
**GENOMICS PROTEOMICS AND
BIOINFORMATICS:**
**Global Gene Expression Profiling in
*Escherichia coli*K12: THE EFFECTS OF
LEUCINE-RESPONSIVE REGULATORY
PROTEIN ,**

She-pin Hung, Pierre Baldi and G. Wesley
Hatfield

J. Biol. Chem. 2002, 277:40309-40323.

doi: 10.1074/jbc.M204044200 originally published online July 18, 2002

Access the most updated version of this article at doi: [10.1074/jbc.M204044200](https://doi.org/10.1074/jbc.M204044200)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2002/10/24/277.43.40309.DC1.html>

This article cites 78 references, 46 of which can be accessed free at
<http://www.jbc.org/content/277/43/40309.full.html#ref-list-1>

Global Gene Expression Profiling in *Escherichia coli* K12

THE EFFECTS OF LEUCINE-RESPONSIVE REGULATORY PROTEIN*[§]

Received for publication, April 25, 2002, and in revised form, July 17, 2002
Published, JBC Papers in Press, July 18, 2002, DOI 10.1074/jbc.M204044200

She-pin Hung^{‡§}, Pierre Baldi^{¶**}, and G. Wesley Hatfield^{‡***‡‡§§}

From the [‡]Departments of Microbiology and Molecular Genetics and of [¶]Biological Chemistry, College of Medicine, the ^{‡‡}Department of Chemical Engineering and Material Sciences, School of Engineering, the ^{¶¶}Department of Information and Computer Science, and the ^{**}Institute for Genomics and Bioinformatics, University of California, Irvine, California 92697

Leucine-responsive regulatory protein (Lrp) is a global regulatory protein that affects the expression of multiple genes and operons in bacteria. Although the physiological purpose of Lrp-mediated gene regulation remains unclear, it has been suggested that it functions to coordinate cellular metabolism with the nutritional state of the environment. The results of gene expression profiles between otherwise isogenic *lrp*⁺ and *lrp*⁻ strains of *Escherichia coli* support this suggestion. The newly discovered Lrp-regulated genes reported here are involved either in small molecule or macromolecule synthesis or degradation, or in small molecule transport and environmental stress responses. Although many of these regulatory effects are direct, others are indirect consequences of Lrp-mediated changes in the expression levels of other global regulatory proteins. Because computational methods to analyze and interpret high dimensional DNA microarray data are still an early stage, much of the emphasis of this work is directed toward the development of methods to identify differentially expressed genes with a high level of confidence. In particular, we describe a Bayesian statistical framework for a posterior estimate of the standard deviation of gene measurements based on a limited number of replications. We also describe an algorithm to compute a posterior estimate of differential expression for each gene based on the experiment-wide global false positive and false negative level for a DNA microarray data set. This allows the experimenter to compute posterior probabilities of differential expression for each individual differential gene expression measurement.

During the last 50 years, a great deal of knowledge about the regulation of gene expression in *Escherichia coli* has been obtained. We now know that the expression of genetic information is regulated at three hierarchical levels: global control of

basal level gene expression by chromosome structure, control of regulons and stimulons by global regulatory proteins, and operon-specific controls (1, 2). At the most general level, the expression of all genes is regulated by DNA supercoiling-dependent mechanisms that affect the topology of the entire chromosome (3). At the next level, large groups of genes are regulated by abundant regulatory proteins with rather degenerate binding site specificity that, in cooperation with operon-specific controls, regulate often-overlapping groups of metabolically related operons, called regulons and stimulons, in response to environmental or metabolic signals. At the most basic level, individual genes or operons are regulated by less abundant proteins that bind in a site-specific manner to one or a few sites to regulate single genes or operons. Isolated examples of each level of control have been described. However, the definition of these hierarchical control levels in a depth sufficient to understand genetic regulatory networks on a global scale, all the way from specific circuits up to the complete regulatory network of the cell, remains to be elucidated (4). Before we can infer and model these regulatory networks, individual components at each hierarchical level must be identified. In other words, a more complete definition of the genes of specific regulons and stimulons must be obtained. It is now possible to obtain much of this information using high-throughput technologies such as DNA microarrays.

The purpose of the work presented here is to identify the network of genes that are differentially regulated by the global *E. coli* regulatory protein, leucine-responsive regulatory protein (Lrp),¹ during steady state growth in a glucose supplemented minimal salts medium. Lrp is a DNA-binding protein that has been reported to affect the expression of approximately 55 genes.² In most cases, Lrp has been reported to activate operons that encode genes for biosynthetic enzymes and repress operons that encode genes for catabolic enzymes (5, 6). The intermediary metabolite, L-leucine, is required for the binding of Lrp at some of its DNA target sites; however, at other sites L-leucine inhibits DNA binding, and at still other sites it exerts no effect at all. Although the physiological purpose of Lrp-mediated gene regulation remains unclear, it has

* This work was supported in part by the Institute of Genomics and Bioinformatics (University of California, Irvine) and National Institutes of Health Grant GM-55073 (to G. W. H.), by a Laurel Wilkening faculty innovation award, and by a Sun Microsystems award (to P. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] The on-line version of this article (available at <http://www.jbc.org>) contains all of the raw and processed data for the experimental results reported here.

[§] Supported by a training grant from the University of California Biotechnology Research and Education Program.

^{§§} To whom correspondence should be addressed: Dept. of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, CA 92697. Tel.: 949-824-5344; Fax: 949-824-8598; E-mail: gwhatfie@uci.edu.

¹ The abbreviations used are: Lrp, leucine-responsive regulatory protein; ATP- γ S, adenosine 5'-O-(thiotriphosphate); MMLV, Moloney murine leukemia virus; BSA, bovine serum albumin; PPDE, posterior probability for differential expression; MOPS, 4-morpholinepropanesulfonic acid; MES, 4-morpholineethanesulfonic acid; dH₂O, distilled water; ORF, open reading frame; AD, average difference; PM, perfect match; MM, mismatch.

² Although the expression levels of approximately 55 genes have been reported to be affected by Lrp, these observations were obtained under a wide variety of environmental and nutritional growth conditions; thus, the expression of some of these genes might not be affected by the presence or absence of Lrp under the growth conditions employed in this study.

been suggested that it might function to coordinate cellular metabolism with the nutritional state of the environment by monitoring the levels of free L-leucine in the cell. The experiments reported here were carried out in the absence of exogenous L-leucine.

Although many data analysis techniques have been applied to DNA microarray data, this field is still evolving and has not yet reached a level of maturity. Therefore, much of the emphasis of the work reported here is directed toward the assessment of methods to identify differentially expressed genes with a high level of confidence. In particular, we apply a Bayesian statistical framework to derive a regularized estimate of the standard deviation of the level of expression of each gene in each condition based on a limited number of replications, and an algorithm to compute a posterior estimate of differential expression for each gene to estimate the global false positive rate specific for each DNA microarray experiment.

MATERIALS AND METHODS

Chemicals and Reagents—Avian myeloblastosis virus reverse transcriptase, ATP γ S, glycogen, and Sephadex G-25 Quickspin columns were obtained from Roche Molecular Biochemicals. Ribonuclease inhibitor III was from Panvera/Takara. Ultrapure deoxynucleoside triphosphates and DNase I were from Amersham Biosciences. Random hexamer oligonucleotides and T4 polynucleotide kinase were from New England Biolabs. [α -³²P]dCTP (2–3000 Ci/mmol) was from PerkinElmer Life Sciences. DNA filter arrays (Panorama *E. coli* gene arrays) were from Sigma-Genosys Biotechnologies. DNA-free kit and 5 M NaCl RNase-free, DNase-free solution were from Ambion, Inc. 16 S rRNA-specific primers, 23 S rRNA-specific primers, and Biotin-Oligo 948 (high performance liquid chromatography-purified) oligonucleotides were from Operon. MMLV reverse transcriptase, dithiothreitol, and ribonuclease H (RNase H) were from Epicentre Technologies. RNeasy total RNA isolation kit and RNA/DNA mini column kit were from Qiagen. Polyethylene oxide-iodoacetyl-biotin, ImmunoPure NeutrAvidin, streptavidin, and 10% Tween 20 were from Pierce. Novex XCell SureLock™ MiniCell and 4–20% TBE gel were from Invitrogen. 5 \times sucrose gel loading dye was from Amresco. SYBR Gold and R-phycoerythrin streptavidin were from Molecular Probes. Acetylated bovine serum albumin (BSA) solution and phosphate-buffered saline (pH 7.2) were from Invitrogen. All other chemicals were obtained from Sigma.

Bacteria Strains and Growth Conditions—Strain IH-G2490 (*ilvP_G::lacZYA*) was constructed by ligating a 515-bp *EcoRI*-*Bam*HI DNA fragment containing a 494-bp *ilvGMEDA*-derived *Hin*FI fragment (base pair position –245 to +249) into the *EcoRI* and *Bam*HI sites of the *lacZ*-truncated pRS551 Δ (yielding the reporter plasmid pRSG2490) and integrating this reporter plasmid construct into the bacterial chromosome of the *polA*-deficient strain, NO3434, as described previously (7). An isogenic *lrp* derivative of strain IH-G2490 was created by generalized P1 transduction of the *lrp*-35::Tn10 allele into this strain according to the methods of Miller (8) to yield strain IH-G2491 (*ilvP_G::lacZYA*, *lrp*::Tn10). The genes of the chromosomal *lac* operon are transcribed from the *ilvP_G* promoter in both strains, which is repressed by the binding of Lrp in the leader-attenuator region upstream of the *ilvG* translational start site. Cells were grown in 50 ml of MOPS medium (9) containing 0.4% glucose in 250-ml Erlenmeyer flasks at 37 °C as described previously (10).

Isolation of Total RNA—Total RNA was isolated from cells at an A_{600} of 0.5–0.6. Ten-ml samples of log phase cells were pipetted directly into 10 ml of boiling lysis buffer (1% SDS, 0.1 M NaCl, 8 mM EDTA) and mixed at 100 °C for 1.5 min. These samples were transferred to 125-ml Erlenmeyer flasks, mixed with an equal volume of hot acid phenol (pH 4.3), and shaken vigorously for 6 min at 64 °C. After centrifugation, the aqueous phase was transferred to a fresh Erlenmeyer flask and the hot acid phenol extraction procedure was repeated. The second aqueous phase was extracted with phenol-chloroform-isoamyl alcohol (25:24:1; pH 4.3) at room temperature, and twice with chloroform-isoamyl alcohol (24:1). Total RNA was precipitated with two volumes of ethanol in 0.3 M NaOAc (pH 5.3), washed with 70% ethanol, and redissolved in a 10 mM Tris, 1 mM EDTA solution (pH 8.0). Residual genomic DNA was removed with the DNA-free kit of Ambion Inc. according to the instructions from the manufacturer. The RNA concentration was determined by absorption at 260 nm. In all cases, independent 10-ml samples from three separate cultures were processed in parallel.

cDNA Synthesis and Target Labeling Conditions for the Nylon Array

Experiments—For random hexamer-primed cDNA synthesis, 20 μ g of total RNA and 37.5 ng of random hexamer primers were heated at 70 °C for 3 min and quickly cooled on ice. cDNA synthesis was performed at 42 °C for 3 h in a 60- μ l reaction mixture containing: RNA and primer mixture; reverse transcriptase buffer (Roche); 1 mM amounts each of dATP, dGTP, and dTTP; 50 μ Ci of [α -³²P]dCTP; 20 units of ribonuclease inhibitor III; and 4 μ l (88 units) of avian myeloblastosis virus reverse transcriptase. Labeled cDNA targets were separated from unincorporated nucleotides on Sephadex G-25 Quickspin columns.

mRNA Enrichment and Target Labeling Conditions for the Affymetrix GeneChip™ Experiments—To enrich the proportion of mRNA in the total RNA preparation, 300 μ g of total RNA from IH-G2490 (*lrp*⁺) or IH-G2491 (*lrp*[–]) was prepared as described above. Each 300- μ g total RNA preparation was split into 12 aliquots to increase the efficiency of the enrichment procedure. All reactions were performed in PCR tubes in a thermocycler. For each reaction, 25 μ g of total RNA were mixed with 70 pmol of a rRNA-specific primer mix in a final volume of 30 μ l. Each specific primer mix included three specific primers for 16 S rRNA (5'-CCTACGGTTACCTTGTT-3', 5'-TTAACCTTGCGGCCGTACTC-3', and 5'-TCCGATTAACGCTTGACCC-3') and five specific primers for 23 S rRNA (5'-cctcagcgttcattagt-3', 5'-CTATAGTAAAGGTTACCGGG-3', 5'-TCGTCATCACGCGCTCAGCC-3', 5'-TCCCACATGTTTCCCA-C-3', and 5'-CATGGAAAACATATTACC-3'). This mixture was heated to 70 °C for 5 min and quickly cooled to 4 °C. 10 μ l of 10 \times MMLV reverse transcriptase buffer (0.5 M Tris-HCl (pH 8.3), 0.1 M MgCl₂, and 0.75 M KCl), 5 μ l of 10 mM dithiothreitol, 2 μ l of 25 mM dNTPs mix, 3.5 μ l of 20 units/ μ l SuperRNasin, 6 μ l of 50 units/ μ l MMLV reverse transcriptase, and water were added to each tube to a final volume of 100 μ l. The reactions were incubated at 42 °C for 25 min, and incubation was continued at 45 °C for 20 min for cDNA synthesis. To remove the rRNA moiety from the rRNA/cDNA hybrid, 5 μ l of 10 units/ μ l RNase H was added and the mixture was incubated at 37 °C for 45 min. RNase H was inactivated by heating at 65 °C for 5 min. Newly synthesized cDNA was removed by incubation with 4 μ l of 2 units/ μ l DNase I and 1.2 μ l of 20 units/ μ l SuperRNasin at 37 °C for 2 h. Four reactions were combined for RNA cleanup with a single Qiagen RNeasy mini column. The quantity of enriched mRNA was measured by absorbance at 260 nm. A typical yield is 10–20 μ g of RNA from 300 μ g of total RNA constituting a 10–20-fold enrichment of mRNA to rRNA.

