrays compared with more quantitative methods (e.g. real-time PCR) and methods that monitor promoter activity (e.g. CAT assays)? The answer seems to be as follows. Although induction ratios >100 can be measured for some genes in some systems, induction ratios, as determined by most microarray methods, are generally <10. This causes confusion because induction ratios of ≈ 2 are often met with scepticism by persons familiar with other assay systems which, for highly induced genes, yield ratios of 10–1000. We could identify no published systematic comparisons of changes in RNA abundance from microarrays versus other methods. Thus, this is currently an unresolved issue for bacterial expression systems. However, reports are available that compare induction ratios measured with arrays with those determined by real-time PCR (Caldwell *et al*., 2001; Helmann *et al*., 2001; Lee *et al*., 2001) and Northern hybridization (Pomposiello *et al*., 2001; Schut *et al*., 2001; Wiegert *et al*., 2001; Khil and Camerini-Otero, 2002). Taken together, these indicate that arrays often, but not always, underestimate induction ratios by two- to 10-fold. The reason for this difference has not been examined thoroughly. Spiking known amounts of *in vitro*-transcribed mRNA into a microarray assay indicates that signal intensity is linear and proportional to the spiked concentration (Caldwell *et al*., 2001), but this validation strategy has not been applied systematically. Many workers in the field suspect that current microarray methods lack the dynamic range of other assay systems, particularly at higher levels of expression. We conclude that arrays most probably underestimate actual mRNA induction ratios, but statistical significance is readily estimated and can be used for ranking genes according to their differential gene expression values.

A common concern is that transcript abundance is not necessarily reflected in the corresponding protein levels. Although this is clearly the case for gene products that are primarily regulated by post-transcriptional processes (e.g. RpoS), we note that, in the three studies in which proteome data were compared directly with transcriptome data, regulation of the majority of genes parallels the levels of their protein products. Although the induction ratios of proteins on two-dimensional gels are almost always lower than the induction ratios of the corresponding mRNAs, the transcriptome data are usually more comprehensive because of the limited number of proteins that can be resolved on two-dimensional gels (Hommais *et al*., 2001; Yoshida *et al*., 2001; Eymann *et al*., 2002). Direct comparisons indicate that two-dimensional gels underestimate the total number of induced proteins/genes by two- to fourfold compared with microarrays. Numerous studies have demonstrated that enzyme levels correlate with their respective gene expression profiles (Arfin *et al*., 2000; Khodursky *et al*., 2000; Smulski *et al*., 2001; Tao *et al*., 2001). We conclude that, in the vast majority of cases, transcript levels parallel their gene product levels in bacterial systems, and we look forward to systematic validation of this statement.

Other limitations of arrays are also apparent. Most arrays do not contain probes for intergenic regions and thus will not monitor genes in regions that were not identified as open reading frames (ORFs), nor will they monitor regulatory RNAs, which are increasingly recognized as being important (Majdalani *et al*., 2001; Wassarman *et al*., 2001; Zhang *et al*., 2002). Likewise, most arrays do not contain probes for stable RNAs, the levels of which increase in proportion to the growth rate, thus complicating analysis of experiments that involve comparing cultures with different growth rates or studies of the stringent response (Smulski *et al*., 2001; Chang *et al*., 2002; Eymann *et al*., 2002). Another limitation concerns the failure of arrays to recognize regulatory events that occur via a post-translational mechanism. Many regulators (e.g. phosphor relays) fall into this group, and so it is not surprising that the fold induction of genes coding for such regulatory factors is often much lower than the fold induction ratios of the genes that they regulate. Therefore, if finding regulators is important, then it is imperative to combine expression studies with mutant screens. However, a close examination of the expression profile may disclose the cognate regulator because genes coding for transcription factors are often clustered on the chromosome with the genes that they regulate. Thus, scanning the local chromosomal anatomy for genes encoding putative regulators that reside near regulated genes is a useful manoeuvre. Lastly, arrays cannot be used to prove mechanisms of gene regulation, nor can arrays easily distinguish direct from indirect regulatory effects. Nevertheless, arrays have proved to be quite successful in describing trends in gene expression patterns that reflect operon, regulon and stimulon organization.

Microarray experimental design

The preceding considerations shape microarray experimental design strategies, which are responsive to the two processes that determine mRNA abundance most significantly: (i) relative mRNA levels change rapidly in

response to changing regulatory signals; and (ii) relative mRNA abundance is controlled dynamically by the rates of transcription and mRNA turnover. In turn, these lead to two experimental design classes: those that measure differences in mRNA abundance between two conditions and those that describe the cellular response to a stepwise change in conditions. The experimental design must allow for true steady-state conditions or include a sampling interval that appropriately reflects the timing of genetic responses and, in both cases, avoid the introduction of artifacts caused by poor sampling technique.

Essentially, there are three versions of the two-condition experimental design class: differential response to growth parameters; treated versus untreated cultures; and wild-type versus mutant strains. Examples of the first of these include differential gene expression studies of *E. coli* grown in minimal versus rich media (Tao *et al*., 1999) and glucose versus acetate as the sole carbon source (Oh *et al*., 2002); and of *Bacillus subtilis* grown on methionine versus methylthioribose as sulphur sources (Sekowska *et al*., 2001) or in conditions that reflect three different modes of anaerobic versus aerobic growth (Ye *et al*., 2000). Each of these experiments involved comparing growth conditions in steady-state cells in mid-logarithmic phase and resulted in lists of significantly induced or repressed genes that appropriately reflected the growth physiology. It is worth noting that, in each of these examples, 'unappreciated differences' in the selected growth conditions (e.g. differences in growth rate) resulted in differential gene expression patterns that required interpretation in the analysis.

