CURRENT OPINION
www.current-opinion.com

Uses and pitfalls of microarrays for studying transcriptional regulation

Virgil A Rhodius* and Robert A LaRossa†

Microarrays provide a powerful new tool for understanding the regulation of gene expression in bacteria. Many recent publications have used microarrays for identifying regulon members and stimulons that describe the complex organismal responses to environmental perturbations. The use of bioinformatics to identify DNA binding sites of transcription factors greatly facilitates the interpretation of these experiments. Understanding the transcriptome of an organism includes identifying all transcripts and mapping their 5′ and 3′ ends. High-density oligonucleotide arrays have enabled the identification of many new transcripts, including small RNAs and antisense RNAs.

Addresses
*Department of Microbiology and Immunology, University of California, Genentech Hall, 600 16th Street, San Francisco, CA 94143-2200, USA
e-mail: rhodius@itsa.ucsf.edu
†DuPont Company, Central Research and Development, Wilmington, DE 19880-0173, USA
e-mail: robert.a.larossa@usa.dupont.com

Current Opinion in Microbiology 2003, **6**:114–119

This review comes from a themed issue on
Cell regulation
Edited by Andrée Lazdunski and Carol Gross

1369-5274/03/$ – see front matter

DOI 10.1016/S1369-5274(03)00034-1

Abbreviations

ECF	extracytoplasmic function
ORF	open reading frame
sRNA	small RNA
UTR	untranslated region

Introduction

Microarrays are increasingly being used to study bacterial gene expression on a global scale. In discussing comprehensive approaches to gene expression profiling the concepts of regulons and stimulons are important. A regulon is a set of transcription units controlled by a single regulatory protein. Regulons can overlap when the upstream regulatory regions are composites of several binding sites for different regulators. Array experiments that define a regulon provide a group of candidate transcription units, of which the upstream regulatory regions are then analyzed to define potential binding sites for the global regulator. Stimulons are groups of transcription units that are differentially expressed in response to environmental perturbation. The study of such functional units provides a complementary perspective, allowing one to see that seemingly unrelated activities are sometimes modulated together. An early example is the stringent response to amino acid starvation, which shuts off transcription of ribosomal and tRNA genes, but also enhances expression of several amino acid biosynthetic operons. Following the extensive study of regulons and stimulons, results from microarrays must be integrated with existing physiological knowledge, providing confirmation, or occasionally surprising new insights. Alternatively, microarray experiments might explore virgin territory and could very well govern further physiological analysis. Here, we review recent reports concentrating on microarray analysis of regulons and stimulons, and on identifying the many different types of transcripts that form the transcriptome of an organism.

Identifying regulons

Genome-wide expression analysis provides a powerful way of identifying regulon members. Over 20 publications have aimed to identify the regulons controlled by 46 *Escherichia coli* and 34 *Bacillus subtilis* transcription factors (Table 1). The standard protocol is to compare the expression profile of strains carrying an active form of the regulator with strains that carry a less active or inactive form. Typically, one compares either the wild type or a strain carrying a null mutation of the regulator with either a wild-type strain grown under inducing conditions [1,2••,3,4,5•,6–10,11•,12,13,14••,15••], an overproducing strain [7,16–18], or a strain carrying a mutation that locks the regulator into the 'on' conformation [19,20]. When appropriate environmental cues are unknown, it is important to perform the expression analysis on cells grown under a variety of conditions.

Experiments are viewed as 'steady state' when the transcription factor is constitutively active in one of the strains. In this situation, genes expressed by the transcription factor might affect the expression of other genes (e.g. other regulators) resulting in cascades that obscure regulatory circuitry. To circumvent this, a time-course monitoring global expression patterns may be attempted, after the level/activity of the transcription factor is induced [2••,14••,21]. Primary target genes for the transcription factor might be those whose expression changes first, whereas those that are indirectly affected would be modified later. However, when induction results in regulatory concentrations exceeding normal titers, the regulator can occupy weak and physiologically irrelevant sites, or related sites normally bound by another transcription factor. It may even oligomerize at some chromosomal

Table 1

Regulons studied with microarrays.