For the RNA fragmentation, a maximum of 20 μ g of RNA was added to a PCR tube containing 10 μ l of 10 \times NEB buffer for T4 polynucleotide kinase in a final volume of 88 μ l. The tube was incubated at 95 °C for 30 min and cooled to 4 °C.

For the RNA 5'-thiolation and biotin-labeling reaction, 2 μ l of 5 mM ATP γ S and 10 μ l of 10 units/ μ l T4 polynucleotide kinase were incubated with the fragmented RNA at 37 °C for 50 min. The reaction was inactivated by heating to 65 °C for 10 min and cooled to 4 °C. Excess ATP γ S was removed by ethanol precipitation. Fragmented thiolated RNA was collected by centrifugation in the presence of glycogen (0.25 μ g/ μ l) and resuspended in 90 μ l of distilled water (dH₂O). 6 μ l of 500 mM MOPS (pH 7.5) and 4.0 μ l of 50 mM polyethylene oxide-iodoacetyl-biotin were added to the fragmented thiolated RNA and incubated at 37 °C for 1 h. The biotin-labeled RNA was isolated by ethanol precipitation, washed twice with 70% ethanol, and dried and dissolved in 20–30 μ l of molecular biology grade water. The quantity of the biotin-labeled RNA was measured by absorbance at 260 nm. The total yield for the entire procedure is typically 2–4 μ g of biotin-labeled RNA from 300 μ g of total RNA. The efficiency of RNA fragmentation and biotin labeling can be monitored with a gel shift assay where the biotin-labeled RNA is pre-incubated with avidin prior to electrophoresis. Biotin-labeled RNA is retarded during electrophoresis because of the avidin-biotin interaction. The position of the RNA in the gel addresses the fragmentation efficiency. The amount of shifted RNA indicates the efficiency of the biotin labeling. Inefficiencies in either of these parameters should be addressed before proceeding to the hybridization step.

Hybridization to Nylon Filters—The nylon filters were soaked in 2 \times SSPE (20 \times SSPE contains 3 M NaCl, 0.2 M NaH₂PO₄, and 25 mM EDTA) for 10 min and prehybridized in 10 ml of prehybridization solution (5 \times SSPE, 2% SDS), 1 \times Denhardt's solution (50 \times Denhardt's solution contains 5 g of Ficoll, 5 g of polyvinylpyrrolidone, 5 g of bovine serum albumin, and H₂O to 500 ml), and 0.1 mg/ml sheared herring sperm DNA) for at least 1 h at 65 °C. 5–7 \times 10⁷ cpm of cDNA targets in 500 μ l of prehybridization solution were heated at 95 °C for 10 min, rapidly cooled on ice, and added to 5.5 ml of prehybridization solution. The prehybridization solution was removed and replaced with the hybridization solution. Hybridization was carried out for 15–18 h at 65 °C. Following hybridization each filter was rinsed with 50 ml of 0.5 \times SSPE containing 0.2% SDS at room temperature for 3 min, followed by three

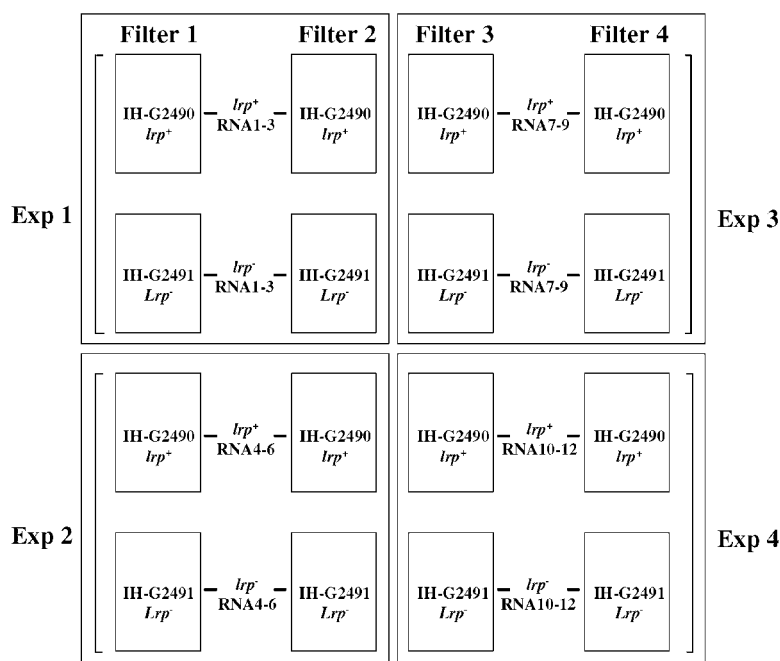


FIG. 1. Experimental design for nylon filter DNA array experiments. See "Materials and Methods" for description.

washes in the same wash solution at 65 °C for 20 min each. The filters were partially air dried, wrapped in Saran Wrap, and exposed to a phosphor screen for 15–30 h. Filters were stripped by microwaving at 30% maximal power (1400 watts) in 500 ml of 10 mM Tris solution (pH 8.0) containing 1 mM EDTA and 1% SDS for 20 min. Stripped filters were wrapped in Saran Wrap and stored in the presence of damp paper towels in sealed plastic bags at 4 °C.

Hybridization to Affymetrix GeneChips—For hybridization of biotinylated RNA targets to the Affymetrix GeneChips, 2–4 μ g of fragmented biotin-labeled RNA of IH-G2490 (*lrp*⁺) and IH-G2491 (*lrp*⁻) were used for each GeneChip. The hybridization solution for each array was prepared with 100 μ l of 2 \times MES hybridization buffer (200 mM MES, 2 M NaCl, 40 mM EDTA, and 0.02% Tween 20), 1 μ l of 100 nM Biotin-Oligo 948 (5'-biotin-GTCAAGATGCTACCGTTCAG-3'), 2 μ l of 10 mg/ml hering sperm DNA, 2 μ l of 50 mg/ml BSA, and 2–4 μ g of fragmented biotin-labeled RNA and brought final volume to 200 μ l with molecular biology grade water. The GeneChip arrays were equilibrated to room temperature immediately before use. The hybridization solution prepared above was added to each GeneChip and incubated in a GeneChip hybridization oven (Affymetrix) at 45 °C for 16 h at a rotation rate of 60 rpm.

Following hybridization, the stain and wash procedures were carried out in an Affymetrix GeneChip Fluidics Station 400 using the ProKGE-WS2 fluidics script to run the machine. Streptavidin solution mix (300 μ l of 2 \times MES stain buffer, 24 μ l of 50 mg/ml BSA, 6 μ l of 1 mg/ml streptavidin, and 270 μ l of dH₂O), antibody solution (300 μ l 2 \times MES stain buffer, 24 μ l of 50 mg/ml BSA, 6 μ l of 10 mg/ml normal goat IgG, 6 μ l of 0.5 mg/ml biotin anti-streptavidin, and 264 μ l of dH₂O) and SAPE solution (300 μ l of 2 \times MES stain buffer, 24 μ l of 50 mg/ml BSA, 6 μ l of 1 mg/ml streptavidin-phycoerythrin, and 270 μ l of dH₂O) were prepared in amber tubes for the staining of each probe array. After hybridization, the hybridization solution was removed and kept at 4 °C. Each GeneChipTM was filled with 300 μ l of nonstringent wash buffer (6 \times SSPE, 0.01% Tween 20, 0.005% Antifoam). The GeneChips were inserted into the fluidics station, and the ProKGE-WS2 protocol was selected to control the staining and washing of the probe arrays. After the procedure was complete, the GeneChips were removed from the fluidics station and checked for large bubbles or air pockets before scanning. The buffer in the GeneChips were drained and refilled with nonstringent buffer if bubbles were present.

Experimental Design for Nylon Filter DNA Array Experiments—The experimental design for the nylon filter DNA array experiments reported here is diagrammed in Fig. 1. In experiment 1, filters 1 and 2 were hybridized with ³³P-labeled, random hexamer-generated cDNA targets complementary to each of three RNA preparations (RNA 1–3) obtained from the cells of three individual cultures of the *lrp*⁺ strain (IH-G2490). These three ³³P-labeled cDNA target preparations were pooled prior to hybridization to the full-length ORF probes on the filters (experiment 1). Following PhosphorImager analysis, these filters were

stripped and again hybridized with pooled, ³³P-labeled cDNA targets complementary to each of another three independently prepared RNA preparations (RNA 1–3) from the *lrp*⁻ (IH-G2491) (experiment 1). This procedure was repeated two more times with filters 1 and 2 using two more independently prepared pools of cDNA targets (experiment 2, RNA 4–6). Another set of filters, filters 3 and 4, were used for experiments 3 and 4 as described for experiments 1 and 2. This protocol results in duplicate filter data for four experiments performed with cDNA targets complementary to four independent prepared sets of pooled RNA. Thus, because each filter contains duplicate spots for each ORF and duplicate filters were used for each experiment, four measurements for each ORF from each experiment were obtained. These four measurements for each experiment were averaged for further statistical analysis.

Experimental Design for Affymetrix GeneChip Experiments—The experimental design for the Affymetrix GeneChip experiments reported here is diagrammed in Fig. 2. The same 24 total RNA preparations used for the nylon filter experiments were pooled into sets of 3 and used for the preparation of biotin-labeled RNA targets for hybridization to Affymetrix GeneChips. For experiments 1–4, four GeneChips were hybridized with biotin-labeled RNA pools 1–3, 4–6, 7–9, and 10–12 prepared from *lrp*⁺ cells, and four GeneChips were hybridized with biotin-labeled RNA pools 1–3, 4–6, 7–9, and 10–12 prepared from *lrp*⁻ cells, respectively. One average difference measurement for each gene probe set on each GeneChip was obtained for subsequent data processing and analysis.

Data Acquisition from the Nylon Filter DNA Array—A commercial software package obtained from Research Imaging Inc. (DNA ArrayVision) was used to grid the 16-bit image file obtained from the Phosphor-Imager, to record the pixel density of each of the 18,432 addresses on each filter, and to perform the background subtractions. 8,580 of the addresses on each filter are spotted with duplicate copies of each of the 4,290 *E. coli* ORFs. The remaining 9,852 empty addresses were used for background measurements. Because the backgrounds were quite constant, a global average background measurement was subtracted from each experimental measurement, although local background calculations are possible. Greater than 4 logs of linearity for the Phosphor-Imager-derived data were observed.

Data Acquisition of Affymetrix GeneChips—Each GeneChip array was scanned twice with an HP GeneArray confocal laser at a 3 μ m resolution, and the intensities at each perfect match (PM) and mismatch (MM) probe cell from both scans were averaged and saved as a *.DAT file. The average intensity of each GeneChip was globally scaled to 2500 and saved as a *.CEL. These probe pair measurements for each probe set were used for subsequent data processing and statistical analysis.

Model-based Oligonucleotide GeneChip Analysis—Gene expression values from Affymetrix GeneChips are based on the average difference (AD) between hybridization signals of PM and MM oligonucleotide probe sets for each gene as described in the expression analysis tech-

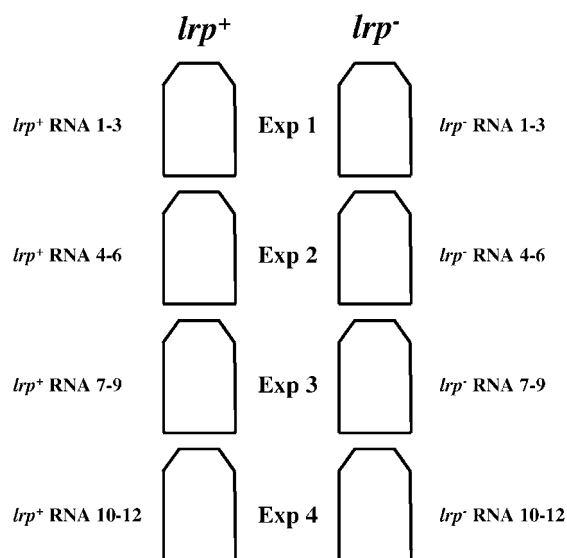


FIG. 2. Experimental design of the Affymetrix GeneChip experiments. See "Materials and Methods" for description.

nical manual from Affymetrix. The AD value of each probe set is calculated as $AD = \Sigma(PM-MM)/\text{number of probe pairs}$. Algorithms incorporated into Affymetrix software remove probe pairs that are out of a given range when calculating AD values for each probe set. In this process, the mean and standard deviation are calculated for intensity differences ($PM - MM$) across the entire probe set (excluding the highest and lowest values), and values within a set number of standard deviations (3 as default) are not included in the calculation. The advantage is that this process minimizes the variance introduced by experimental or biological error by removing the outliers present in each probe set. The disadvantage is that this process does not always remove the same probe pairs for the calculation of the AD values among GeneChips. This can lead to the misinterpretation of the gene expression profiles obtained from GeneChip experiments. To alleviate this problem, a model-based method incorporated into a program called dChip has been described by Li and Wang (11). This method maintains constant probe pair set identities across all GeneChips while excluding outliers because of cross-hybridization, contamination during hybridization, or manufacturing defects that affect probe set measurements. For all of the GeneChip experiments reported here, each probe pair set from the *.CEL files was modeled by the dChip software prior to statistical analysis.