Comparison of treated versus untreated cultures is a reasonable strategy for measuring differential gene expression resulting from exposure to agents causing deleterious effects on growth or induction of global regulatory networks. Differential gene expression after the addition of acetate (Arnold *et al*., 2001) or acivicin (Smulski *et al*., 2001) to growing *E. coli* cells and addition of DL-norvaline to growing *B. subtilis* cells (Eymann *et al*., 2002) led to the identification of differentially expressed genes involved, respectively, in the acid tolerance response, starvation for histidine and starvation for isoleucine and leucine. Interestingly, the last two treatments led to the induction of the stringent response, as reflected in the gene expression profiles. We note that, in each of these examples, the identification of genes by the expression profiling studies was complemented by excellent physiological and biochemical experiments, which significantly enhanced the value of the publications. Although the experimental strategy of 'splitting the culture' before treatment requires culture manipulations that can cause inadvertent changes in the gene expression profiles, this seems to have been avoided by comparison of the experimental condition with a control that had been treated similarly (Smulski *et al*., 2001; Zheng *et al*., 2001). Still, it is possible that culture handling might induce stress response genes in both conditions. Sampling from the same culture before and after the treatment would serve to identify expression artifacts of this kind (Arnold *et al*., 2001; Wiegert *et al*., 2001).

Comparison of mutant strains with the wild type is a powerful tool for the elucidation of regulatory networks. Gene expression profiling of *E. coli* mutants defective in the global, pleiotropic regulatory proteins, IHF (Arfin *et al*., 2000) and HN-S (Hommais *et al*., 2001), identified genes that respond either positively or negatively to the absence of these transcription factors. In both these comparisons of mutant and wild-type logarithmic phase cultures, 3–5% of the genes on the genome were differentially regulated, yet substantial growth defects were not identified. Not only were the bacteria able to tolerate a significant perturbation of gene expression, the absence of an effect on growth would seem to simplify data analysis by reducing the significance of growth-related changes. However, neither the use of transcription factor mutants nor the absence of an effect of these mutations on growth excludes the possibility that indirect, downstream effects might contribute to the expression profile. Further analysis of this issue may be warranted. In cases in which the transcription factor binding site has been identified, correlation between the differential expression of target genes and the occurrence of the appropriate binding site is a reasonable strategy for dissecting direct from indirect effects (Arfin *et al*., 2000).

A simple comparison of steady-state mutant and wild-type cultures is not a reasonable design for investigating all regulatory factors. For example, with inducible stress response networks, the loss of the regulatory factor would cause a failure to respond to the stress signal. Here, comparison of treated versus untreated wild-type cells reveals the normal global response to the treatment, whereas the same experiment run in the regulatory mutant identifies target genes in specific regulons. This strategy has been used very successfully to identify genes under OxyR control during hydrogen peroxide stress (Zheng *et al*., 2001) and genes of the σ^W regulon that respond to alkaline shock (Wiegert *et al*., 2001). The value of both publications was enhanced by incorporating molecular analysis of the regulatory sites to confirm the participation of the regulatory factors at the promoters of newly identified targets. An alternative, and possibly complementary, experimental strategy would use microarray transcript profiling to identify genes regulated by investigator-initiated expression of a transcription factor gene from an inducible promoter.

Each variation of the two-condition experiment generates relatively long lists of genes that respond either directly or indirectly to the experimental variable. Aside from the fact that long lists are boring, there are dangers

inherent in the two-condition experimental strategy that must be considered during analysis. Concern that the list could contain 'false positives' is minimized by proper experimental replication and statistical analysis (Arfin *et al*., 2000; Tao *et al*., 2001; Conway *et al*., 2002). This concern is further avoided when the list is placed in an appropriate biological context by inclusion of corroborating physiological, biochemical and genetic experiments.

A very real and underappreciated danger of gene expression profiling can result from a phenomenon known as the 'small-world' network (Watts and Strogatz, 1998). Wherever there is a high degree of communication between members of small regulatory networks and a small number of interconnections between members of different regulatory networks, a small-world network is created where seemingly unrelated regulatory networks respond to the same experimental variable. A real-world example was uncovered during the systematic perturbation of genes encoding enzymes and transcription factors of the yeast galactose utilization pathway. These mutations were expected to impact Gal gene expression, but also had a number of unanticipated effects on genes whose participation in galactose metabolism is unclear (Ideker *et al*., 2001). The solution to the small-world network effect seems to lie in integrating experimental approaches and cluster analysis of multiple experimental variables.