Factor	Function	Organism	References
ComA	Two-component regulator; activated by cell density	*Bacillus subtilis*	[18]
ComK	Development of competence in stationary phase	*Bacillus subtilis*	[6,10]
DegU	Two-component regulator; exoprotease production, competence, motility	*Bacillus subtilis*	[18]
EtrA	Fnr homolog; regulates energy metabolism, transcription factors, biosynthesis	*Shewanella oneidensis*	[9]
EvgA	Two-component regulator; involved in acid and multidrug resistance	*Escherichia coli*	[7]
HrpL	Alternative sigma factor; regulates virulence type III protein secretion genes	*Pseudomonas syringae*	[43,44]
IHF	Global regulator	*Escherichia coli*	[12]
LexA	Regulator of the SOS response	*Escherichia coli*	[20]
LrhA	Regulator of flagella, motility and chemotaxis genes	*Escherichia coli*	[8]
Lrp	Leucine-responsive regulatory protein; regulation of stationary phase transitions	*Escherichia coli*	[3,4]
MarA	Oxidative stress and multiple antibiotic resistance	*Escherichia coli*	[21]
Nac	Controls assimilation of nitrogen; regulator of σ^{70}-dependent genes	*Escherichia coli*	[19]
NtrC	Growth under nitrogen limiting conditions; regulator of σ^{54}-dependent genes	*Escherichia coli*	[19]
PhoP	Two-component regulator; induced by phosphate starvation	*Bacillus subtilis*	[18]
SdiA	Cell division, DNA replication/repair, drug sensitivity, macromolecular metabolism	*Escherichia coli*	[16]
Spo0A	Entry into sporulation	*Bacillus subtilis*	[1]
σ^B	ECF sigma factor; general stress response factor	*Bacillus subtilis*	[14••]
σ^E	ECF sigma factor; sporulation	*Bacillus subtilis*	[1]
σ^E	ECF sigma factor; required for heat shock, exposure to SDS and oxidative agents, macrophage survival	*Mycobacterium tuberculosis*	[5•]
σ^F	ECF sigma factor; sporulation	*Bacillus subtilis*	[1]
σ^H	Stationary-phase sigma factor	*Bacillus subtilis*	[2••]
σ^W	ECF sigma factor; induced by alkali shock	*Bacillus subtilis*	[15••]
SoxS	Superoxide stress	*Escherichia coli*	[21]
TCS	Twenty four two-component systems (TCS)	*Bacillus subtilis*	[17]
TCS	Thirty six two-component systems (TCS)	*Escherichia coli*	[11•]
TrpR	Regulation of tryptophan biosynthesis and transport genes	*Escherichia coli*	[13]

Factor, transcription factor; Function, physiological role of the transcription factor; Organism, species used in study.

regions. Such artefacts could lead to the incorrect assignment of alarmingly large numbers of transcription units to a regulon.

Understanding stimulons

The transcriptional response of organisms to environmental perturbations can be large and complex, involving multiple regulons, including cascades of expression of transcription factors and their regulons. The objective in stimulon studies is the description of the integrated response of an organism to a stress. A variety of conditions have been explored, including heat shock in several species [22••,23•,24,25], anaerobiosis [26], growth on rich media versus minimal media [27,28], exponential growth versus stationary phase growth [28], exposure to inhibitors [29] and exposure to light and dark [30] (Table 2). To dissect these responses, it is often preferable to perturb the cells and monitor the changes in gene expression at one or more time points afterwards. The most developed understanding comes from when the mediating transcription factors and their regulons are known. In these cases, a typical approach is to first identify the stimulon gene members by comparing gene expression in a wild-type strain after an environmental perturbation with an unperturbed control. The experiment is then repeated under the same conditions, but with a strain deleted for a relevant transcription factor, enabling the subset of genes that are directly or indirectly dependent on that transcription factor to be identified. [5•,22••,31,32••]. This is an extremely powerful iterative approach that enables stimulons to be dissected systematically.