Statistical Methods—As described above, the experimental design employed in this study consists of ³³P-labeled cDNA target preparations for each of two genotypes hybridized to nylon filters, or three biotinylated mRNA target preparations hybridized to Affymetrix GeneChips. The designs for these experiments are depicted in Figs. 1 and 2. For each measurement, a background subtracted estimate of expression level for each gene was obtained and scaled to total counts by dividing each individual gene expression value by the total of all values on the filter or GeneChip. Thus, each normalized gene level is expressed as a fraction of the total mRNA hybridized to each DNA array.

For any given measurement, a value greater than zero (indicating an expression level) or a zero (indicating an expression level lower than background) is obtained. Only those genes exhibiting an expression level greater than zero in all experiments were used for statistical analysis. Gene measurements containing zero expression values were set aside. Among this set of genes, those with zero expression values for all measurements in one genotype, and all values greater than zero for all measurements of another genotype for each experiment were identified. The significance of these results was analyzed by ranking these genes in ascending order according to their coefficients of variance of the four greater than zero measurements. The remaining genes were analyzed both by a simple *t* test and a regularized *t* test based on a Bayesian statistical framework described under "Results and Discussion."

Data Accession—All of the raw and processed data for the experimental results reported here are available in tabular format as supplemental data in the on-line version of this article.

RESULTS AND DISCUSSION

An ad Hoc Method for the Estimation of Global False Positive Levels—To interpret the results of a high dimensional DNA array experiment, it is necessary to determine the global false positive level inherent in the data set being analyzed. The global false positive level reflects all sources of experimental and biological variation inherent in a DNA array experiment. The basic idea is to infer the false positive level in the control *versus* treatment situation from the false positive level observed with the control *versus* control (and/or treatment *versus*, treatment) comparison. With this information, a global level of confidence can be calculated for differentially expressed genes measured at any given statistical significance level. For example, consider an experiment comparing the gene expression profiles of two genotypes, where an average of 10 genes are observed to be differentially expressed with a *p* value less than 0.0001 when gene expression profiles from one genotype are compared with data of the same genotype (e.g. *lrp*⁺ *versus* *lrp*⁺ or *lrp*⁻ *versus* *lrp*⁻). Because no differential expression is expected in these comparisons, these 10 genes are clearly false positives generated by chance occurrences driven by experimental errors and biological variance. Now, if 100 genes are differentially expressed with a *p* value less than 0.0001 when the data from one genotype (*lrp*⁺) are compared with the data from the other genotype (*lrp*⁻), it is reasonable to infer that we can be only 90% confident that the differential expression of any one of these 100 genes is biologically meaningful because 10 false positives are expected from this data set. This example demonstrates that, although the confidence level based on the measurement for an individual gene may exceed 99.99% for two treatment conditions (local confidence of 0.0001), the confidence that this gene is differentially expressed might be only 90% (global confidence of 0.9). This example defines an *ad hoc* method of comparing control to control data to derive an estimate of an experiment-wide false positive level.

We applied this *ad hoc* method for the estimation of false positive levels of the experiments described here by averaging the four measurements for each gene from the duplicate control filters of each experiment hybridized with labeled targets from the control strain IH-G2490 (*lrp*⁺) and comparing these averaged values of control data from experiments 1 and 3 to the averaged values of control data from experiments 2 and 4 (Fig. 1). In another analysis, we compared control data from experiments 1 and 4 to the averaged values of control data from experiments 2 and 3. Equivalent comparisons were performed with filters hybridized with labeled targets from the experimental strain (IH-G2491 (*lrp*⁻)). These particular two-by-two (control *versus* control or experimental *versus* experimental) comparisons were chosen because they average across experimental errors and biological differences both among filters and RNA preparations. The results of a simple *t* test analysis of these data were ranked in ascending order of the *p* values for each gene measurement based on the *t* test distribution. The results of these statistical analyses are shown in Table I.

The data in Table I show that, among the control *versus* control or experimental *versus* experimental comparisons, no genes exhibited a *p* value less than 0.0001. However, an examination of the *p* values observed when the control data were compared with the experimental data shows that 12 genes were differentially expressed with a *p* value less than 0.0001. Thus, we can be fairly certain that these 12 genes are differentially expressed because of biological reasons and not by chance occurrences driven by experimental error and biological variance. On the other hand, we know from the literature that more than 12 genes are regulated by Lrp (5, 6). This demon-

TABLE 1
Determination of confidence level for differentially expressed genes

<i>p</i> value	No. of genes ^a		% Confidence (<i>ad hoc</i>)	PPDE (< <i>p</i>) ^b
	Control <i>vs.</i> control and experimental <i>vs.</i> experimental	Control <i>vs.</i> experimental		
<0.0001	0	12	≈100	0.989
<0.0005	0.25	30	99.2	0.980
<0.001	1	44	97.7	0.975
<0.005	3.75	134	97.2	0.955
<0.01	7.25	208	96.5	0.944

^a Calculated by averaging the control or experimental measurements and comparing experiments 1 and 3 *versus* 2 and 4 or experiments 1 and 4 *versus* 2 and 3 that average data across filters and RNA preparations.

^b Ref. 17.

strates that, given the experimental errors inherent in this experiment, the differentially expressed levels of most genes cannot pass this stringent statistical test. Therefore, to identify other differentially expressed genes, we must lower the stringency of our statistical criterion. The data in Table I show that, as the *p* value is raised to 0.005, we observe an additional 122 genes that are differentially expressed at this threshold level. At the same time, raising the statistical threshold to 0.005 reveals an average of 3.75 genes that appear differentially expressed with a *p* value equal to or less than 0.005 when the control or experimental data sets are compared with themselves. This means that, given this complete data set from four replicate experiments, we expect at least 3.75 false positives among the 134 genes differentially expressed with a *p* value equal to or less than 0.005. Therefore, our global confidence in the identification of any one of these 134 genes as differentially expressed genes is estimated to be 97%.

It should be emphasized that relaxing the *p* value threshold rapidly increases the average number of false positives in the control (*lrp*⁺ *versus* *lrp*⁺ or *lrp*⁻ *versus* *lrp*⁻) data sets relative to the number of genes differentially expressed at the same *p* value in the experimental (*lrp*⁺ *versus* *lrp*⁻) data set and, therefore, decreases the confidence with which differentially expressed genes can be identified.

Improved Statistical Inference from DNA Array Data Using a Bayesian Statistical Framework—A simple *t* test evaluates the distance between the means of two groups normalized in terms of the within-group standard deviations. The result is that large differences between genotypes for any given ORF will be declared nonsignificant if the expression level of that ORF is unreplicable within experimental treatments. Conversely, small differences in expression will be determined to be statistically significant for a given ORF if expression levels for that ORF are replicable within treatments. In short, the *t* test statistic is constructed by scaling the difference in gene expression levels between genotypes relative to the observed variances within genotypes. *p* values based on the *t* test statistic range from 1.0 for gene expression levels with identical values associated with the null hypothesis to very small *p* values for differential gene expression levels that are highly significant.

In a perfect world, all DNA microarray experiments would be highly replicated. Such replication would allow accurate estimates of the variance within experimental treatments to be obtained, and the *t* test would perform well, *i.e.* the variance for each gene measurement would be based on many measurements for that gene. However, DNA microarray experiments are expensive and time-consuming to carry out. As a result, the level of replication within experimental treatments is often low. This results in poor estimates of variance and a correspondingly poor performance of the *t* test itself. On the other hand, we have shown that the confidence in the interpretation of DNA microarray data with a low number of replicates can be

improved by using a Bayesian statistical approach that incorporates information of within treatment measurements (12, 13). This results in a more consistent set of differentially expressed genes identified with fewer replicates. The Bayesian approach is based on the observation that genes of similar expression levels exhibit similar variance. Thus, more robust estimates of the variance of a gene can be derived by pooling neighboring genes with comparable expression levels. For the analysis of the data reported here, we ranked the mean gene expression levels of the replicate experiments in ascending order, used a sliding window of 101 genes, and assigned the average standard deviation of the 50 genes ranked below and above each gene as the background standard deviation for that gene. The variance of any gene within any given treatment then can be estimated by the weighted average of the treatment-specific background variance and the treatment-specific empirical variance across experimental replicates. In the Bayesian approach employed in this study, the weight given to the within experiment gene variance estimate is a function of the number of experimental replicates. This leads to the desirable property that the Bayesian approach employing such a regularized *t* test converges on the same set of differentially expressed genes as the simple *t* test but with fewer replicates (12).

A comparison of the results of statistical analyses employing a simple *t* test and a regularized *t* test is shown in Table II. Here, the simple *ad hoc* method of comparing controls to controls was used to demonstrate that the number of false positives expected at a given *p* value is lower when the Bayesian statistical framework is employed. For example, only 2 false positives are expected at a *p* value threshold less than 0.005 with the Bayesian regularization, whereas 3.75 false positives are expected at this same *p* value threshold with the *t* test alone. At the same time, 188 differentially expressed genes with a *p* value less than 0.005 are observed with the regularized *t* test, whereas only 134 genes are identified at this same threshold with the simple *t* test (Table II). Thus, more genes are identified with a lower false positive level and a higher global confidence level. In other words, complementing the empirical variance of the four experimental measurements for each gene with the corresponding background variance within an experiment improves our confidence in the identification of differentially expressed genes and the number of genes that can be identified at a given *p* value threshold based on a *t* test distribution.

Although the data in Table II show that the Bayesian statistical approach using a regularized *t* test identifies more genes with a higher level of global confidence than the simple *t* test, the natural question that arises is whether these genes are true positives, *i.e.* whether these are Lrp-regulated genes. This question is addressed by the data shown in Fig. 3. For example, of the 44 genes differentially expressed between *lrp*⁺ and *lrp*⁻ strains with a *p* value less than 0.001 identified by a simple *t* test, 10 are known to be Lrp-regulated (Table III).

TABLE II
Comparison of nylon filter DNA array data analyzed with a simple t test and a regularized t test

p value	t -test			p value	Regularized t -test		
	No. of genes		% Confidence		No. of genes		% Confidence
	Control vs. control	Control vs. experimental		Control vs. control	Control vs. experimental		
<0.0001	0	12	≈100	<0.0001	0	39	≈100
<0.0005	0.25	30	99.2	<0.0005	0.25	62	99.6
<0.001	1	44	97.7	<0.001	0.5	79	99.4
<0.005	3.75	134	97.2	<0.005	2	188	98.9
<0.01	7.25	208	96.5	<0.01	3.75	268	98.6

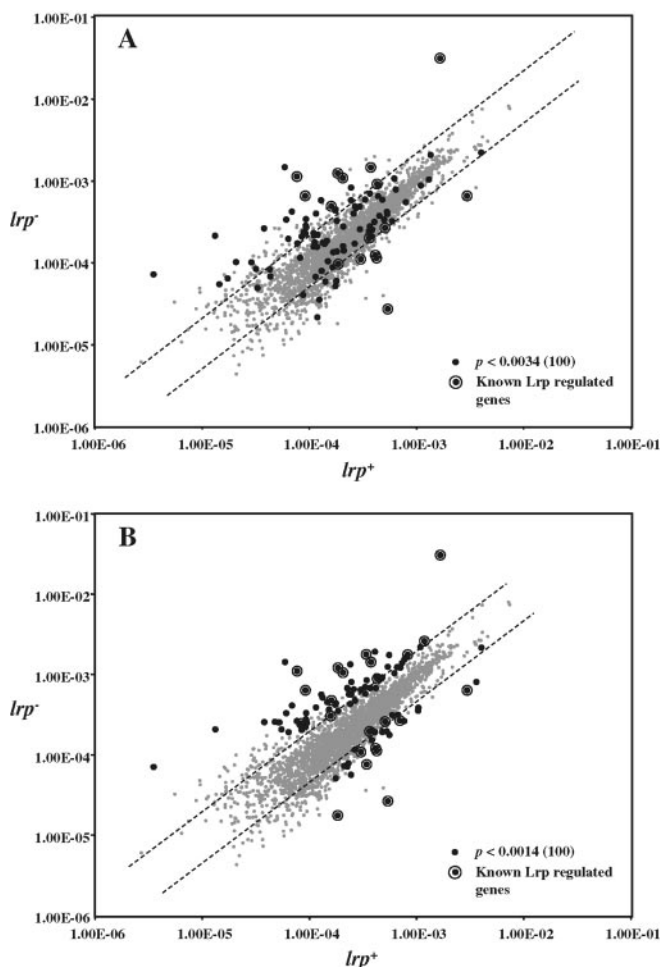


FIG. 3. Scatter plot showing the mean of the fractional mRNA levels obtained from eight filters hybridized with ^{33}P -labeled cDNA targets prepared from three pooled RNA preparations extracted from *Escherichia coli* K12 strains IH-G2490 (lrp^+) and IH-G2491 (lrp^-). A, the larger black dots identify 100 genes differentially expressed between strains IH-G2490 and IH-G2491 with p values less than 0.0034 based on a simple t test distribution. The circled black dots identify genes known to be regulated by Lrp. The gray spots represent the relative expression levels of each of the 2,758 genes expressed at a level above background in all experiments. The dashed lines demarcate the limits of 2-fold differences in expression levels. B, the larger black dots identify 100 genes differentially expressed between strains IH-G2490 and IH-G2491 with p values less than 0.0014 based on a regularized t test. The circled black dots identify genes known to be regulated by Lrp. The gray spots represent the relative expression levels of each of the 2,758 genes expressed at a level above background in all experiments. The dashed lines demarcate the limits of 2-fold differences in expression levels.