Cluster analysis of several different strains and growth conditions involving tryptophan metabolism (Khodursky *et al*., 2000) and nitrogen limitation (Zimmer *et al*., 2000) provided new insights into the underlying regulatory networks. The latter paper thoughtfully used a set of selection criteria for assigning nitrogen limitation-induced genes to regulatory classes. Time course analysis of a strain in which the native *E. coli* mixed acid fermentation pathway had been replaced by one for ethanol production explained why the engineered strain could grow more rapidly than the wild type (Tao *et al*., 2001). The time course of induction of the SOS response after exposure to UV light was measured and, in a complementary experiment, the LexA regulon was characterized in a *lexA* mutant (Courcelle *et al*., 2001). The latter example revealed differences in induction kinetics for various UV-induced genes. Time courses with sampling intervals that pace the dynamic aspects of transcription and mRNA turnover rates, while considering the timing of the genetic response, can be used to distinguish direct from indirect effects. We used this strategy to dissect the numerous systems either induced or repressed during growth arrest caused by glucose–lactose diauxie and after exposure to a growth-inhibitory dose of hydrogen peroxide (Chang *et al*., 2002). Cluster analysis of these data sets allowed us to distinguish condition specific genetic responses in these two experiments and revealed those systems that are fundamental to the genetic programme underlying the physiological adaptation to growth arrest. Moreover, the return to steady-state gene expression in cells during recovery from growth arrest emphasized the principle that steady-state growth is accompanied by steady-state gene expression.

Papers that blend array technology and more traditional life science research methodologies lend confidence to biological interpretation. Proper replication of array experiments maximizes the statistical significance of the data and, hence, the quality of the analysis. Multiple array data sets lend themselves to cluster analysis and other analysis strategies that illuminate the system response to experimental variables while minimizing false positives, small-world networks and indirect effects. Robust array data sets reveal biological meaning, as indicated above, by consideration of the laboratory experiments. Armed with this new functional genomics technology and a working knowledge of its application, we can proceed to the question – what are bacteria really doing in their native environments? Below, we explore this theme by examining how microarray expression analysis can be used in the study of bacterial pathogenesis.

Functional genomics and the study of bacterial pathogenesis

Capturing the in vivo *transcriptome: opportunities and impediments*

The focus of most current pathogenesis research is the identity, functional properties and regulation of individual virulence determinants. In contrast, the ultimate goal of genome-wide expression studies is to measure, on a whole-genome scale, the host-adapted transcriptional responses of infecting bacteria. From the genome-scale perspective, *in vivo* growth will be shown to entail a significant reprogramming of the transcriptome, compared with exponentially growing planktonic cultures in rich laboratory media. Beyond the induction of genes encoding virulence determinants, the whole-genome expression profile of tissue-grown bacteria will encompass other response classes that are likely to be essential for *in vivo* survival, but do not injure host tissues or subvert, inactivate or adapt the organism to host defences, and thus are not virulence determinants *per se*. These we designate 'essential *in vivo* genes' to indicate those that are likely to encode functions required for growth *in vivo*, although they may be dispensable for *in vitro* growth. Among these are genes that: encode metabolic pathways for the scavenging of nutrients and essential growth factors from the host; mediate the surface mode-of-growth (including biofilm formation) of bacteria attached to mucous membranes and other surfaces; or respond to the presence of other

microbes via cell density-dependent regulation, particularly in the crowded and multispecies niches of the respiratory, gastrointestinal and urogenital tracts. These adaptations will, in turn, dictate *in vivo* growth rates that will be evident in the gene expression profile and, when correctly interpreted, could disclose for the first time the physiological state of bacteria *in vivo* in a site- and stage-specific manner.

Technical issues currently impede *in vivo* transcriptional profiling experiments and, as a result, only two of the 24 pathogenicity studies listed in Table 1 provide expression results from *in vivo* grown organisms. In one of these, the authors cleverly inoculated dialysis chambers implanted in the rat peritoneal cavity with *Borrelia burgdorferi*, the causative agent of Lyme disease, to achieve an *in vivo* growth environment free of host cells (Revel *et al*., 2002). In the other experiment, investigators took advantage of two features of the rice water stools of cholera patients to obtain *in vivo* expression profiles of *Vibrio cholerae* (Merrell *et al*., 2002). First, such stools generally contain few host cells or commensal faecal flora. Secondly, their transit time from the site of infection in the small bowel to the waiting receptacle is brief, and thus their transcriptome in the collected samples probably reflects their adaptation to the small intestine rather than to more distal sites of the gastrointestinal tract. However, the transcriptome of bacteria shed in stools could differ from those still attached to host epithelial cells.

The use of samples from *in vivo* sites that lack host cells and commensal flora provides an enormous technical advantage for microarray expression profiling purposes, because samples containing RNA from the pathogen can be prepared that are not mixed with RNA from the host or other bacterial species. We anticipate that other microarray studies of this kind will be conducted, for example with bacterially infected cerebrospinal fluid (meningitis) and

Table 1. Microarray studies of pathogenic bacteria.