Identifying transcription factor binding sites

Genome-wide expression experiments should be combined with bioinformatic analysis to identify regulon members and DNA-binding sites of particular transcription factors. The upstream regulatory sequences of regulon and stimulon members are enriched relative to the whole genomic sequence for the binding sites of specific transcription factors. This facilitates their detection using algorithms to search for over-represented sequence patterns [7] (Table 3). In some cases, the binding-site profile is already known for the transcription factor in question, enabling models to be constructed that describe the sequence preferences and from this make genome-wide predictions [1,2••,5•,10,14••,15••,20,31,33]. Such binding-site analysis can aid interpretation of array data by helping to differentiate genes that are directly or indirectly regulated by a particular transcription factor.

Prediction of DNA-binding sites may not completely correlate with binding-site occupancy *in vivo* and hence expression data. The reasons for this are that many transcription factors might regulate certain genes only in the presence or absence of other factors at that promoter, and therefore only express some genes under

Table 2

Stimulons studied with microarrays.

Perturbation	Regulon(s)	Organism	References
Sporulation	Spo0A, σ^E, σ^F	*Bacillus subtilis*	[1]
Heat shock from 37 to 48°C	σ^B, HrcA, CtsR, AhrC	*Bacillus subtilis*	[23•]
Heat shock from 37 to 50°C	σ^{32}	*Escherichia coli*	[24]
Heat shock from 37 to 45°C	HspR, HrcA, σ^H, σ^E	*Mycobacterium tuberculosis*	[22••]
Heat shock; 29°C versus 37°C	Fur + others	Group A *Streptococcus*	[25]
Aerobic versus anaerobic growth	FNR, ResDE	*Bacillus subtilis*	[26]
Exponential versus transition to stationary phase	σ^S + others	*Escherichia coli*	[28]
Minimal versus rich media	σ^S, FadR, Lrp + others	*Escherichia coli*	[27,28]
Glycolytic versus gluconeogenic growth	CcpA, IolR	*Bacillus subtilis*	[45]
Acetate versus glucose	CRP + others	*Escherichia coli*	[46]
Light to dark transition		*Synechocystis* sp.	[30]
Acid shock		*Mycobacterium tuberculosis*	[47]
SDS treatment	σ^E + others	*Mycobacterium tuberculosis*	[5•]
Oxidizing agent diamide	σ^H + others	*Mycobacterium tuberculosis*	[31]
Antibiotic inhibition of cell wall synthesis	σ^M and σ^W	*Bacillus subtilis*	[48]
Auto-inducer 2-stimulated quorum sensing	σ^{54} regulators + others	*Escherichia coli*	[49]
Hydrogen peroxide	OxyR, SoxRS, (σ^{32}?), (LexA?) + others	*Escherichia coli*	[32••]
Acivicin	RelA, OxyR, SoxRS, (σ^{32}?), + others	*Escherichia coli*	[29]

Perturbation, environmental perturbation; Regulon(s), known or predicted regulons involved in the response; Organism, species used in study.

certain conditions, or bind to target sites in a cooperative fashion with the same or different transcription factors. In these cases, the target sites are often weak and can be difficult to identify. For example, in identifying the σ^W regulon of *B. subtilis*, Cao *et al.* [15••] compared genome-wide sequence searches using a promoter consensus sequence and reasonable variations derived from known σ^W-binding sites, genome-wide expression analysis comparing wild-type and σ^W mutant cells, and run-off transcription/macroarray analysis (ROMA) to identify σ^W-dependent transcripts. They found that at least 50% of the σ^W operons could be identified by any one method, but that no single approach identifies more than 80%. This illustrates that a variety of approaches and conditions should be used to identify regulon members.