However, among the 39 genes differentially expressed between lrp^+ and lrp^- strains with a p value less than 0.0001 identified by the Bayesian approach, 17 are known to be Lrp-regulated (Table IV).

Why does the regularized t test identify more Lrp-regulated

genes? The answer to this question lies in the fact that all of the genes identified to be differentially expressed with a p value less than 0.005 with the regularized t test exhibit -fold changes greater than ~ 1.7 -fold (Fig. 3B). However, many genes identified to be differentially expressed with a p value less than 0.005 with the simple t test exhibit -fold changes as small as ~ 1.2 -fold (Fig. 3A). Furthermore, the 100 genes with the lowest p values identified as differentially expressed by both methods contain only 43 genes in common. Thus, many of the genes identified by the simple t test that are excluded by the Bayesian approach are genes that show small -fold changes. In general, these genes with small -fold changes identified by the simple t test are associated with “too good to be true” within treatment variance estimates, reflecting underestimates of the within treatment variance when the number of observations is small. The elimination of this source of false positives by the Bayesian approach improves the identification of true positives. However, although this is desired, genes that are truly differentially expressed with small -fold changes in the range of ~ 1.2 – 1.7 -fold will also be eliminated by the Bayesian approach. For example, of the 16 genes of the top 100 with the lowest p values identified by the simple t test that are known to be regulated by Lrp, one was not identified by the Bayesian method. This Lrp-regulated gene that did not pass the regularized t test was the *sdaC* gene, previously reported to be regulated by Lrp 3-fold (14, 15) and measured to be regulated 1.9-fold in the experiment performed with the DNA arrays. Nevertheless, although this gene is lost, the overall performance of the regularized t test surpasses that of the simple t test.

At first glance it might appear that the Bayesian approach validates the often-used 2-fold rule for the identification of differentially expressed genes (16), *i.e.* the identification of genes differentially expressed between two experimental treatments with a -fold change greater than 2 in, for example, three of four experiments. This type of reasoning is based on the intuition that larger observed -fold changes can be more confidently interpreted as a stronger response to the experimental treatment than smaller observed -fold changes, which of course is not necessarily the case. An implicit assumption of this reasoning is that the variance among replicates within treatments is the same for every gene. In reality, the variance varies among genes and it is critical to incorporate this information into a statistical test (12). Clearly, with a background standard deviation of, for example, 50, differential expression measurements of 200/100 and 20,000/10,000 have different significance. This point is further emphasized by simply examining the scatter plots in Fig. 3. Here, many genes that appear differentially expressed greater than 2-fold do not exhibit p values less than 0.005 and a global confidence level of at least 97%. This does not mean that these might not be Lrp-regulated genes; it simply means that they are false negatives that cannot be identified at this level of confidence.

Commonly used software packages do not possess algorithms for implementing Bayesian statistical methods. Therefore, we

TABLE III
Genes differentially expressed between *lrp*⁺ and *lrp*⁻ (control vs. experimental) *E. coli* strains with a *p* value less than 0.001 identified with a simple *t* test

The data are presented as the average (mean) and S.D. of four independent gene expression measurements expressed as a fraction of the total hybridization signal (total mRNA) on each DNA microarray filter.

Gene name ^a	Control	Experimental	Control	Experimental	<i>p</i> value	PPDE (< <i>p</i>)	Fold
	<i>mean</i>	<i>mean</i>	<i>S.D.</i>	<i>S.D.</i>			
<i>yecI</i>	3.11E-05	8.48E-05	3.35E-06	7.49E-06	8.62E-06	0.99516	2.73
<i>uvrA</i>	1.28E-03	1.04E-03	1.5E-05	3.37E-05	1.70E-05	0.99386	-1.23
<i>gdhA</i>	9.16E-05	2.73E-04	1.52E-05	2.16E-05	2.18E-05	0.99329	2.98
<i>oppB</i> *	7.51E-05	1.14E-03	2.12E-05	3.79E-04	2.48E-05	0.99298	15.12
<i>b2343</i>	2.82E-05	1.02E-04	3.75E-06	1.92E-05	2.67E-05	0.99280	3.61
<i>artP</i>	6.73E-05	4.23E-04	1.24E-05	1.16E-04	3.60E-05	0.99200	6.28
<i>b1810</i>	1.07E-04	2.32E-04	3.65E-06	3.20E-05	4.47E-05	0.99136	2.17
<i>oppC</i> *	2.01E-04	1.08E-03	2.34E-05	3.61E-04	5.44E-05	0.99074	5.38
<i>gltD</i> *	5.28E-04	2.74E-05	1.28E-04	1.42E-05	5.87E-05	0.99049	-19.27
<i>b1330</i>	1.07E-04	1.58E-04	5.44E-06	9.60E-06	7.16E-05	0.98981	1.47
<i>uup</i>	2.02E-04	1.60E-04	6.65E-06	5.72E-06	7.66E-05	0.98956	-1.26
<i>oppA</i> *	1.62E-03	3.16E-02	7.63E-04	1.03E-02	8.45E-05	0.98920	19.44
<i>malE</i> *	3.56E-04	2.01E-04	2.32E-05	2.17E-05	1.16E-04	0.98793	-1.78
<i>oppD</i> *	8.97E-05	6.55E-04	2.76E-05	2.05E-04	1.16E-04	0.98793	7.30
<i>galP</i>	3.75E-04	2.11E-04	2.25E-05	2.40E-05	1.31E-04	0.98740	-1.78
<i>lysU</i> *	1.81E-04	1.24E-03	7.48E-05	2.78E-04	1.44E-04	0.98697	6.87
<i>hybA</i>	3.53E-04	2.47E-04	2.11E-05	1.50E-05	1.49E-04	0.98682	-1.43
<i>hybC</i>	3.54E-04	2.34E-04	2.20E-05	1.81E-05	1.61E-04	0.98646	-1.51
<i>yhcB</i>	4.25E-05	6.84E-05	2.84E-06	6.28E-06	1.68E-04	0.98625	1.61
<i>yifM_2</i>	1.12E-04	6.74E-05	5.35E-06	7.61E-06	1.81E-04	0.98589	-1.66
<i>ilwG_1</i> *	4.21E-04	9.15E-04	7.55E-05	6.85E-05	2.54E-04	0.98411	2.17
<i>grxB</i>	5.95E-05	3.38E-04	1.92E-05	1.07E-04	2.92E-04	0.98332	5.68
<i>phoP</i>	8.29E-05	2.10E-04	1.20E-05	4.42E-05	3.16E-04	0.98285	2.54
<i>ydjA</i>	1.10E-04	1.79E-04	1.26E-05	1.06E-05	3.54E-04	0.98216	1.62
<i>ydaA</i>	2.61E-04	4.88E-04	3.53E-05	5.45E-05	3.55E-04	0.98214	1.87
<i>yddG</i>	1.77E-04	3.25E-04	2.52E-05	3.37E-05	3.84E-04	0.98165	1.84
<i>emrA</i>	3.58E-04	2.78E-04	2.43E-05	4.57E-06	3.95E-04	0.98147	-1.29
<i>b1685</i>	3.71E-05	2.64E-04	1.20E-05	1.22E-04	4.13E-04	0.98118	7.10
<i>glpA</i>	1.28E-04	8.01E-05	8.54E-06	9.26E-06	4.71E-04	0.98029	-1.59
<i>manA</i>	8.71E-05	2.40E-04	2.16E-05	4.08E-05	4.80E-04	0.98016	2.75
<i>ybeD</i>	1.13E-04	4.01E-04	1.70E-05	1.48E-04	5.15E-04	0.97967	3.55
<i>cfa</i>	2.89E-04	4.89E-04	2.83E-05	6.08E-05	5.16E-04	0.97966	1.69
<i>b3914</i>	6.23E-05	1.97E-04	8.70E-06	5.46E-05	5.44E-04	0.97928	3.16
<i>ybiK</i>	2.03E-04	2.76E-04	1.50E-05	1.47E-05	5.78E-04	0.97884	1.36
<i>yggB</i>	1.73E-04	4.50E-04	3.57E-05	8.50E-05	6.05E-04	0.97850	2.61
<i>amn</i>	4.31E-04	6.51E-04	4.47E-05	4.72E-05	6.07E-04	0.97848	1.51
<i>b1976</i>	1.30E-04	1.77E-04	1.21E-05	6.38E-06	7.56E-04	0.97677	1.36
<i>speB</i>	1.21E-04	3.56E-05	2.09E-05	1.08E-05	7.73E-04	0.97659	-3.40
<i>hdeA</i>	2.40E-04	8.29E-04	8.46E-05	9.90E-05	8.12E-04	0.97619	3.45
<i>lrp</i> *	2.96E-04	1.11E-04	6.21E-05	2.22E-05	8.27E-04	0.97604	-2.67
<i>pheA</i>	9.11E-05	3.41E-04	3.78E-05	4.17E-05	8.36E-04	0.97595	3.75
<i>gst</i>	3.44E-06	7.24E-05	4.05E-06	2.41E-05	8.57E-04	0.97574	21.01
<i>proC</i>	1.76E-04	6.04E-05	5.17E-05	6.99E-06	8.89E-04	0.97543	-2.91
<i>sdaC</i> *	1.82E-04	9.71E-05	2.02E-05	1.76E-05	8.96E-04	0.97536	-1.87

^a Known Lrp-regulated genes are identified by an asterisk.

developed a statistical program, CyberT, which does accommodate this approach. We use the statistical tools incorporated into CyberT to compare and analyze the gene expression data from the experiments described here. CyberT is available for on-line use on the genomics web site of the University of California, Irvine.

A Computational Method to Determine False Positive Levels—A computational version of our *ad hoc* method for estimating false-positive levels has been recently described (17). The basic idea is to consider the *p* values as a new data set and to build a probabilistic model for this new data. When control data sets are compared with one another (*i.e.*, no differential gene expression), it is easy to see that the *p* values ought to have a uniform distribution between 0 and 1. In contrast, when data sets from different genotypes or treatment conditions are compared with one another, the distribution of *p* values will tend to cluster more closely to 0 than 1, *i.e.*, there will be a subset of differentially expressed genes with “significant” *p* values. One can use a mixture of β distributions to model this distribution of *p* values in the form shown in Equation 1.

$$P(p) = \sum_{i=0}^K \lambda_i \beta(p; r_i, s_i) \quad (\text{Eq. 1})$$

For $i = 0$, we use $r_0 = s_0 = 1$ to implement the uniform distribution as a special case of a β distribution. Thus, $K + 1$ is the number of components in the mixture and the mixture coefficients λ_i represent the prior probability of each component. In many cases, two components ($K = 1$) are sufficient but sometimes additional components are needed. In general, the mixture model can be fit to the *p* values using the Expectation Maximization algorithm or other iterative optimization methods to determine the values of the λ , r , and s parameters (17). From the mixture mode given n genes, the estimate of the number of genes for which there is a true difference is $n(1 - \lambda_0)$.

In the case of the data reported here, the parameters of the mixture model of Equation 1 with two β components are given by the following: $\lambda_0 = 0.56$, $\lambda_i = 0.44$, $r_0 = 1$, $s_0 = 1$, $r_1 = 0.45$, $s_1 = 3.01$. For any *p* value threshold T , the mixture model allows us to estimate the rate of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)

TABLE IV
Genes differentially expressed between lrp^+ and lrp^- (control vs. experimental) *E. coli* strains with a p value less than 0.0001 identified with a regularized t test.

The data are presented as the average (mean) and S.D. of four independent gene expression measurements expressed as a fraction of the total hybridization signal (total mRNA) on each DNA microarray filter.