Species	Experimental classification[a]	Growth condition	Transcription factor or developmental programme	Reference
E. coli K-12, MG1655	A, C	Oxidative stress, H_2O_2	*oxyR*	Zheng *et al*. (2001)
E. coli O157:H7	A	Acetate-induced acid tolerance	NA	Arnold *et al*. (2001)
H. pylori NTU-D1	A	Acid-induced gene expression	NA	Ang *et al*. (2001)
Pasteurella multocida	A	Iron limitation	NA	Paustian *et al*. (2001)
M. tuberculosis	A, C	Iron limitation	IdeR	Rodriguez *et al*. (2001)
M. tuberculosis	A	Low O_2	NA	Sherman *et al*. (2001)
M. tuberculosis	A	Nutrient limitation	NA	Betts *et al*. (2002)
M. tuberculosis	A	Acid pH	NA	Fisher *et al*. (2002)
S. pyogenes	A	Temperature-regulated genes	NA	Smoot *et al*. (2001)
Borrelia burgdorferi	B	Rat, peritoneal cavity (implanted dialysis chambers)	NA	Revel *et al*. (2002)
V. cholerae O1, El Tor	B	Human cholera stools	NA	Merrell *et al*. (2002)
M. tuberculosis	A, C	Stress response: SDS	σ^E	Manganelli *et al*. (2001)
M. tuberculosis	A, C	Thiol-specific oxidative stress (diamide)	σ^H	Manganelli *et al*. (2002)
M. tuberculosis	A, C	Growth phase-dependent σ^H-dependent genes	σ^H	Kaushal *et al*. (2002)
S. aureus	A, C	Cell density-dependent gene regulation	*arg* and/or SarA gene expression	Dunman *et al*. (2001)
P. syringae pv. tomato DC3000	C	AB medium	HrpL alternative sigma factor	Fouts *et al*. (2002)
V. cholerae O1, El Tor	A, C	AKI medium	*luxO*	Zhu *et al*. (2002)
S. pneumoniae	D	Induction of competence by CSP	NA	Rimini *et al*. (2000)
S. pneumoniae	D	Response to autoinducer peptide		De Saizieu *et al*. (2000)
E. coli O157:H7	D	Cell density-dependent gene regulation, genes induced by autoinducer II		Sperandio *et al*. (2001)
E. coli K-12 (*luxS*::Tc^R)	D	AI-2 regulated genes		DeLisa *et al*. (2001)
M. tuberculosis	E	Genes induced by isoniazide		Wilson *et al*. (1999)
P. aeruginosa	E	Biofilm-induced antibiotic resistance		Whiteley *et al*. (2001)

a. Five experimental classifications are denoted by A, B, C, D and/or E:
A, RNA obtained from bacteria grown *in vitro* under conditions intended to simulate microenvironments of the host; B, RNA obtained from bacteria in a clinical sample from an infected person or animal model; C, Regulon membership of a transcription factor deduced from the comparison of expression profiles from the wild-type parent and a mutant that either does not express the transcription factor or overexpresses the transcription factor; grown under conditions that activate the transcription factor in question; D, Factor-induced developmental programmes characterized by a temporally programmed physiological process; E, Antibiotic-induced expression profiles disclose novel drug targets, compound mode-of-action or mechanisms of resistance.

urine (cystitis and pyelonephritis). However, for many infection syndromes, the relatively small numbers of the infectious agent within infected tissues and the presence of host cells and normal flora will preclude expression profiling without the use of genome-specific primers to prepare and/or amplify bacterial cDNA (Talaat *et al*., 2000) or the development of methods that separate bacterial mRNA from host RNA.

Unlike many of the exquisitely designed *in vitro* experiments described in the preceding sections, which sought to control all variables except the one under study, the expression profile of an *in vivo* expression experiment is the composite of many contemporaneous responses that reflect the physico-chemical features of the *in vivo* site of infection. This naturally raises the question of how can the multifaceted *in vivo* expression profile – and, for that matter, multivariable *in vitro* experiments – be deconstructed into subsets comprising each of the individual responses? The answer would seem to come from a combination of physiological intuition, the use of bioinformatic tools and empirical investigations. The last of these will include *in vitro* expression profiling experiments aimed at discovering the signature profiles of the organism to individual conditions of growth that are hypothesized to prevail *in vivo*. To date, there is no such example of a comprehensive study of this kind, but the studies published thus far, and summarized in Table 1 suggest the following experimental strategy.

Based on the annotated functions of the differentially expressed genes, the *in vivo* expression profile is scrutinized by an experienced bacterial physiologist to identify the physiological and metabolic features of the response. Here, it is reasonable to assume that the functional unit of the response is not the collection of genes that comprise the expression profile, but rather the biochemical pathways that these genes encode. Accordingly, from this pathway perspective, deconstruction of the *in vivo* expression profile can be facilitated by superimposing expression data onto species-specific pathway databases. In this manner, the regulated pathways are highlighted against a background of other pathways, the expression of which is not differentially regulated *in vivo* (Karp *et al*., 1999).

Further information comes from correlating these hypothesized physiological adaptations with the annotated functions of differentially regulated genes that reside in clusters on the genome. Clusters of this kind frequently specify mixed catabolic, transport and regulatory functions for the same metabolic pathway. Accordingly, the possible role of genes in the cluster that lack an annotated function is disclosed by reference to the company that they keep.