Functional transcription factor binding sites *in vivo* can also be identified by chromatin immunoprecipitation followed by cDNA microarray hybridization. Using this strategy, Laub *et al.* [34••] identified binding sites of the cell cycle regulator CtrA in *Caulobacter crescentus*. This approach does not actually identify binding sites; instead, it maps them to a specific intergenic region or open reading frame (ORF). However, if used in conjunction with genome-wide expression analysis and bioinformatics to search for DNA sites within the experimentally identified sequences, this method provides a powerful additional tool for discovering functional binding sites of transcription factors.

Table 3

Algorithms to search for over-represented patterns in DNA sequences.

Algorithm	URL	References
AlignAce	http://atlas.med.harvard.edu/	[50]
BioProspector	http://bioprospector.stanford.edu/	[51]
CONSENSUS	http://bioweb.pasteur.fr/seqanal/interfaces/consensus.html	[52]
Gibbs Motif Sampler	http://bayesweb.wadsworth.org/gibbs/gibbs.html	[53]

Transcriptome analysis: operons, untranslated regions, small RNAs and antisense transcripts

Fully understanding the transcriptome of an organism requires the identification of transcripts and mapping of their 5′ and 3′ ends. However, even with the most well-annotated genomes, our knowledge of transcripts is limited to the ORFs. Little is known about genes that are co-transcribed as multigene transcripts, the extent of the upstream or downstream untranslated regions (UTRs) of each transcript, short untranslated transcripts that form small RNAs (sRNAs), and antisense transcripts of ORFs. The use of microarrays, especially high-density oligonucleotide arrays, has enabled rapid progress in this area.

Several strategies have been used to identify genes that are co-transcribed to form operons. Sabatti *et al.* [35] used publicly available *E. coli* genome-wide expression datasets, performed under a variety of conditions, to search for adjacent genes having correlated expression profiles to suggest that they are co-transcribed. Their approach relies on genes being differentially expressed, and as a consequence there still remains a significant number of adjacent genes in which no expression correlation can be determined. The Affymetrix high-resolution oligonucleotide arrays provide a useful tool in mapping transcripts with reasonable resolution [36]. The *E. coli* array contains

25-mer oligonucleotides that are spaced every 60 nucleotides for ORFs representing the coding strand, and every six nucleotides for intergenic regions >40 nt in length representing both strands. Tjaden and co-workers [37,38•] used these high-density oligonucleotide arrays to analyse *E. coli* expression data from different growth conditions to identify correlated expression patterns of intergenic probes and adjacent ORF probes. This enabled them to identify many new, unanticipated operons, as multigene expression results in transcription of the intervening intergenic region. As the arrays only contain probes for large intergenic regions, not all potential multigene transcripts could be identified. Nevertheless, both methods provide valuable data that can be used to refine genome annotations and suggest the possibility of fundamental change in our understanding of transcriptional circuitry.

The intergenic probes on the Affymetrix arrays can also be used to map the limits of 5′ UTRs of transcripts by analyzing the expression patterns of intergenic probes upstream of ORFs [37]. Given the spacing between intergenic probes, the start of a particular 5′ UTR can only be mapped to within a distance range, but this is often sufficient to identify promoter sequences. The high-density probes on the Affymetrix arrays can also be used to identify short transcripts and others that do not correspond to ORFs. For example, both Tjaden *et al.* [38•] and Wassarman *et al.* [39] identified sRNAs. Extensive transcriptome analysis of *E. coli* by Tjaden *et al.*, identified 4052 coding transcripts and 1102 additional transcripts comprising large intergenic regions that are continuously transcribed from flanking genes to form operons, long 5′ and 3′ UTRs, new ORFs, sRNAs and 317 novel transcripts with unknown function.