Gene name ^a	Control	Experimental	Control	Experimental	p value	PPDE ($< p$)	-Fold
	<i>mean</i>	<i>mean</i>	<i>S.D.</i>	<i>S.D.</i>			
<i>oppA</i> *	1.62E-03	3.16E-02	7.63E-04	1.03E-02	5.14E-13	1.00000	19.44
<i>lysU</i> *	1.81E-04	1.24E-03	7.48E-05	2.78E-04	8.88E-10	0.99999	6.87
<i>oppB</i> *	7.51E-05	1.14E-03	2.12E-05	3.79E-04	1.02E-09	0.99999	15.12
<i>oppC</i> *	2.01E-04	1.08E-03	2.34E-05	3.61E-04	3.26E-09	0.99998	5.38
<i>oppD</i> *	8.97E-05	6.55E-04	2.76E-05	2.05E-04	2.69E-08	0.99995	7.30
<i>serA</i> *	2.90E-03	6.56E-04	1.14E-03	1.12E-04	4.08E-08	0.99994	-4.41
<i>ftn</i>	2.36E-04	1.38E-03	1.29E-04	5.46E-04	2.27E-07	0.99984	5.84
<i>rmf</i>	5.79E-05	1.47E-03	4.68E-05	3.35E-04	2.75E-07	0.99982	25.43
<i>hdeA</i>	2.40E-04	8.29E-04	8.46E-05	9.90E-05	2.99E-07	0.99982	3.45
<i>ilvP_G::lacY</i> ^{*b}	3.68E-04	1.47E-03	4.56E-05	8.10E-04	3.39E-07	0.99980	3.99
<i>hdeB</i>	3.99E-04	1.98E-03	2.58E-04	5.59E-04	4.50E-07	0.99977	4.96
<i>ilvP_G::lacA</i> ^{*b}	3.31E-04	1.83E-03	1.74E-04	7.48E-04	5.64E-07	0.99974	5.53
<i>artP</i>	6.73E-05	4.23E-04	1.24E-05	1.16E-04	1.42E-06	0.99957	6.28
<i>artI</i>	1.26E-04	5.80E-04	3.79E-05	2.80E-04	2.01E-06	0.99948	4.60
<i>gltD</i> *	5.28E-04	2.74E-05	1.28E-04	1.42E-05	2.34E-06	0.99943	-19.27
<i>ilvG-I</i> *	4.21E-04	9.15E-04	7.55E-05	6.85E-05	4.71E-06	0.99916	2.17
<i>ilvK</i> *	4.16E-04	1.15E-04	1.47E-04	3.22E-05	6.18E-06	0.99903	-3.61
<i>ybeD</i>	1.13E-04	4.01E-04	1.70E-05	1.48E-04	8.55E-06	0.99884	3.55
<i>livH</i> *	4.05E-04	1.24E-04	8.18E-05	5.50E-05	9.07E-06	0.99880	-3.26
<i>uspA</i>	5.42E-04	1.80E-03	3.07E-04	7.43E-04	9.99E-06	0.99874	3.32
<i>pheA</i>	9.11E-05	3.41E-04	3.78E-05	4.17E-05	1.47E-05	0.99844	3.75
<i>grxB</i>	5.95E-05	3.38E-04	1.92E-05	1.07E-04	1.61E-05	0.99836	5.68
<i>b2253</i>	4.24E-04	9.00E-04	8.15E-05	1.50E-04	1.77E-05	0.99827	2.12
<i>hdhA</i>	1.30E-05	2.14E-04	1.22E-05	2.75E-05	1.86E-05	0.99822	16.49
<i>gst</i>	3.44E-06	7.24E-05	4.05E-06	2.41E-05	2.32E-05	0.99800	21.01
<i>oppF</i> *	1.57E-04	4.90E-04	2.82E-05	2.32E-04	2.59E-05	0.99787	3.13
<i>rpoE</i>	1.71E-04	4.35E-04	4.80E-05	7.29E-05	2.67E-05	0.99784	2.55
<i>yhjE</i>	5.44E-04	1.82E-04	6.88E-05	1.20E-04	2.91E-05	0.99773	-2.98
<i>yggB</i>	1.73E-04	4.50E-04	3.57E-05	8.50E-05	2.91E-05	0.99773	2.61
<i>rpoS</i>	3.35E-04	8.77E-04	1.21E-04	3.07E-04	3.03E-05	0.99768	2.62
<i>b1685</i>	3.71E-05	2.64E-04	1.20E-05	1.22E-04	3.66E-05	0.99743	7.10
<i>livM</i> *	6.80E-04	2.74E-04	1.38E-04	1.55E-04	4.24E-05	0.99721	-2.48
<i>rseA</i>	2.41E-04	5.82E-04	5.61E-05	1.32E-04	4.51E-05	0.99712	2.42
<i>ilvP_G::lacZ</i> ^{*b}	8.10E-04	1.81E-03	4.51E-05	6.17E-04	4.60E-05	0.99709	2.24
<i>gdhA</i>	9.16E-05	2.73E-04	1.52E-05	2.16E-05	5.44E-05	0.99681	2.98
<i>livJ</i> *	1.16E-03	2.69E-03	5.03E-04	4.42E-04	5.80E-05	0.99669	2.32
<i>fimA</i> *	3.35E-04	7.82E-05	1.46E-04	3.08E-05	6.35E-05	0.99652	-4.29
<i>trxA</i>	9.05E-05	2.84E-04	2.99E-05	4.29E-05	7.43E-05	0.99621	3.13
<i>ydaR</i>	5.15E-05	2.62E-04	2.61E-05	5.61E-05	8.40E-05	0.99595	5.08

^a Known Lrp-regulated genes are identified by an asterisk.

^b *lac* genes under the control of the Lrp-regulated *ilvP_G* promoter-regulatory region.

consistent with the statistical assumptions made to derive the original set of p values. More precisely, in a general mixture of β models, we have as follows.

$$TP = P(p < T \text{ and change}) = \sum_{i=1}^K \lambda_i \int_0^T \beta(p; r_i, s_i) dp \quad (\text{Eq. 2})$$

$$TN = P(p > T \text{ and no change}) = \lambda_0(1 - T) \quad (\text{Eq. 3})$$

$$FP = P(p < T \text{ and no change}) = \lambda_0 T \quad (\text{Eq. 4})$$

$$FN = P(p > T \text{ and change}) = \sum_{i=1}^K \lambda_i \int_T^1 \beta(p; r_i, s_i) dp \quad (\text{Eq. 5})$$

If we set a threshold T below which p values are considered significant and representative of change, we can estimate the rates of false positives and false negatives. The posterior probability for differential expression (PPDE) then can be calculated for each gene in the experiment with p value p as PPDE(p) according to Equation 6.

$$PPDE(p) = P(\text{change}|p) = \frac{\sum_{i=1}^K \lambda_i \beta(p; r_i, s_i)}{\sum_{i=0}^K \lambda_i \beta(p; r_i, s_i)} = \frac{\sum_{i=1}^K \lambda_i \beta(p; r_i, s_i)}{\lambda_0 + \sum_{i=1}^K \lambda_i \beta(p; r_i, s_i)} \quad (\text{Eq. 6})$$

Alternatively, one can calculate a posterior probability of differential expression PPDE($< p$) for values below a certain threshold p according to Equation 7.

$$PPDE(< p) = P(\text{change}|p < T) = \frac{\sum_{i=1}^K \lambda_i \int_0^T \beta(p; r_i, s_i) dp}{\sum_{i=0}^K \lambda_i \int_0^T \beta(p; r_i, s_i) dp} \quad (\text{Eq. 7})$$

The distribution of p values from our lrp^+ versus lrp^- data is shown in Fig. 4, and a plot of PPDE(p) and PPDE($< p$) values versus p values is shown in Fig. 5. A comparison of the *ad hoc* method for determining the global significance for the differ-

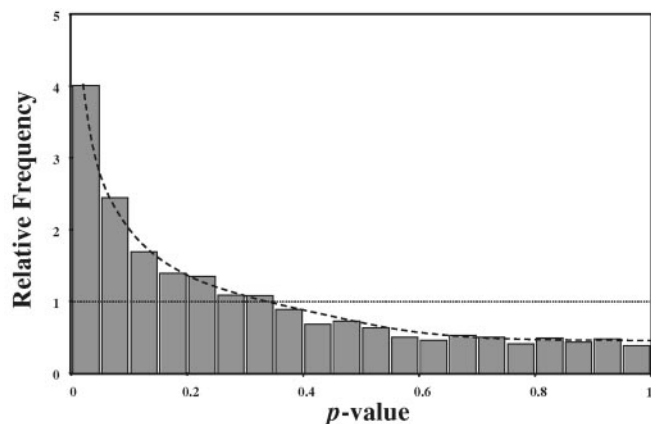


FIG. 4. **Distribution of the p values from the lrp^+ versus lrp^- data.** The fitted model (dashed curve) is a mixture of a β and the uniform distribution (dotted line).

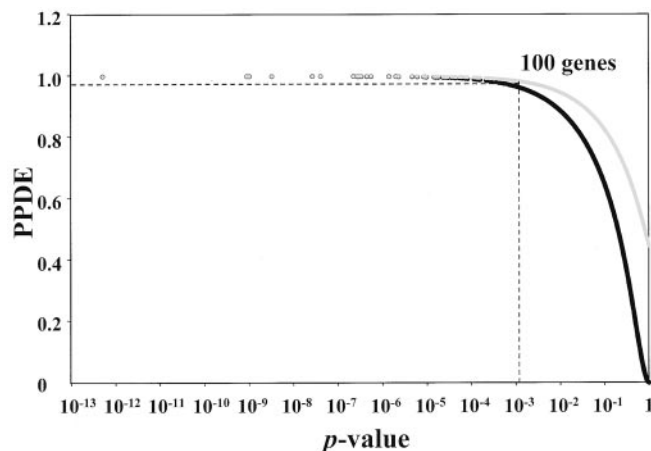


FIG. 5. **Relationship between PPDE and p value.** PPDE ($< p$), gray points; PPDE(p), black points. The dotted line correlates the number of genes differentially expressed with PPDE ($< p$) of 0.97 that are measured with $p < 0.0014$.

ential expression of a given gene and the computational method is presented in Table V. It is satisfying to see that these data compare well.

It is clear from the data of Fig. 5 that for each p value threshold T , there is a tradeoff between the rates of true and false positives. A low conservative p value threshold leads to few FP but may also reduce the TP rate. A large p value threshold ultimately allows one to recover all the TP but at the cost of increasing the FP rate. This fundamental tradeoff is usually captured in statistics by using a receiver operating characteristic curve obtained by plotting the true hit rate (or sensitivity) defined by $TP/(TP + FN)$ versus the false hit rate, $FP/(FP + TN)$ (86). In the mixture model above, a simple calculation shows that for a given p value threshold T ,

$$\frac{FP}{FP + TN} \text{ and } \frac{TP}{TP + FN} = \frac{\sum_{i=1}^K \lambda_i \int_0^T \beta(p; r, s_i) dp}{1 - \lambda_0} \quad (\text{Eq. 8})$$

With two components in the mixture ($K = 1$), the last expression reduces to the following.

$$\frac{TP}{TP + FN} = \int_0^T \beta(p; \gamma_1, s_1) dp \quad (\text{Eq. 9})$$

Thus, for our Lrp data the receiver operating characteristic curve in Fig. 6 is simply the distribution function of the second β component in the mixture. For instance, this curve demonstrates that with a 76% true hit rate we can expect a 20% false hit rate.

*The Functional Classes of Genes Differentially Expressed in lrp^+ and lrp^- *E. coli* Strains*—To facilitate the following discussions, we limit our considerations to the 100 genes differentially expressed with the lowest p value based on a regularized t test. The 100 genes differentially expressed between lrp^+ and lrp^- *E. coli* strains with a p value less than 0.0014 and a PPDE greater than 0.98 are listed in Table VI. In the text we simply refer to the -fold change for each gene. However, as emphasized above, it should be kept in mind that reporting -fold changes is incomplete and can be misleading. For this reason, mean expression levels, standard deviations, p values, and PPDE values for the 39 genes with p values less than 0.0001 are reported in Table IV. Additional statistical data for the remaining 61 genes with a p value less than 0.0014 as well as all genes expressed at a level of above background in all four experiments can be found in the supplemental data (available in the on-line version of this article).

Because the physiological purpose of Lrp is presumed to be the coordination of gene expression levels with the nutritional and environmental conditions of the cell (5, 6), it was pleasing to discover that most of the genes affected by Lrp are ones that encode products involved in small molecular and macromolecule synthesis or degradation, as well as gene systems involved in small molecule transport and environmental stress responses. These genes can be sorted into the functional groups shown in Fig. 7; they also are listed in Table VI and discussed below.

Small Molecule Biosynthesis—Among the genes differentially expressed between lrp^+ and lrp^- strains, 11 are genes required for amino acid biosynthesis. Of these, the *ilvG*, *ilvM*, *leuB*, and *serA* genes are members of operons previously reported to be regulated by Lrp.

The *ilvG* and *ilvM* genes, the first two genes of the *ilvG-MEDA* operon, encode the two subunits of acetohydroxy acid synthase II, one of three isoenzymes catalyzing the first step of the parallel pathway for L-valine and L-isoleucine biosynthesis. We have previously used an *ilvP_G::lacZ* construct to measure β -galactosidase activities in the same isogenic lrp^+ and lrp^- strains employed in this study (7). These results showed that β -galactosidase was increased 2.5-fold in the lrp^- mutant strain. The data reported here are consistent with this earlier report. We have also described the presence of a constitutive internal promoter, *ilvP_E*, located between the *ilvM* and *ilvE* genes of this operon (18). This affect of the internal promoter is apparent in our DNA microarray data; the expression of the operon distal *ilvEDA* genes is decreased only 1.2-fold.