Informed by a physiological perspective of the microbe, the analytical process discussed above naturally leads to the idea that the organism's adaptation to the host is the sum of independent responses to each of several features of the host microenvironment. Therefore, to 'deconstruct' the *in vivo* expression profile further, it is logical to perform condition-specific microarray expression experiments *in vitro*. For example, analysis of an *in vivo* expression profile might suggest that the microbe has adapted to a surface mode-of-growth in an acidic, microaerophilic, low-iron environment, within which it uses fatty acids for carbon and energy and generates reductants to defend its internal redox state. These suppositions can then be tested by performing *in vitro* microarray experiments to determine whether each of the condition-specific profiles can be located within the *in vivo* expression profile. In this example, expression profiles would be obtained from bacteria grown as a biofilm on plastic or glass and from planktonic bacteria grown in a variety of media devised to mimic each of the hypothesized conditions: acidic pH, low O_2, iron-depleted medium, a medium containing fatty acids rather than glucose or glycerol as the sole carbon source and growth during exposure to oxidants. Wherever possible, such experiments should be performed in a manner that does not affect the growth rate or otherwise result in pleiotropic effects. The accumulation of many *in vitro* expression profiles of this kind is not only important in the expression profile deconstruction process, but also provides a valuable archive of condition-specific expression signatures. Once these signatures are recognized within the *in vivo* expression profile, their presence can be used to deduce the physico-chemical features of a host compartment – from the perspective of the microbe. In this manner, expression signatures can serve as bioprobes of the host microenvironment. For example, if an *in vivo* expression profile is found to contain a gene set induced during *in vitro* growth in iron-depleted medium, then it is logical to conclude that iron availability is limited in the corresponding host microenvironment.

From a bioinformatics perspective, the availability of an archive of condition-specific expression signatures provides an additional dividend. The accumulation of expression profiles from a large number of separate conditions facilitates the use of clustering algorithms, the power of which to reveal co-ordinately regulated genes increases with the number of different conditions clustered (Eisen *et al*., 1998). It is obvious to us that a large number of laboratories will contribute gene expression profile data to the community effort of understanding pathogenesis. Thus, there is a need for microarray standards and centralized databases, several of which were reviewed recently (Rhodius *et al*., 2002). One published standard and guideline was developed by the Microarray Gene Expression Data (MGED) group and is available at http://www.mged.org/Workgroups/MIAME/miame_checklist.html. We note, however, that these databases

have to date received a minimal number of submissions and have yet to gain wide acceptance.

Tricking the pathogen to disclose its in vivo *transcriptome during* in vitro *growth: expression profiling studies in simulated microenvironments of the host*

Implicit in the foregoing discussion is the assumption that the *in vivo* transcriptome, as a whole, is merely the sum of its parts and that each of the individual responses that compose it can occur independently of the others in a condition-specific manner. However, the pathogenesis field is replete with examples in which exposure to a simple chemical cue, often produced by modifications of standard culture conditions, elicits the co-ordinate regulation of genes that not only encode virulence determinants but, in some cases, metabolic pathways as well (Mekalanos, 1992). Thus, it appears that an organism does not need to sense all features of a host microenvironment to mount a multifaceted adaptive response. We can understand this phenomenon as the consequence of an evolutionary process during which fitness for a particular habitat became keyed to a characteristic and faithful physico-chemical feature of the habitat. This feature then serves as a sentinel cue that acts via 'hard-wired' regulatory circuits to orchestrate the organism's adaptive response co-ordinately to this and other features of the environment. These two related concepts – regulatory networks and cue-driven induction of co-ordinately regulated genes – have enormous significance for the design of microarray expression studies of pathogenesis: exposure of the microbe during *in vitro* growth to a sentinel cue activates specific virulence determinants and adapts the organism for *in vivo* growth; these can be captured in a single microarray expression experiment. Of the 24 microarray pathogenesis experiments summarized in Table 1, 14 have used this experimental strategy and are designated '*in vitro* conditions intended to simulate microenvironments of the host'. The chemical cues that were tested in these studies include nutrient privation, oxidative stress, temperature shifts, acidic pH, hypoxia, iron limitation and detergent-mediated cell envelope stress.

Deducing regulatory circuits from condition-specific expression profiles

The analysis of condition-specific expression profiles often betrays the cognate regulator because genes encoding transcription factors are frequently contiguous to the genes that they regulate. Additionally, genes encoding transcription factors and the genes they govern are often co-regulated at the transcriptional level by the same cue and will thus be clustered together in the expression profile. This is true even for many transcription factors that are mainly activated by post-translational events. Where such studies suggest the identity of the cognate regulator, a logical next step is its mutational inactivation. A comparison of the expression profiles of the wild type and regulatory mutant will lead to the identification of cue-specific genes that require the regulator for their expression. Further analysis using bioinformatics tools, microarray-based, *in vivo* binding site analysis (also known as genome-wide location analysis) (Laub *et al*., 2002) and gel shift assays will disclose a subset of these genes that has a conserved promoter motif and binds the cognate regulator. In this manner, the more comprehensive studies in this group have used microarray expression profiling to define the transcriptional response to an environmental mimetic of the host microenvironment, to identify the cognate regulator, to define regulon membership and to identify a subset of regulon membership genes that the transcription factor regulates directly. Studies denoted by 'C' in Table 1 illustrate this experimental paradigm.