An interesting observation is the prevalence of antisense transcripts. Several groups report read-through of convergent gene pairs, which results in antisense expression of one of the genes [10,20,23•]. In addition, an Affymetrix high-density array in which the probes for the ORFs were complementary to the non-coding strand enabled the detection of antisense transcripts from 3000 to 4000 ORFs [36]. The physiological significance of this remarkably high-level of antisense expression in *E. coli* needs to be addressed. It is important to note that arrays comprising double-stranded PCR products for each ORF detect both sense and antisense gene expression and are unable to distinguish between the two. By contrast, oligonucleotide arrays only detect expression of the target strand, unless oligonucleotides are designed for both coding and non-coding strands of ORFs.

mRNA decay

Our understanding of the contribution of mRNA decay to the regulation of protein production at a global level is extremely limited. Bernstein *et al.* [40] derived mRNA stability half-life measurements for most mRNAs by quantifying their levels with DNA microarrays at multiple time points after transcription arrest by the addition of rifampicin to *E. coli* cultures. Surprisingly, the authors found no correlation of mRNA stability with transcript abundance, the number of putative RNase E cleavage sites or the predicted folded state of the 5′ or 3′ UTR, concluding that the mechanism of RNA decay is still largely unknown. In addition, the authors suggest that, globally, synthesis rather than degradation is the predominant factor in determining the steady-state mRNA level.

Use of genomic controls

With the increasing amount of available bacterial microarray data, it is useful to be able directly to compare and contrast gene-expression datasets produced by different research groups. This requires that experimental samples are compared with a standard control, such as a designated control strain grown under standard and reproducible conditions, or a pool of mixed RNAs derived from different time points or conditions. Although this might be sufficient for one laboratory, it is too problematic to be used by many different laboratories. Talaat *et al.* [41], extending the idea of Wei *et al.* [28], used genomic normalization to enhance cross-comparability between microarray experiments in which they compared mRNA-derived gene expression levels with hybridized genomic DNA from the same organism as a standard control. The trade-off is that while using hybridized genomic DNA as a control requires two slides to compare the expression patterns of two conditions, time-course or complex multi-comparison experiments are easily analyzed with genomic DNA as the common denominator.

Conclusions

Over the next few years there will be a wealth of genomic expression data, enabling visualization of the complexities of gene regulation and transcriptional networks at a global level. For detailed analysis, this requires a high standard of microarray experimentation, using well-characterized strains, exploration of different expression conditions to identify the complete breadth of a regulon, experimental repetition and statistical analysis to identify significantly differentially regulated genes, and bioinformatics to predict and identify regulatory sequences. In addition, to capitalize on the opportunities and precedents offered by studies of bacterial physiology it is important to facilitate microarray data retrieval for analysis by other research groups. This can be achieved by adapting the guidelines of the MIAME (minimum information about a microarray experiment) protocol laid out by Brazma *et al.* [42].

Acknowledgements

Owing to space limitations, we apologize to the authors of many studies that we were unable to cite. We thank Joyce West for critical reading of the manuscript and acknowledge support from the National Institutes of Health (NIH), Grant GM57755.

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

2. •• Britton RA, Eichenberger P, Gonzalez-Pastor JE, Fawcett P, Monson R, Losick R, Grossman AD: **Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis***. *J Bacteriol* 2002, **184**:4881-4890.

The authors defined the σ^H regulon using a combined microarray approach in which they compared the genome-wide expression profile of a wild-type strain with a *sigH* mutant and also performed time-course expression analysis after overexpression of *sigH*. From these datasets, the authors identified genes that are directly regulated by σ^H by searching for σ^H-binding sites with hidden Markov models trained from known σ^H sites. Their strategy is very similar to that of Price *et al.* (2001) [14••].

3. Hung SP, Baldi P, Hatfield GW: **Global gene expression profiling in *Escherichia coli* K-12. The effects of leucine-responsive regulatory protein**. *J Biol Chem* 2002, **277**:40309-40323.

4. Tani TH, Khodursky A, Blumenthal RM, Brown PO, Matthews RG: **Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis**. *Proc Natl Acad Sci USA* 2002, **99**:13471-13476.

5. • Manganelli R, Voskuil MI, Schoolnik GK, Smith I: **The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages**. *Mol Microbiol* 2001, **41**:423-437.