It is interesting that two other genes of the aspartate family of amino acids previously unknown to be regulated by Lrp appear in this list (19). These are *thrL*, the leader polypeptide of the threonine operon (20–23), and *asd*, the structural gene for aspartyl-semialdehyde dehydrogenase (24, 25). This enzyme is involved in the conversion of oxaloacetate to homoserine, a precursor of threonine and isoleucine. These findings suggest the possibility that all of the genes of the aspartate family that convert the TCA cycle intermediate, oxaloacetate, to amino acids might be sensitive to Lrp-mediated regulatory effects.

The *serA* gene encodes phosphoglycerate dehydrogenase, the first enzyme specific for serine biosynthesis. Newman and colleagues (6, 15) have reported that the transcriptional level of *serA* is decreased 6-fold in a lrp^- strain. Our results exhibit a similar transcriptional regulation. Newman and colleagues

TABLE V
Determination of confidence level for differentially expressed genes with a regularized *t* test

<i>p</i> value	No. of genes		% Confidence (<i>ad hoc</i>)	PPDE ($< p$)
	Control vs. control	Control vs. experimental		
< 0.0001	0	39	≈ 100	0.996
< 0.0005	0.25	62	99.6	0.989
< 0.001	0.5	79	99.4	0.985
< 0.005	2	188	98.9	0.964
< 0.01	3.75	268	98.6	0.947

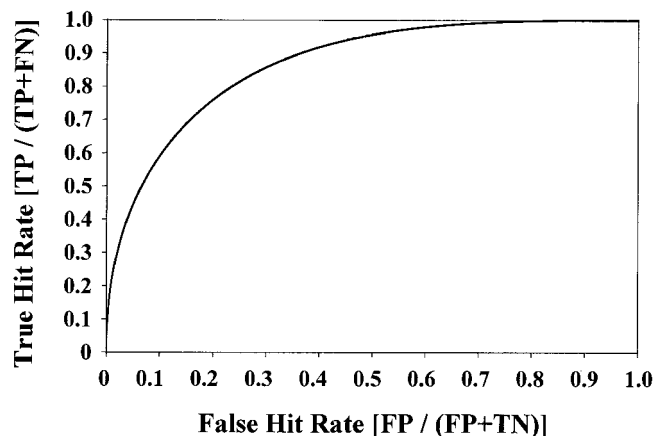


FIG. 6. Receiver operating characteristic curve. This plot correlates the fraction of correctly identified differentially expressed genes (*y* axis) with the fraction of falsely identified differentially expressed genes (*x* axis).

(15, 26) also have reported that the expression level of the *leu* operon is decreased 11-fold in a *lrp*⁻ strain and showed that the growth rate of a *lrp*⁻ strain is increased by adding leucine to the growth medium. On the other hand, Landgraf *et al.* (27) have suggested that Lrp-mediated regulation of the *leu* operon is indirect and reported a much smaller effect (1.4-fold). Our studies agree with those of Landgraf *et al.*

Because these known Lrp-regulated genes identified by our experiments are detected with a high level of measurement accuracy and confidence, we can be similarly confident that the expression of other genes in this group are also members of the Lrp regulon. However, the differential expression of these newly identified genes could be the consequences of either primary or secondary Lrp effects. An obvious way to discern whether or not the operons containing these genes are directly regulated by Lrp would be to search for Lrp binding sites in their promoter-regulatory regions (10). Unfortunately, because of the degeneracy of the consensus Lrp binding sequence, this is not possible. Even when a 3 of 15 mismatch is allowed, 60% of all regions 500 base pairs upstream of all *E. coli* ORFs contain at least one putative Lrp binding site. Thus, it is difficult to determine at this time whether the differential expression of these genes is directly or only indirectly affected by Lrp.

Small Molecule Transport—22 of the 100 genes listed in Table VI are involved in small molecule transport. Of these, 11 have been documented to be regulated by Lrp. Products of the *livJ* and *livKHMGF* genes are components of two transport systems with high affinity for leucine. The *livJ* gene product binds leucine, isoleucine, and valine, whereas the *livK* gene product is specific for leucine alone. These two systems share a set of membrane components, products of the *livHMGF* genes. Haney *et al.* (28) have reported that both of these operons are repressed by Lrp in the presence of high concentration of leucine. Bhagwat *et al.* (29) have reported that in the absence of leucine the expression of the *livJ* gene is unaffected by Lrp and

that the expression of the *livKHMGF* operons is activated approximately 15-fold. Because the experiments reported here were also performed in the absence of leucine, we would expect similar results and, in fact, we observe a 2.5–10-fold activation of the genes for the *livKHMGF* operon. However, our results suggest that Lrp is also responsible for a 2-fold repression of *livJ* under these growth conditions.

The *oppABCDF* operon contains genes encoding a periplasmic binding protein and transport permease proteins for a wide range of tripeptide transport systems. Austin *et al.* (30) have reported that this operon exhibits high constitutive expression in a *lrp*⁻ strain. Accordingly, our results show that the expression of the *oppA* and *oppB* genes is increased 15- and 20-fold, respectively, but that the expression of the *oppC*, *oppD*, and *oppF* genes are increased to a lesser extent. These data suggest the possibility of an unidentified internal promoter between *oppB* and *oppC*.

Four proteins, the *malEFG* and *-K* gene products, are required for maltose uptake in *E. coli*. These four genes are arranged in two operons, *malEFG* and *malK-lamB-malM*. Tchetina *et al.* (31) have reported that transcription of both of these operons are decreased 50–70% in a *lrp*⁻ strain grown in glycerol. Our results demonstrate that the transcription level of both operons also is decreased approximately 80% in a *lrp*⁻ strain grown in a glucose-supplemented minimal MOPS medium. However, because *malE* is the only gene of either operon that passed our statistical cut-off, we cannot be as confident that Lrp also affects the expression of the other genes of these operons (see supplemental data, available in on-line version of this article).

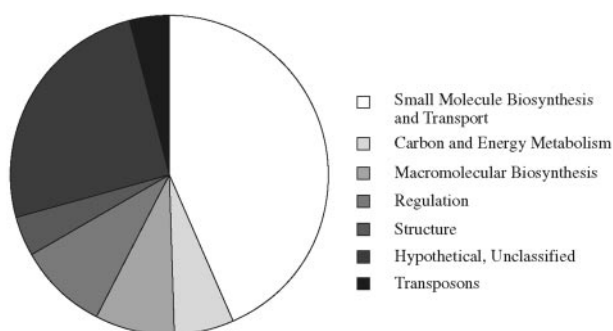
Of the remaining newly identified genes of this class, we find two examples of two genes in the same operon, *artP* and *artI* of the *artPIQMJ* operon and *potH* and *potG* of the *potFGHI* operon. *artP* and *artI* are involved in arginine transport (32). The expression of these genes is increased in the *lrp*⁻ strain 6.3- and 4.6-fold, respectively. The *potH* and *potG* genes are members of the *potFGHI* operon involved in the transport of putrescine (33). The expression of these genes is decreased ~3-fold in the *lrp*⁻ strain. In addition to these previously documented systems, our results suggest that transport systems involved in the transport of various dipeptides, carbohydrates, organic acids, alcohols, and inorganic compounds are also influenced by Lrp (34–42).

Carbon and Energy Metabolism—Besides its influence on the transport of organic acids, Lrp also influences the expression of genes involved in the metabolism of these compounds. For example, the levels of expression of the structural genes for *malP* and *manA* are both altered in a *lrp*⁻ strain (43–45). The remaining genes of this group listed in Table VI are involved either in the catabolism of glucose under aerobic conditions, or alternative carbon sources under anaerobic conditions (46–50).

Macromolecular Biosynthesis—8 of the 100 genes listed in Table VI are involved in macromolecule synthesis. Three of these genes, *hns*, *hupB*, and *himD*, encode proteins that influence the structure and DNA topology of the *E. coli* chromosome. The remaining five are involved in protein synthesis or degradation. Of these, only one has been previously identified as a

TABLE VI
Functional groups for *Lrp*-regulated genes

Small molecule biosynthesis and transport							
<i>Amino acid biosynthesis</i>							
<i>serA</i>	-4.41	<i>ilvG-1</i>	2.17	<i>gdhA</i>	2.98	<i>ilvP_G::lacZ</i>	2.24
<i>leuB</i>	-1.86	<i>ilvG-2</i>	2.04	<i>pheA</i>	3.75	<i>ilvP_G::lacY</i>	3.99
<i>proB</i>	-1.79	<i>ilvM</i>	3.07	<i>dapA</i>	1.89	<i>ilvP_G::lacA</i>	5.53
<i>thrL</i>	3.24	<i>hisG</i>	2.53	<i>asd</i>	1.98		
<i>Co-factor biosynthesis</i>							
<i>folE</i>	2.26	<i>gst</i>	21.01	<i>grxB</i>	5.68	<i>trxA</i>	3.13
<i>Central intermediary metabolism</i>							
<i>gltD</i>	-19.27	<i>hdhA</i>	16.49				
<i>Transport</i>							
<i>livK</i>	-3.61	<i>sbp</i>	-2.16	<i>artP</i>	6.28	<i>oppA</i>	19.44
<i>livH</i>	-3.26	<i>galP</i>	-1.78	<i>artI</i>	4.60	<i>oppB</i>	15.12
<i>livG</i>	-9.90	<i>dppB</i>	-3.01	<i>livJ</i>	2.32	<i>oppC</i>	5.38
<i>livM</i>	-2.48	<i>ascF</i>	-2.80	<i>glnH</i>	2.22	<i>oppD</i>	7.30
<i>potH</i>	-3.27	<i>malE</i>	-1.78	<i>bcp</i>	3.63	<i>oppF</i>	3.13
<i>potG</i>	-2.03			<i>ftn</i>	5.84		
<i>Carbon and energy metabolism</i>							
<i>Carbon compound catabolism</i>							
<i>malP</i>	-2.38	<i>manA</i>	2.75				
<i>Energy metabolism</i>							
<i>nirB</i>	-4.12	<i>glpD</i>	-2.77	<i>ppc</i>	1.98	<i>gltA</i>	2.14
<i>Macromolecular biosynthesis</i>							
<i>DNA structure and synthesis</i>							
<i>hns</i>	2.00	<i>himD</i>	4.09	<i>hupB</i>	2.39		
<i>Translation</i>							
<i>lysU</i>	6.87	<i>clpA</i>	1.75	<i>pepD</i>	1.93	<i>rpmI</i>	1.82
<i>rmf</i>	25.43						
<i>Regulation</i>							
<i>Regulatory</i>							
<i>rpoS</i>	2.62	<i>rpoE</i>	2.55	<i>rseA</i>	2.42	<i>phoP</i>	2.54
<i>Stress response</i>							
<i>sodA</i>	-2.35	<i>ahpC</i>	2.09	<i>dnaK</i>	2.13	<i>uspA</i>	3.32
<i>CpxP</i>	3.16						
<i>Cell structure</i>							
<i>fimA</i>	-4.29	<i>ompT</i>	-4.26	<i>ompX</i>	2.65	<i>slp</i>	5.67
<i>Hypothetical or unclassified</i>							
<i>yhjE</i>	-2.98	<i>yccA</i>	2.54	<i>yhbH</i>	2.75	<i>hdeA</i>	3.45
<i>b0703</i>	-2.78	<i>b0667</i>	1.71	<i>yhiX</i>	5.16	<i>hdeB</i>	4.96
<i>yibJ</i>	-2.56	<i>ydaA</i>	1.87	<i>yggV</i>	3.08	<i>b1685</i>	7.10
<i>b2253</i>	2.12	<i>yeeX</i>	2.43	<i>ydhD</i>	3.16	<i>ydaR</i>	5.08
<i>b2254</i>	1.83	<i>b2294</i>	2.53	<i>ybeD</i>	3.55	<i>yggB</i>	2.61
<i>yadF</i>	2.20	<i>b2595</i>	2.01	<i>b1839</i>	3.96	<i>yafK</i>	2.98
<i>yajC</i>	2.00						
<i>Transposons</i>							
<i>rhsB</i>	-2.76	<i>insB</i> 6	-2.55	<i>tra5_4</i>	-2.01	<i>trs5_11</i>	-1.77

FIG. 7. Distribution of functions for genes differentially expressed between *lrp*⁺ and *lrp*⁻ *Escherichia coli* strains.

Lrp-regulated gene. This gene, *lysU*, encodes one of the two lysyl-tRNA synthetases in *E. coli*. Gazeau *et al.* (51, 52) have reported that this gene is repressed 9-fold by *Lrp*. Under the conditions of our experiments, *Lrp* represses the expression of the *lysU* gene 7-fold. Of the four newly identified genes of this group, two, *clpA* and *pepD*, are proteases involved in protein degradation (53, 54). The remaining two genes, *rpmI* and *rmf*, encode ribosome-associated proteins (55–58).

The *hns*, *hupB*, and *himD* genes are the structural genes for the H-NS, HU, and IHF proteins that are important for the condensation of the chromosome into a nucleoid structure, for

restraining negative supercoils, and in several cases for the regulation of gene expression (3, 59–61). In each case, these genes are repressed by *Lrp*. It is likely that the effect of *Lrp*-mediated effects on the expression levels of these global regulatory gene products might be responsible for many secondary changes in gene expression levels observed in *lrp*⁻ strains.