Transcriptional portraits of pathogenic developmental programmes

Some aspects of pathogenesis can be portrayed as a developmental programme that entails a temporally ordered series of transcriptional responses, which affect stage-specific physiological and morphological adaptations. Examples of this for medically important bacteria include sporulation by *Bacillus anthracis*, the intracellular developmental stages of *Chlamydia trachomatis* and the shifts that occur as *B. burgdorferi* and *Yersinia pestis* move between their arthropod vectors and mammalian hosts. Although the potential value of such studies is great, with the exception of *B. burgdorferi* (Revel *et al*., 2002), few studies of this kind with microbial pathogens have been published. Microbial pheromone or signalling molecules also initiate developmental programmes. Included in this group are the development of competence by *Streptococcus pneumoniae* (Rimini *et al*., 2000) and the cell density-dependent quorum-sensing response by many pathogenic species (Miller and Bassler, 2001). After the addition of a signalling molecule to a bacterial culture, the most informative experiments of this kind have monitored the transcriptional response during a time course that parallels the morphological and physiological changes that characterize the developmental programme. Alternatively, where a cell density responsive regulator is known or inferred, the regulator-dependent quorum-sensing response has been defined using microarray expression profiling to compare wild type and regulator mutants grown under conditions that cause the accumulation of quorum-sensing molecules. A particularly informative study of this kind defined the LuxO regulon and

the position of this transcription factor in a regulatory hierarchy controlling virulence gene expression in *V. cholerae* (Zhu *et al*., 2002). Biofilm formation, including transitions between the planktonic and surface-attached populations, probably entails a complex developmental programme involving surface sensing and attachment, horizontal surface spread and vertical growth of the consortium with the formation of the architectural features that typify mature biofilms (O'Toole *et al*., 2000). Microarray expression profiling studies of this developmental process provide an exciting opportunity to compare the stage-specific transcriptome with stage-specific images of biofilms obtained by scanning confocal laser microscopy.

The study of developmental programmes by microarray expression profiling raises special technical issues not encountered in the study of homogeneous suspensions of planktonic cultures. As microarray methods entail the collection of RNA from all individuals in a sample, the expression results represent the average value of each transcript species in the RNA extract. However, because the essence of developmental programmes is change across temporal and/or spatial dimensions, the heterogeneity of developmental stages in non-synchronized cultures is unavoidable. Estimating the extent of heterogeneity and quantifying the presence, size and functional significance of minority populations are therefore important in the interpretation of microarray expression data in this experimental category. In the case of developmental programmes that are mainly spatial in nature, it may be necessary to use reporter constructs, such as green fluorescent protein fusions, which visualize the transcriptional state of genes of interest while allowing the protein fusion product to be imaged in three-dimensional space at the single-cell level of resolution in a real-time and non-destructive manner. Although such methods have been developed and used for the study of biofilms (Geesey, 2001), in addition to other applications, they have never been used to refine microarray expression data. Ultimately, a detailed picture of the molecular processes that describes how a bacterium interacts with its host or environment will require experimental designs that incorporate the best of traditional research methodologies with the emerging genome-based technologies.

Acknowledgements

The authors gratefully acknowledge support for research in our laboratories from the NIH, RO1-AI48945-01 (to T.C.) and RO1-AI44826 (to G.S.). T.C. wishes to thank Darren Smalley and Frank Wu for assistance with literature, and April Anderson for critical reading of the manuscript.

References

Ang, S., Lee, C.-Z., Peck, K., Sindici, M., Matrubutham, U., Gleeson, M.A., *et al.* (2001) Acid-induced gene expression *Helicobacter pylori*; study in genomic scale by microarray. *Infect Immun* **69:** 1679–1686.

Arfin, S.M., Long, A.D., Ito, E.T., Tolleri, L., Riehle, M.M., Paegle, E.S., and Hatfield, G.W. (2000) Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J Biol Chem* **275:** 29672–29684.

Arnold, C.N., McElhanon, J., Lee, A., Leonhart, R., and Siegele, D.A. (2001) Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response. *J Bacteriol* **183:** 2178–2186.

Baldwin, D., Crane, V., and Rice, D. (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr Opin Plant Biol* **2:** 96–103.

Baracchini, E., and Bremer, H. (1987) Determination of synthesis rate and lifetime of bacterial mRNAs. *Anal Biochem* **167:** 245–260.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99:** 9697–9702.

Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., and Duncan, K. (2002) Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* **43:** 717–731.

Caldwell, R., Sapolsky, R., Weyler, W., Maile, R.R., Causey, S.C., and Ferrari, E. (2001) Correlation between *Bacillus subtilis scoC* phenotype and gene expression determined using microarrays for transcriptome analysis. *J Bacteriol* **183:** 7329–7340.

Chang, D., Smalley, D.J., and Conway, T. (2002) Gene expression profiling of *E. coli* growth transitions: an expanded stringent response model. *Mol Microbiol* **45:** 289–306.

Conway, T., Kraus, B., Tucker, D.L., Smalley, D.J., Dorman, A.F., and McKibben, L. (2002) DNA array analysis in a Microsoft Windows environment. *Biotechniques* **32:** 110, 112–110,114–116, 118–119.

Courcelle, J., Khodursky, A., Peter, B., Brown, P.O., and Hanawalt, P.C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* **158:** 41–64.

Cummings, C.A., and Relman, D.A. (2000) Using DNA microarrays to study host–microbe interactions. *Emerg Infect Dis* **6:** 513–525.

DeLisa, M.P., Wu, C.F., Wang, L., Valdes, J.J., and Bentley, W.E. (2001) DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*. *J Bacteriol* **183:** 5239–5247.

De Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., *et al.* (2000) Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J Bacteriol* **182:** 4696–4703.

Dudley, A.M., Aach, J., Steffen, M.A., and Church, G.M. (2002) Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc Natl Acad Sci USA* **99:** 7554–7559.

Dunman, P.M., Murphy, E., Haney, S., Placios, D.,

Tucker-Kellogg, G., Wu, S., *et al.* (2001) Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci. *J Bacteriol* **183:** 7341–7353.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95:** 14863–14868.

Eymann, C., Homuth, G., Scharf, C., and Hecker, M. (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *J Bacteriol* **184:** 2500–2520.

Fisher, M.A., Plikaytis, B.B., and Shinnick, T.M. (2002) Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol* **184:** 4025–4032.

Fouts, D.E., Abramovitch, R.B., Alfano, J.R., Baldo, A.M., Buell, C.R., Cartinhour, S., *et al.* (2002) Genomewide identification of *Pseudomonas syringae* pv. tomato DC3000 promoters controlled by the HrpL alternative sigma factor. *Proc Natl Acad Sci USA* **99:** 2275–2280.

Geesey, G.G. (2001) Bacterial behavior at surfaces. *Curr Opin Microbiol* **4:** 296–300.

Helmann, J.D., Wu, M.F., Kobel, P.A., Gamo, F.J., Wilson, M., Morshedi, M.M., *et al.* (2001) Global transcriptional response of *Bacillus subtilis* to heat shock. *J Bacteriol* **183:** 7318–7328.

Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J.P., *et al.* (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol Microbiol* **40:** 20–36.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292:** 929–934.

Karp, P.D., Krummenacker, M., Paley, S., and Wagg, J. (1999) Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol* **17:** 275–281.

Kaushal, D., Schroeder, B.G., Tyagi, S., Yoshimatsu, T., Scott, C., Ko, C., *et al.* (2002) Reduced immunopathology and mortality despite tissue persistence in a *Mycobacterium tuberculosis* mutant lacking alternative sigma factor, SigH. *Proc Natl Acad Sci USA* **99:** 8330–8335.

Khil, P.P., and Camerini-Otero, R.D. (2002) Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. *Mol Microbiol* **44:** 89–105.

Khodursky, A.B., Peter, B.J., Cozzarelli, N.R., Botstein, D., Brown, P.O., and Yanofsky, C. (2000) 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* **97:** 12170–12175.

Kushner, S.R. (1996) mRNA decay. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., Curtiss, R., III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., *et al.* (eds). Washington, DC: American Society for Microbiology Press, pp. 849–860.

Kushner, S.R. (2002) mRNA decay in *Escherichia coli* comes of age. *J Bacteriol* **184:** 4658–4665.

Laub, M.T., McAdams, H.H., Feldblyum, T., Fraser, C.M., and Shapiro, L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290:** 2144–2148.

Laub, M.T., Chen, S.L., Shapiro, L., and McAdams, H.H. (2002) Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci USA* **99:** 4632–4637.

Lee, J.M., Zhang, S., Saha, S., Santa, A.S., Jiang, C., and Perkins, J. (2001) RNA expression analysis using an antisense *Bacillus subtilis* genome array. *J Bacteriol* **183:** 7371–7380.

Majdalani, N., Chen, S., Murrow, J., St John, K., and Gottesman, S. (2001) Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol Microbiol* **39:** 1382–1394.

Manganelli, R., Voskuil, M.I., Schoolnik, G.K., and Smith, I. (2001) The *Mycobacterium tuberculosis* ECF sigma factor sigma E: role in global gene expression and survival in macrophages. *Mol Microbiol* **41:** 423–437.

Manganelli, R., Voskuil, M.I., Schoolnik, G.K., Dubnau, E., Gomez, M., and Smith, I. (2002) Role of the extracytoplasmic-function sigma factor sigma H in *Mycobacterium tuberculosis* global gene expression. *Mol Microbiol* **45:** 365–374.

Mekalanos, J.J. (1992) Environmental signals controlling the expression of virulence determinants in bacteria. *J Bacteriol* **174:** 1–7.

Merrell, D.S., Butler, S.M., Qadri, F., Dolganov, N.A., Alam, A., Cohen, M.B., *et al.* (2002) Host-induced epidemic spread of the cholera bacterium. *Nature* **417:** 642–645.

Miller, M.B., and Bassler, B.L. (2001) Quorum sensing bacteria. *Annu Rev Microbiol* **55:** 165–199.

Oh, M.K., Rohlin, L., Kao, K.C., and Liao, J.C. (2002) Global expression profiling of acetate-grown *Escherichia coli*. *J Biol Chem* **277:** 13175–13183.

O'Toole, G., Kaplan, H.B., and Kolter, R. (2000) Biofilm formation as microbial development. *Annu Rev Microbiol* **54:** 49–79.

Paustian, M.L., May, B.J., and Kapur, V. (2001) *Pasteurella multocida* gene expression in response to iron limitation. *Infect Immun* **69:** 4109–4115.

Pedersen, S., and Reeh, S. (1978) Functional mRNA half lives in *E. coli*. *Mol Gen Genet* **166:** 329–336.

Pomposiello, P.J., Bennik, M.H., and Demple, B. (2001) Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. *J Bacteriol* **183:** 3890–3902.