This paper illustrates the use of microarrays for bacteria other than *Escherichia coli* and *Bacillus subtilis*. The authors studied the global response to SDS-induced membrane stress that is mediated by σ^E, which in turn regulates σ^B. The authors also identified members of the stimulon that require σ^E (and σ^B) and performed global sequence searches for σ^E promoter elements.

6. Ogura M, Yamaguchi H, Kobayashi K, Ogasawara N, Fujita Y, Tanaka T: **Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK**. *J Bacteriol* 2002, **184**:2344-2351.

7. Masuda N, Church GM: ***Escherichia coli* gene expression responsive to levels of the response regulator EvgA**. *J Bacteriol* 2002, **184**:6225-6234.

8. Lehnen D, Blumer C, Polen T, Wackwitz B, Wendisch VF, Unden G: **LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in *Escherichia coli***. *Mol Microbiol* 2002, **45**:521-532.

9. Beliaev AS, Thompson DK, Fields MW, Wu L, Lies DP, Nealson KH, Zhou J: **Microarray transcription profiling of a *Shewanella oneidensis* etrA mutant**. *J Bacteriol* 2002, **184**:4612-4616.

10. Berka RM, Hahn J, Albano M, Draskovic I, Persuh M, Cui X, Sloma A, Widner W, Dubnau D: **Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK**. *Mol Microbiol* 2002, **43**:1331-1345.

11. • Oshima T, Aiba H, Masuda Y, Kanaya S, Sugiura M, Wanner BL, Mori H, Mizuno T: **Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12**. *Mol Microbiol* 2002, **46**:281-291.

The authors analysed 36 two-component regulators in *Escherichia coli* by comparing deletion mutant strains with wild type to provide an 'array' of array data. On the basis of genome-wide expression data the two-component regulators were divided into three groups: those with little or no change; those with significant changes; and those with drastic changes. This approach for studying uncharacterized two-component regulators provides valuable preliminary information on the roles of some of these regulators under standard conditions; however, known regulon members of some of the other two-component systems were not identified. This illustrates the need to explore different physiological conditions when performing these comparisons to ensure that the regulators are in the active 'induced' state in the wild-type strain.

12. Arfin SM, Long AD, Ito ET, Tolleri L, Riehle MM, Paegle ES, Hatfield GW: **Global gene expression profiling in *Escherichia coli* K-12. The effects of integration host factor**. *J Biol Chem* 2000, **275**:29672-29684.

13. 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.

14. •• Price CW, Fawcett P, Ceremonie H, Su N, Murphy CK, Youngman P: **Genome-wide analysis of the general stress response in *Bacillus subtilis***. *Mol Microbiol* 2001, **41**:757-774.

The authors studied the general stress response mediated by σ^B using two approaches: comparison of the expression profiles between wild type and a *sigB* mutant strain using time-course analysis after exposure to ethanol; and induction of σ^B in the absence of its antisigma factor, RbsW, followed by time-course expression analysis to identify members of the σ^B regulon. From these datasets, the authors identified genes directly regulated by σ^B by searching for σ^B binding sites using hidden Markov models trained from known σ^B sites. Their strategy is very similar to that of Britton *et al.* (2002) [2••].

15. •• Cao M, Kobel PA, Morshedi MM, Wu MF, Paddon C, Helmann JD: **Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches**. *J Mol Biol* 2002, **316**:443-457.

The authors compared three approaches to identify regulon members. They concluded that no single technique identifies all members, suggesting therefore that a combination of complementary approaches should be used to achieve maximal coverage when defining bacterial regulons.

16. Wei Y, Lee JM, Smulski DR, LaRossa RA: **Global impact of *sdiA* amplification revealed by comprehensive gene expression profiling of *Escherichia coli***. *J Bacteriol* 2001, **183**:2265-2272.

17. Kobayashi K, Ogura M, Yamaguchi H, Yoshida K, Ogasawara N, Tanaka T, Fujita Y: **Comprehensive DNA microarray analysis of *Bacillus subtilis* two-component regulatory systems**. *J Bacteriol* 2001, **183**:7365-7370.