Regulatory Proteins of Stress Responses—The expression levels of several proteins involved in cellular adaptation to nutritional and environmental assaults are increased 2–3-fold in the *lrp*⁻ strain. These include the alternative sigma factors *rpoS* and *rpoE*. The *rpoS* sigma factor, σ^{38} , is a central regulator of many stationary phase-responsive genes. Although it is induced to high levels in early stationary phase cells, it also is expressed, albeit at a lower level, during the exponential growth phase, where it functions as a general stress response element essential for prolonged cell survival (62, 63). It is involved in the induction of several genes important for osmotic, oxidative, heat, and DNA damage stress responses (64). The *rpoE* gene encodes another sigma element, σ^{24} , that also is expressed at a higher level in the *lrp*⁻ strain. Although the major functions of coping with thermal stress are encoded by genes transcribed by σ^{32} , genes transcribed by σ^{24} are necessary for survival under extreme temperature stress conditions (65). Interestingly, although the expression level of the *rpoH* gene for σ^{32} is unaffected in the *lrp*⁻ strain, several genes regulated by this sigma factor, such as *dnaK*, *dnaJ*, *clpA*, *clpB*,

TABLE VII

Genes differentially expressed between *lrp*⁺ and *lrp*⁻ (control vs. experimental) *E. coli* strains with four measurements below background obtained from *lrp*⁺ or *lrp*⁻ strains

The data are presented as the average (mean) and S.D. of four independent gene expression measurements expressed as a fraction of the total hybridization signal (total mRNA) on each DNA microarray filter. NA, not available.

Gene name ^a	Control	Experimental	Control	Experimental	Coefficient of variance
	<i>mean</i>	<i>mean</i>	<i>S.D.</i>	<i>S.D.</i>	
<i>yceB</i>	0.00E-00	3.27E-05	NA	3.05E-06	0.09
<i>gltB</i> *	1.28E-04	0.00E-00	2.37E-05	NA	0.19
<i>msyB</i>	0.00E-00	3.86E-05	NA	9.57E-06	0.25
<i>gcvH</i> *	2.03E-05	0.00E-00	5.43E-06	NA	0.27
<i>osmC</i>	0.00E-00	7.05E-05	NA	3.03E-05	0.43
<i>ilvH</i> *	4.13E-05	0.00E-00	2.11E-05	NA	0.51
<i>yaiB</i>	0.00E-00	2.15E-04	NA	1.09E-04	0.51
<i>yacL</i>	0.00E-00	1.99E-05	NA	1.28E-05	0.64
<i>yljA</i>	0.00E-00	1.30E-04	NA	8.67E-05	0.67
<i>b1720</i>	0.00E-00	1.61E-05	NA	1.17E-05	0.72
<i>fhuF</i>	3.09E-05	0.00E-00	2.69E-05	NA	0.87
<i>ibpA</i>	0.00E-00	3.43E-06	NA	3.20E-06	0.93
<i>fimC</i>	6.07E-05	0.00E-00	5.81E-05	NA	0.96
<i>stpA</i> *	2.15E-05	0.00E-00	2.07E-05	NA	0.96
<i>ribE</i>	0.00E-00	5.63E-06	NA	5.41E-06	0.96
<i>yhiE</i>	0.00E-00	1.40E-05	NA	1.38E-05	0.98
<i>b1431</i>	0.00E-00	5.04E-06	NA	5.32E-06	1.06
<i>b1438</i>	0.00E-00	6.39E-06	NA	7.28E-06	1.14
<i>yedL</i>	4.74E-06	0.00E-00	5.61E-06	NA	1.18
<i>fimB</i> *	0.00E-00	7.76E-06	NA	9.33E-06	1.20
<i>kdtB</i>	1.64E-05	0.00E-00	2.07E-05	NA	1.27
<i>relE</i>	0.00E-00	1.63E-05	NA	2.57E-05	1.58
<i>hisH</i>	0.00E-00	2.57E-05	NA	4.17E-05	1.62

^a Known Lrp-regulated genes are identified by an asterisk.

clpP, *htpG*, *htpX*, *gapA*, and *grpE*,³ all exhibit 2–3-fold increased expression levels in the *lrp*⁻ strain.

The *rpoE* and *rseA* genes are both members of *rpoE-rseABC* operon, and the *resA* gene product is a negative regulator of this operon (66). The remaining genes of this group, *phoP*, *cpxP*, *aspC*, and *uspA*, also up-regulated 2–3-fold, are similarly involved in stress responses. *phoP* is a regulatory protein involved in a variety of environmental stress signals including magnesium starvation and nutritional deprivation (67). *cpxP* encodes a periplasmic protein important for PH tolerance (68, 69). *ahpC* and *uspA* encode proteins involved in the oxidative stress response (70–73). *sodA* encodes a superoxide dismutase also required for survival during oxidative stress conditions (37, 74). However, the expression of this last gene of this group is decreased 2.4-fold in a *lrp*⁻ strain.

Cell Structure—Of the four genes of this group listed in Table VI, only the *fimA* gene has been reported to be regulated by Lrp. The gene product of the *fimA* gene is the major fimbrial subunit of type I pili. The expression of this gene is controlled by a cis-acting DNA element (switch). Several reports have shown that switching frequency is reduced in IHF and *lrp*⁻ strains (75, 76). In agreement with these reports, our data show that *fimA* expression is reduced 4.3-fold in a *lrp*⁻ strain.

The remaining genes of this group, *ompT*, *ompX*, and *slp*, are outer membrane proteins involved in nutritional or environmental stress responses. The *ompT* gene encodes an outer membrane endopeptidase associated with pathogenicity in certain Gram-negative bacteria (77). Its activity is increased during conditions of temperature stress (78). *ompX* encodes an outer membrane protein required for σ^E activity during temperature stress in some *E. coli* strains (79). Finally, *slp* encodes the starvation lipoprotein induced during nutritional deprivation (80).

Examples of Genes Only Expressed in Either Strain IH-G2490 or Strain IH-G2491—Only those genes exhibiting an expression level greater than zero in all experiments were used for statistical analysis as described above. Gene measurements containing zero expression values were set aside and discussed here. Among this set of genes, 23 genes with zero expression values for all measurements of one genotype, and all values greater than zero for all measurements of another genotype for each experiment, were identified. The significance of these results (Table VII) was analyzed by ranking these genes in ascending order according to their coefficients of variance of the four greater than zero measurements.

Four of the 23 genes in Table VII are known Lrp-regulated genes contained in the *gltBDF*, *ilvIH*, *gcvTHP*, and *stpA* operons. The genes of the *gltBDF* operon encode a regulatory protein (*gltF*) and the two subunits of glutamate synthase (*gltB* and *D*), an enzyme involved in ammonia assimilation. Ernsting *et al.* (81, 82) have reported that this enzyme activity is very low or missing in a *lrp*⁻ strain. In agreement with this report, our results show that the mRNA level of *gltB* is below detection (Table VII) and *gltD* mRNA (Table VI) is reduced 19-fold in the *lrp*⁻ strain. On the other hand, the mRNA level of the regulatory *gltF* gene is reduced only 1.9-fold (see supplemental data, available in the on-line version of this article). This result suggests the presence of an internal, *lrp*⁻ independent promoter between the *gltD* structure gene and *gltF* regulatory gene of this operon.

Wang *et al.* (83) have shown that the *ilvIH* operon that encodes acetohydroxy acid synthase III of the branched chain amino acid pathway is repressed 30-fold in a *lrp*⁻ strain. The repressed level of transcripts of this operon are undetectable above background in the experiments reported here (Table VII). Lin *et al.* (15, 84, 85) reported that the *gcvTHP* operon that encodes proteins that cleave glycine to produce one-carbon units and ammonia is repressed 20-fold in a *lrp*⁻ strain. We find the transcript for one of the genes of this operon (*gcvH*) is undetectable (Table VII), and the remaining two genes measured with *p* values higher than our statistical cut-off level are

³ Of these genes only the *dnaK* and *clpA* genes pass our stringent statistical test. However, none of the remaining genes in this list possesses a *p* value greater than 0.01 (see supplemental data available in the on-line version of this article).

both repressed (see supplemental data, available in the on-line version of this article). The transcriptional regulation of the *stpA* gene, encoding the *E. coli* H-NS-like protein StpA, is regulated by a variety of environmental conditions and several global transcription factors, including Lrp. Free and Dorman (87) have shown the transcription of *stpA* is significantly decreased in a *lrp*⁻ strain growing in minimal medium. Our results demonstrate that the expression level of the *stpA* gene is not detected in a *lrp*⁻ strain.

Nylon Filter Data Versus Affymetrix GeneChip Data—When different array formats are used that require different target preparation methods, the magnitudes and sources of experimental errors are surely different. This raises the question of whether or not results obtained from experiments performed with different DNA array formats can be compared with one another, or indeed even whether comparable results can be obtained. To address this question, we have assessed the results obtained from DNA array experiments performed with pre-synthesized nylon filters hybridized with ³³P-labeled cDNA targets prepared from total RNA with random hexamer primers and *in situ* synthesized Affymetrix GeneChips hybridized with enriched, biotin-labeled, mRNA targets obtained from the same total RNA preparations.

For the GeneChip experiments, the exact same four control and experimental pairs of pooled RNA preparations used in the *lrp*⁺ versus *lrp*⁻ nylon filter experiments described above (Fig. 1) were used for hybridization to four pairs of *E. coli* Affymetrix GeneChips. However, because of economic considerations, each experiment was not performed in duplicate; hence, only one measurement for each gene was obtained on each chip. Thus, instead of having four measurements for each gene expression level for each experiment (Fig. 1), only one measurement was obtained from each GeneChip (Fig. 2). On the other hand, this single measurement is the average of the difference between hybridization signals from ~15 perfect match and mismatch probe pairs.⁴ Although these are not equivalent to duplicate measurements because different probes are used, these data do increase the reliability of each gene expression level measurement.

Because only one measurement from one GeneChip was obtained for each genotype for each experiment, it was not possible to distinguish sources of error contributed by differences among GeneChips and from differences among target preparations as we have previously reported for the filter data (10). Nevertheless, it was possible to use the *ad hoc* control versus control and PPDE computational methods to compare data among the four control GeneChips hybridized with independent biotin-labeled mRNA targets from *E. coli* strain IH-G2490 (*lrp*⁺). These methods were used to estimate the number of false positives expected at given *p* value thresholds. These results for the GeneChip data, as well as the nylon filter data, are presented in Tables VIII and V, respectively.

It is clear from these results that the filter data identifies more differentially expressed genes with lower *p* values and higher confidence levels than the GeneChip data. This is not surprising because, as explained above, each gene measurement level in the filter data set is the average of four duplicate measurements from two separate filters, whereas each gene measurement in the GeneChip data set is based on a single measurement from each experiment. In fact, when the top 100 genes with the lowest *p* values from the nylon filter and GeneChip experiments are compared, only 17 genes are common to both lists. This lack of correspondence is likely the result of the

TABLE VIII
Differential gene expression data for Affymetrix GeneChip experiments using CyberT with a regularized *t* test

Calculated with 3,515 control and experimental gene expression measurements (AD values from *.CEL file with negative values converted to 0) containing four non-zero values for four experiments.

<i>p</i> values	No. of genes ^a		% Confidence (<i>ad hoc</i>)	PPDE (< <i>p</i>)
	Control vs. control	Control vs. experimental		
< 0.0001	1	21	95.2	0.985
< 0.0005	2.75	32	91.4	0.943
< 0.001	5.5	37	85.1	0.903
< 0.005	19	54	64.8	0.672
< 0.01	31.3	62	49.6	0.527
< 0.05	164	140		0.195

^a Calculated by averaging the control or experimental measurements and comparing experiments 1 and 3 versus 2 and 4 or experiments 1 and 4 versus 2 and 3.

TABLE IX
Number of differentially expressed genes identified by Affymetrix Microarray Suite software 4.0

No. of replicates	No. of differentially expressed genes ⁿ
1	416–682
2	118–184
3	68–95
4	55

greater variance among the GeneChip measurements and the fact that fundamentally different DNA array formats are compared. However, when a Bayesian statistical framework is applied to the analysis of each data set, the correspondence is nearly doubled and 27 genes are found to be common to both lists. These results further strengthen the conclusions of Long *et al.* (12) that statistical analyses performed with a Bayesian prior identify genes that are up- or down-regulated more reliably than approaches based on a simple *t* test when only a few experimental replications are possible.

The GeneChip results described above were obtained from raw data that were background subtracted and normalized to the total signal on each DNA array, and analyzed with the CyberT statistical software. Affymetrix has developed its own empirical algorithms for the analysis of GeneChip data that are commercially available in a software package, Microarray Suite 4.0. Below we compare the identification of differentially expressed genes identified with the CyberT and Microarray Suite 4.0 software.

Because the Affymetrix software allows the comparison of only one GeneChip pair at a time, it was run on each of the four independent experiments comparing *lrp* genotypes. Each comparison identified between 500 and 700 genes that the Affymetrix software calls as marginally increased or decreased, or increased or decreased (Table IX). However, filtering the results from these four independent experiments identified only 55 genes that the Affymetrix software called differentially expressed in all four experiments. Remarkably, comparison of these 55 genes to the 55 genes exhibiting the lowest *p* values identified by the CyberT software employing a Bayesian statistical framework revealed 39 genes in common with both lists. Among these were 21 known Lrp-regulated genes.