Record, M.T., Reznikoff, W.S., Craig, M.L., McQuade, K.L., and Schlax, P.J. (1996) *Escherichia coli* RNA polymerase (Eσ70), promoters, and the kinetics of the steps of transcription initiation. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., Curtiss, R., III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., *et al.* (eds). Washington, DC: American Society for Microbiology Press, pp. 792–821.

Revel, A.T., Talaat, A.M., and Norgard, M.V. (2002) DNA microarray analysis of differential gene expression in *Borrelia burgdorferi*, the Lyme disease spirochete. *Proc Natl Acad Sci USA* **99:** 1562–1567.

Rhodius, V., Van Dyk, T.K., Gross, C., and LaRossa, R.A. (2002) Impact of genomic technologies on studies of bacterial gene expression. *Annu Rev Microbiol* **56:** 599–624.

Richardson, J.P., and Greenblatt, J. (1996) Control of RNA chain elongation and termination. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., Curtiss, R., III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., *et al.* (eds). Washington, DC: American Society for Microbiology Press, pp. 822–848.

Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., and Blattner, F.R. (1999) Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* **27:** 3821–3835.

Rimini, R., Jansson, B., Feger, G., Roberts, T.C., de Francesco, M., Gozzi, A., *et al.* (2000) Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Mol Microbiol* **36:** 1279–1292.

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., and Smith, I. (2002) *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* **70:** 3371–3381.

Schut, G.J., Zhou, J., and Adams, M.W. (2001) DNA microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for a new type of sulfur-reducing enzyme complex. *J Bacteriol* **183:** 7027–7036.

Sekowska, A., Robin, S., Daudin, J.J., Henaut, A., and Danchin, A. (2001) Extracting biological information from DNA arrays: an unexpected link between arginine and methionine metabolism in *Bacillus subtilis*. *Genome Biol Res* **2:** 0019.1–0019.12.

Sherman, D.R., Voskuil, M., Schnappinger, P., Liao, R.L., Harrell, M.I., and Schoolnik, G.K. (2001) Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc Natl Acad Sci USA* **98:** 7534–7539.

Smoot, L.M., Smoot, J.C., Graham, M.R., Somerville, G.A., Sturdevant, D.E., Migliaccio, C.A., *et al.* (2001) Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. *Proc Natl Acad Sci USA* **98:** 10416–10421.

Smulski, D.R., Huang, L.L., McCluskey, M.P., Reeve, M.J., Vollmer, A.C., Van Dyk, T.K., and LaRossa, R.A. (2001) Combined, functional genomic-biochemical approach to intermediary metabolism: interaction of acivicin, a glutamine amidotransferase inhibitor, with *Escherichia coli* K-12. *J Bacteriol* **183:** 3353–3364.

Sperandio, V., Torres, A.G., Giron, J.A., and Kaper, J.B. (2001) Quorum sensing is a global regulatory mechanism in enterohemorrhagic *Escherichia coli* O157:H7. *J Bacteriol* **183:** 5187–5197.

Talaat, A.M., Hunter, P., and Johnston, S.A. (2000) Genome-directed primers for selective labeling of bacterial transcripts for DNA microanalysis. *Nature Biotechnol* **18:** 679–682.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181:** 6425–6440.

Tao, H., Gonzalez, R., Martinez, A., Rodriguez, M., Ingram, L.O., Preston, J.F., and Shanmugam, K.T. (2001) Engineering a homo-ethanol pathway in *Escherichia coli*: increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation. *J Bacteriol* **183:** 2979–2988.

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **15:** 1637–1651.

Watts, D.J., and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature* **393:** 440–442.

Wei, Y., Lee, J.M., Richmond, C., Blattner, F.R., Rafalski, J.A., and LaRossa, R.A. (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol* **183:** 545–556.

Whiteley, M., Bangera, M.G., Bumgarner, R.E., Parsek, M.R., Teitzel, G.M., Lory, S., *et al.* (2001) Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413:** 860–864.

Wiegert, T., Homuth, G., Versteeg, S., and Schumann, W. (2001) Alkaline shock induces the *Bacillus subtilis* sigma (W) regulon. *Mol Microbiol* **41:** 59–71.

Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O., and Schoolnik, G.K. (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* **96:** 12833–12838.

Winzeler, E.A., Schena, M., and Davis, R.W. (1999) Fluorescence-based expression monitoring using microarrays. *Methods Enzymol* **306:** 3–18.

Ye, R.W., Tao, W., Bedzyk, L., Young, T., Chen, M., and Li, L. (2000) Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J Bacteriol* **182:** 4458–4465.

Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C.M., Matsunaga, M., Yamaguchi, H., *et al.* (2001) Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. *Nucleic Acids Res* **29:** 683–692.

Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C., and Storz, G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell* **9:** 11–22.

Zheng, M., Wang, X., Templeton, L.J., Smulski, D.R., LaRossa, R.A., and Storz, G. (2001) DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J Bacteriol* **183:** 4562–4570.

Zhu, J., Miller, M.B., Vance, R.E., Dziejman, M., Bassler, B.L., and Mekalanos, J.J. (2002) Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc Natl Acad Sci USA* **99:** 3129–3134.

Zimmer, D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, A.B., Peter, B.J., *et al.* (2000) Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci USA* **97:** 14674–14679.