18. Ogura M, Yamaguchi H, Yoshida K, Fujita Y, Tanaka T: **DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B. subtilis* two-component regulatory systems**. *Nucleic Acids Res* 2001, **29**:3804-3813.

19. 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.

20. Courcelle J, Khodursky A, Peter B, Brown PO, Hanawalt PC: **Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli***. *Genetics* 2001, **158**:41-64.

21. Pomposiello PJ, Bennik MH, Demple B: **Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate**. *J Bacteriol* 2001, **183**:3890-3902.

22. •• Stewart GR, Wernisch L, Stabler R, Mangan JA, Hinds J, Laing KG, Young DB, Butcher PD: **Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays**. *Microbiology* 2002, **148**:3129-3138.

This paper illustrates the use of microarrays for bacteria other than *Escherichia coli* and *Bacillus subtilis*. The authors obtained an overview of the heat-shock response in wild-type cells by comparing expression profiles of cells grown at 37°C with those shifted to 45°C. The heat-shock response was then dissected by repeating the experiments in strains deleted for the mediating transcription factors, HspR and HrcA. Finally, the authors identified genes from the expression data that are likely to be directly regulated by HspR or HrcA, by searching for binding sites.

23. • Helmann JD, Wu MF, Kobel PA, Gamo FJ, Wilson M, Morshedi MM, Navre M, Paddon C: **Global transcriptional response of *Bacillus subtilis* to heat shock**. *J Bacteriol* 2001, **183**:7318-7328.

The authors identified the heat-shock stimulon by comparing the expression profiles at multiple-time points of cells grown at 37°C with those after shifting to 48°C. The time-course analysis enabled the authors to identify cascades of gene expression mediated by different transcription factors after heat-shock.

24. 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.

25. Smoot LM, Smoot JC, Graham MR, Somerville GA, Sturdevant DE, Migliaccio CA, Sylva GL, Musser JM: **Global differential gene expression in response to growth temperature alteration in group A *Streptococcus***. *Proc Natl Acad Sci USA* 2001, **98**:10416-10421.

26. Ye RW, Tao W, Bedzyk L, Young T, Chen M, Li L: **Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions**. *J Bacteriol* 2000, **182**:4458-4465.

27. 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.

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

29. Smulski DR, Huang LL, McCluskey MP, Reeve MJ, Vollmer AC, Van Dyk TK, LaRossa RA: **Combined, functional genomic-biochemical approach to intermediary metabolism: interaction of acivicin, a glutamine amidotransferase inhibitor, with *Escherichia coli* K-12**. *J Bacteriol* 2001, **183**:3353-3364.

30. Gill RT, Katsoulakis E, Schmitt W, Taroncher-Oldenburg G, Misra J, Stephanopoulos G: **Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. strain PCC6803**. *J Bacteriol* 2002, **184**:3671-3681.

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

32. •• 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.

The authors analysed the response of *Escherichia coli* to hydrogen peroxide, mediated by OxyR, by comparing hydrogen peroxide treated and untreated wild-type cells to identify genes sensitive to peroxide stress and also *oxyR* deleted strains to identify OxyR-dependent peroxide stress genes. Using this approach, the authors identified members of the peroxide-dependent OxyR regulon and also OxyR-independent hydrogen peroxide regulated genes that might be regulated by other peroxide regulators such as the SoxRS system and an unknown regulator of the *isc* operon involved in Fe-S cluster formation. The genome-wide expression analysis of the OxyR regulon was combined with computer-directed searches for OxyR binding sites in [33].

33. Zheng M, Wang X, Doan B, Lewis KA, Schneider TD, Storz G: **Computation-directed identification of OxyR DNA binding sites in *Escherichia coli***. *J Bacteriol* 2001, **183**:4571-4579.

34. •• Laub MT, Chen SL, Shapiro L, McAdams HH: **Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle**. *Proc Natl Acad Sci USA* 2002, **99**:4632-4637.