These results illustrate several important points. First, they stress the importance of replication when only two conditions are compared. Little can be learned about those genes regulated by Lrp from the analysis of only one experiment with one GeneChip pair because an average of 600 genes were identified as differentially expressed, only 55 of which can be reproduced in four independent experiments. Furthermore, in the absence of statistical analysis, it is not possible to determine the confi-

⁴ The number of probe pairs for each ORF and inter-ORF regions ranges from 3 to 298.

dence level and rank the reliability of any differentially expressed gene measurement identified with the Affymetrix software. This is, of course, important for prioritizing genes to be examined by additional experimental approaches. Finally, and most importantly, these results demonstrate that remarkably similar answers can be obtained from fundamentally different DNA microarray formats when the raw data from each set of experiments are analyzed by the statistical methods employed here.

Summary—It is indicative of the power of gene expression profiling experiments that two thirds of the genes measured here with a 97% global confidence were previously unknown to be members of the Lrp regulatory network. Furthermore, nearly one third of these genes are genes of unknown function. As more experiments of this type are performed, as more functions are assigned to the gene products of hypothetical ORFs, and as bioinformatics methods to identify degenerate protein binding sites typical of proteins that bind to many DNA sites are developed, an even clearer picture of the Lrp genetic regulatory network in *E. coli* will emerge. However, even at this early stage in the development and execution of DNA array technologies and data analysis methods, the results presented here support previous suggestions that the physiological role of Lrp is to monitor the nutritional state of the cell to adjust its metabolism to changing nutritional conditions and, in cooperation with other regulatory networks, to coordinate these changes with the physical environment of the cell.

Acknowledgments—We acknowledge the many helpful discussions, advice, and computational assistance received from Dr. Suzanne B. Sandmeyer and the members of the Functional Genomics Group (University of California, Irvine). We acknowledge expert technical assistance from Dr. Stuart M. Arfin and Elaine Ito and from Dr. Denis Heck and Kim Nguyen of the DNA Microarray Core Facility (University of California, Irvine). We are also grateful to Cambridge University Press for permission to reproduce materials that will appear in a forthcoming book by P. B. and G. W. H. titled *DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*.

REFERENCES

- Neidhardt, F. C., and Savageau, M. A. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 1310–1324, ASM Press, Washington, D. C.
- Schaechter, M. (2001) *Microbiol. Mol. Biol. Rev.* **65**, 119–130
- Hatfield, G. W., and Benham, C. J. (2002) *Annu. Rev. Genet.* **36**, 175–203
- Ideker, T., Galitski, T., and Hood, L. (2001) *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372
- Calvo, J. M., and Matthews, R. G. (1994) *Microbiol. Rev.* **58**, 466–490
- Newman, E. B., Lin, R. T., and D'Ari, R. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 1513–1525, ASM Press, Washington, D. C.
- Rhee, K. Y., Parekh, B. S., and Hatfield, G. W. (1996) *J. Biol. Chem.* **271**, 26499–26507
- Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) *J. Bacteriol.* **119**, 736–747
- Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehle, M. M., Paegle, E. S., and Hatfield, G. W. (2000) *J. Biol. Chem.* **275**, 29672–29684
- Li, C., and Wong, W. H. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 31–36
- Long, A. D., Mangalam, H. J., Chan, B. Y., Tolleri, L., Hatfield, G. W., and Baldi, P. (2001) *J. Biol. Chem.* **276**, 19937–19944
- Baldi, P., and Long, A. D. (2001) *Bioinformatics* **17**, 509–519
- Shao, Z., Lin, R. T., and Newman, E. B. (1994) *Eur. J. Biochem.* **222**, 901–907
- Lin, R., D'Ari, R., and Newman, E. B. (1992) *J. Bacteriol.* **174**, 1948–1955
- Schena, M., Shalun, D., Davis, R. W., and Brown, P. O. (1995) *Science* **270**, 467–470
- Allison, D. B., Gadbury, G. L., Heo, M., Fernandez, J. R., Lee, C. K., Prolla, T. A., and Weindrich, R. (2002) *Comput. Stat. Data Anal.* **39**, 1–20
- Wek, R. C., and Hatfield, G. W. (1986) *Nucleic Acids Res.* **14**, 2763–2777
- Patte, J. C. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 528–541, ASM Press, Washington, D. C.
- Gardner, J. F. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1706–1710
- Gardner, J. F. (1982) *J. Biol. Chem.* **257**, 3896–3904
- Lynn, S. P., Burton, W. S., Donohue, T. J., Gould, R. M., Gumpport, R. I., and Gardner, J. F. (1987) *J. Mol. Biol.* **194**, 59–69
- Saint-Girons, I., and Margarita, D. (1978) *Mol. Gen. Genet.* **162**, 101–107
- Haziza, C., Stragier, P., and Patte, J. C. (1982) *EMBO J.* **1**, 379–384
- Porco, A., and Isturiz, T. (1991) *Acta Cient. Venez.* **42**, 270–275
- Ambartsoumian, G., D'Ari, R., Lin, R. T., and Newman, E. B. (1994) *Microbiology* **140**, 1737–1744
- Landgraf, J. R., Boxer, J. A., and Calvo, J. M. (1999) *J. Bacteriol.* **181**, 6547–6551
- Haney, S. A., Platko, J. V., Oxender, D. L., and Calvo, J. M. (1992) *J. Bacteriol.* **174**, 108–115
- Bhagwat, S. P., Rice, M. R., Matthews, R. G., and Blumenthal, R. M. (1997) *J. Bacteriol.* **179**, 6254–6263
- Austin, E. A., Andrews, J. C., and Short, S. A. (1989) *Abstract Molecular Genetics of Bacteria and Phage*, p. 153, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Tchetina, E., and Newman, E. B. (1995) *J. Bacteriol.* **177**, 2679–2683
- Wissenbach, U., Six, S., Bongaerts, J., Ternes, D., Steinwachs, S., and Uden, G. (1995) *Mol. Microbiol.* **17**, 675–686
- Pistocchi, R., Kashiwagi, K., Miyamoto, S., Nukui, E., Sadakata, Y., Kobayashi, H., and Igarashi, K. (1993) *J. Biol. Chem.* **268**, 146–152
- Sirko, A., Zatyka, M., Sadowy, E., and Hulanicka, D. (1995) *J. Bacteriol.* **177**, 4134–4136
- Macpherson, A. J., Jones-Mortimer, M. C., Horne, P., and Henderson, P. J. (1983) *J. Biol. Chem.* **258**, 4390–4396
- Bauminger, E. R., Treffry, A., Hudson, A. J., Hechel, D., Hodson, N. W., Andrews, S. C., Levi, S., Nowik, I., Arosio, P., Guest, J. R., et al. (1994) *Biochem. J.* **302**, 813–820
- Touati, D., Jacques, M., Tardat, B., Bouchard, L., and Despied, S. (1995) *J. Bacteriol.* **177**, 2305–2314
- Ghrist, A. C., and Stauffer, G. V. (1998) *J. Bacteriol.* **180**, 1803–1807
- Abouhamad, W. N., Manson, M., Gibson, M. M., and Higgins, C. F. (1991) *Mol. Microbiol.* **5**, 1035–1047
- Jeong, W., Cha, M. K., and Kim, I. H. (2000) *J. Biol. Chem.* **275**, 2924–2930
- King, N. D., and O'Brian, M. R. (1997) *J. Bacteriol.* **179**, 1828–1831
- Hall, B. G., and Xu, L. (1992) *Mol. Biol. Evol.* **9**, 688–706
- Guest, J. R., and Roberts, R. E. (1983) *J. Bacteriol.* **153**, 588–596
- Fraenkel, D. G. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 189–198, ASM Press, Washington, D. C.
- Debarbouille, M., Cossart, P., and Raibaud, O. (1982) *Mol. Gen. Genet.* **185**, 88–92
- Spencer, M. E., and Guest, J. R. (1982) *J. Bacteriol.* **151**, 542–552
- Cronan, J. E. J., and Laporte, D. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 206–216, ASM Press, Washington, D. C.
- Cole, J. A., Newman, B. M., and White, P. (1980) *J. Gen. Microbiol.* **120**, 475–483
- Weichert, D., Lange, R., Henneberg, N., and Hengge-Aronis, R. (1993) *Mol. Microbiol.* **10**, 407–420
- Oh, M. K., and Liao, J. C. (2000) *Biotechnol. Prog.* **16**, 278–286
- Gazeau, M., Delort, F., Dessen, P., Blanquet, S., and Plateau, P. (1992) *FEBS Lett.* **300**, 254–258
- Gazeau, M., Delort, F., Fromant, M., Dessen, P., Blanquet, S., and Plateau, P. (1994) *J. Mol. Biol.* **241**, 378–389
- Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., and Gottesman, S. (1990) *J. Biol. Chem.* **265**, 12536–12545
- Miller, C. G., and Schwartz, G. (1978) *J. Bacteriol.* **135**, 603–611
- Lesage, P., Chiaruttini, C., Graffe, M., Dondon, J., Milet, M., and Springer, M. (1992) *J. Mol. Biol.* **228**, 366–386
- Yamagishi, M., Matsushima, H., Wada, A., Sakagami, M., Fujita, N., and Ishihama, A. (1993) *EMBO J.* **12**, 625–630
- Maki, Y., Yoshida, H., and Wada, A. (2000) *Genes Cells* **5**, 965–974
- Ishihama, A. (1999) *Genes Cells* **4**, 135–143
- Yasuzawa, K., Hayashi, N., Goshima, N., Kohno, K., Imamoto, F., and Kano, Y. (1992) *Gene (Amst.)* **122**, 9–15
- Ussery, D., Larsen, T. S., Wilkes, K. T., Friis, C., Worning, P., Krogh, A., and Brunak, S. (2001) *Biochimie* **83**, 201–212
- Laurent-Winter, C., Ngo, S., Danchin, A., and Bertin, P. (1997) *Eur. J. Biochem.* **244**, 767–773
- Zinser, E. R., and Kolter, R. (2000) *J. Bacteriol.* **182**, 4361–4365
- Levinthal, M., and Powner, T. (1996) *Res. Microbiol.* **147**, 333–342
- Hengge-Aronis, R. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 1497–1512, ASM Press, Washington, D. C.
- Gross, C. A. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) Vol. 1, 2nd Ed., pp. 1382–1399, ASM Press, Washington, D. C.
- Missiakas, D., Mayer, M. P., Lemaire, M., Georgopoulos, C., and Raina, S. (1997) *Mol. Microbiol.* **24**, 355–371
- Bearson, S., Bearson, B., and Foster, J. W. (1997) *FEMS Microbiol. Lett.* **147**, 173–180
- Raivio, T. L., Popkin, D. L., and Silhavy, T. J. (1999) *J. Bacteriol.* **181**, 5263–5272
- Danese, P. N., and Silhavy, T. J. (1998) *J. Bacteriol.* **180**, 831–839

70. Cha, M. K., Kim, H. K., and Kim, I. H. (1995) *J. Biol. Chem.* **270**, 28635–28641
71. Blankenhorn, D., Phillips, J., and Slonczewski, J. L. (1999) *J. Bacteriol.* **181**, 2209–2216
72. Ferrante, A. A., Augliera, J., Lewis, K., and Klibanov, A. M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7617–7621
73. Nystrom, T., and Neidhardt, F. C. (1994) *Mol. Microbiol.* **11**, 537–544
74. Walkup, L. K., and Kogoma, T. (1989) *J. Bacteriol.* **171**, 1476–1484
75. Blomfield, I. C. (2001) *Adv. Microb. Physiol.* **45**, 1–49
76. Blomfield, I. C., Calie, P. J., Eberhardt, K. J., McClain, M. S., and Eisenstein, B. I. (1993) *J. Bacteriol.* **175**, 27–36
77. Stathopoulos, C. (1998) *Membr. Cell Biol.* **12**, 1–8
78. Gill, R. T., DeLisa, M. P., Shiloach, M., Holoman, T. R., and Bentley, W. E. (2000) *J. Mol. Microbiol. Biotechnol.* **2**, 283–289
79. Mecsas, J., Welch, R., Erickson, J. W., and Gross, C. A. (1995) *J. Bacteriol.* **177**, 799–804
80. Alexander, D. M., and St John, A. C. (1994) *Mol. Microbiol.* **11**, 1059–1071
81. Ernsting, B. R., Atkinson, M. R., Ninfa, A. J., and Matthews, R. G. (1992) *J. Bacteriol.* **174**, 1109–1118
82. Ernsting, B. R., Denninger, J. W., Blumenthal, R. M., and Matthews, R. G. (1993) *J. Bacteriol.* **175**, 7160–7169
83. Wang, Q., and Calvo, J. M. (1993) *J. Mol. Biol.* **229**, 306–318
84. Stauffer, L. T., and Stauffer, G. V. (1999) *Microbiology* **145**, 569–576
85. Stauffer, L. T., and Stauffer, G. V. (1994) *J. Bacteriol.* **176**, 6159–6164
86. Baldi, P., Brunak, S., Chauvin, Y., and Nielsen, H. (2000) *Bioinformatics* **16**, 412–424
87. Free, A., and Dorman, C. J. (1997) *J. Bacteriol.* **179**, 909–918