The spread of microarray technology to other bacteria is again illustrated here. Perhaps, more importantly, the authors identify functional binding sites of specific transcription factors *in vivo* by chromatin immunoprecipitation followed by hybridization of the precipitated DNA to arrays containing PCR products of both ORFs and intergenic regions. It is possible that similar results might be achieved for bacteria with short intergenic regions with arrays only containing PCR products of ORFs.

35. Sabatti C, Rohlin L, Oh MK, Liao JC: **Co-expression pattern from DNA microarray experiments as a tool for operon prediction**. *Nucleic Acids Res* 2002, **30**:2886-2893.

36. Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ, Church GM: **RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array**. *Nat Biotechnol* 2000, **18**:1262-1268.

37. Tjaden B, Haynor DR, Stolyar S, Rosenow C, Kolker E: **Identifying operons and untranslated regions of transcripts using *Escherichia coli* RNA expression analysis**. *Bioinformatics* 2002, **18**:S337-S344.

38. • Tjaden B, Saxena RM, Stolyar S, Haynor DR, Kolker E, Rosenow C: **Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays**. *Nucleic Acids Res* 2002, **30**:3732-3738.

The authors use high-density oligonucleotide arrays for identifying the transcripts that make up the transcriptome. Tjaden *et al.* (2002) [37] published a related paper in which they use hidden Markov models to estimate the boundaries of mRNAs of transcribed genes to define operons and untranslated regions.

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

40. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN: **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* 2002, **99**:9697-9702.

41. Talaat AM, Howard ST, Hale W, Lyons R, Garner H, Johnston SA: **Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis***. *Nucleic Acids Res* 2002, **30**:e104.

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

43. Zwiesler-Vollick J, Plovanich-Jones AE, Nomura K, Bandyopadhyay S, Joardar V, Kunkel BN, He SY: **Identification of novel hrp-regulated genes through functional genomic analysis of the *Pseudomonas syringae* pv. tomato DC3000 genome**. *Mol Microbiol* 2002, **45**:1207-1218.

44. Fouts DE, Abramovitch RB, Alfano JR, Baldo AM, Buell CR, Cartinhour S, Chatterjee AK, D'Ascenzo M, Gwinn ML, Lazarowitz SG *et al.*: **Genomewide identification of *Pseudomonas syringae* pv. tomato DC3000 promoters controlled by the HrpL alternative sigma factor**. *Proc Natl Acad Sci USA* 2002, **99**:2275-2280.

45. Yoshida K, Kobayashi K, Miwa Y, Kang CM, Matsunaga M, Yamaguchi H, Tojo S, Yamamoto M, Nishi R, Ogasawara N *et al.*: **Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis***. *Nucleic Acids Res* 2001, **29**:683-692.

46. Oh MK, Rohlin L, Kao KC, Liao JC: **Global expression profiling of acetate-grown *Escherichia coli***. *J Biol Chem* 2002, **277**:13175-13183.

47. Fisher MA, Plikaytis BB, Shinnick TM: **Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes**. *J Bacteriol* 2002, **184**:4025-4032.

48. Cao M, Wang T, Ye R, Helmann JD: **Antibiotics that inhibit cell wall biosynthesis induce expression of the *Bacillus subtilis* sigma(W) and sigma(M) regulons**. *Mol Microbiol* 2002, **45**:1267-1276.

49. DeLisa MP, Wu CF, Wang L, Valdes JJ, Bentley WE: **DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli***. *J Bacteriol* 2001, **183**:5239-5247.

50. Roth FP, Hughes JD, Estep PW, Church GM: **Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation**. *Nat Biotechnol* 1998, **16**:939-945.

51. Liu X, Brutlag DL, Liu JS: **BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes**. *Pac Symp Biocomput* 2001:127-138.

52. Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences**. *Bioinformatics* 1999, **15**:563-577.

53. Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC: **Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment**. *Science* 1993, **262**:208-214.