DMEM. H1299 cells were a gift from J. Chen and were grown in RPMI. All transfections were carried out using Lipofectamine 2000 (Invitrogen), Oligofectamine (Invitrogen) or Geneporter 2 (Gene Therapy Systems) according to manufacturer's recommendations. To assess the effect of COP1 on steady-state levels of p53, Saos-2 cells were transfected with increasing amounts of Flag–COP1 or Flag–COP1ΔRING with 250 ng pcDNA3.1+p53, and U2-OS cells were transfected with or without increasing amounts (0.5, 1 and 2 μg) of pCMV–Flag–COP1 or pCMV–Flag–COP1ΔRING and treated with 50 μM ALLN for 6 h before cell collection where indicated. For reporter assays, Saos-2 or H1299 cells were transiently transfected with 150 ng of pcDNA3.1+p53 or pcDNA3.1+p53R175H, 100 ng of *p21*–Luc, *bax*–Luc, *COP1*–Luc, *COP1*mut–Luc or NS–Luc, and 10 ng of pCMV–β-Gal, with or without increasing amounts (0.5, 1 and 2 μg) of pCMV–Flag–COP1 or pCMV–Flag–COP1ΔRING. Luciferase assays were carried out according to manufacturer's instructions (Promega). For p53-induced cell-death assays, Saos-2 cells were transiently transfected with 1 μg of enhanced green fluorescent protein (EGFP) and 5 μg of pcDNA3.1+, pcDNA3.1+p53, pCMV–Flag–COP1 or pcDNA3.1+p53, and 15 μg of pCMV–Flag–COP1 for 48 h. Cells were harvested and stained with propidium iodide for analysis by fluorescence-activated cell sorting (FACS). COP1 siRNA1 (AACUGACCAA GAUAACCUUGA), COP1 siRNA1 inverted (AAAGUUCCAAUAGAACCAGUC), COP1 siRNA2 (AAGACUUGGAGCAGUGUUACU), COP1 siRNA3 (AAGAGGUGUUGGAG UGUUGAC), Pirh2 siRNA1 (AACTGTGGAATTTGTAGG), Pirh2 B inverted (AAGGAUGUUUAAGGUGUCAA), Pirh2 siRNA2 (AAUGUAACUUAUGCCUAGCUA), Pirh2 siRNA2 inverted (AAAUCGAUCCGUAUUCAAUGU) and MDM2 (AAGGAAUUUAGACAACCUGAA) siRNA oligonucleotides with 3′ dTdT overhangs were synthesized by Genentech or Dharmacon. Control siRNA in experiments refers to a mixture of inverted siRNA oligonucleotides. U2-OS, H1299, Saos-2 and BJ cells were transfected with siRNA oligonucleotides three times at 24–36 h intervals and expanded as necessary to prevent contact inhibition.

Immunoprecipitation, GST pull-down assays and pulse–chase analysis

Cells were lysed in immunoprecipitation (IP) lysis buffer (1% Triton X-100, 150 mM NaCl, 50 mM Tris, pH 7.4, and protease inhibitor mix) or radioimmunoprecipitation assay (RIPA) buffer (0.1% SDS, 1% NP-40, 150 mM NaCl, 0.5% deoxycholate, 50 mM Tris, pH 7.4, and protease inhibitor mix), pre-cleared and immunoprecipitated with target antibody and protein A/G PLUS beads. Identification of COP1-interacting proteins was carried out as previously described[21], except that U2-OS cells stably expressing Flag–COP1 were generated. GST pull-down assays were carried out with GST or GST–p53 combined with *in vitro* translated HA–COP1 in PBST (PBS with 0.1% Tween 20) and incubated on ice for 1 h. GST-bound proteins were subject to SDS–PAGE and immunoblot with anti-HA and anti-GST. IPs were washed in lysis buffer with high salt as required. Pulse–chase experiments were carried out as previously described[15], except that HEK293T cells were transfected with pCMV–Flag6a or pCMV–Flag–COP1 for 24 h, and U2-OS cells were transfected with siRNA oligonucleotides as indicated.

In vitro ubiquitination assays

For *in vitro* ubiquitination reactions, *in vitro*-translated p53 was immunoprecipitated with anti-p53 (DO-1 and FL-393) and washed five times with IP lysis buffer, and reactions were carried out on protein A/G beads. Ten micrograms of Flag–ubiquitin (Sigma), 20 ng of UbcH5b (A.G. Scientific), 20 ng of rabbit E1 (Sigma) and 500 ng of GST–COP1 (E3), which was pre-incubated with 20 μM $ZnCl_2$ for 30 min at room temperature, were incubated in a buffer containing 50 mM Tris, pH 7.5, 2 mM ATP, 5 mM $MgCl_2$, 20 μM $ZnCl_2$ and 2 mM DTT. After incubation for 2 h at 30 °C with gentle agitation, reactions were boiled in PBST with 1% SDS for 5 min and reduced to 0.1% SDS with PBST for re-immunoprecipitation with anti-p53 (DO-1 and FL-393). Finally, samples were subjected to SDS–PAGE followed by immunoblotting with anti-Flag–HRP (M2) to detect ubiquitinated species of p53.

Received 23 December 2003; accepted 29 March 2004; doi:10.1038/nature02514.
Published online 21 April 2004.

1. Seo, H. S. *et al.* LAF1 ubiquitination by COP1 controls photomorphogenesis and is stimulated by SPA1. *Nature* **424,** 995–999 (2003).
2. Ang, L. H. *et al.* Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of *Arabidopsis* development. *Mol. Cell* **1,** 213–222 (1998).
3. Yi, C., Wang, H., Wei, N. & Deng, X. W. An initial biochemical and cell biological characterization of the mammalian homologue of a central plant developmental switch, COP1. *BMC Cell Biol.* **3,** 30 (2002).
4. Neff, M. M., Fankhauser, C. & Chory, J. Light: an indicator of time and place. *Genes Dev.* **14,** 257–271 (2000).
5. Ma, L. *et al.* Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell* **13,** 2589–2607 (2001).
6. Ma, L. *et al.* Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in *Arabidopsis. Plant Cell* **14,** 2383–2398 (2002).
7. Hardtke, C. S. & Deng, X. W. The cell biology of the COP/DET/FUS proteins. Regulating proteolysis in photomorphogenesis and beyond? *Plant Physiol.* **124,** 1548–1557 (2000).
8. Bianchi, E. *et al.* Characterization of human constitutive photomorphogenesis protein 1, a RING finger ubiquitin ligase that interacts with Jun transcription factors and modulates their transcriptional activity. *J. Biol. Chem.* **278,** 19682–19690 (2003).
9. Oren, M. Decision making by p53: life, death and cancer. *Cell Death Differ.* **10,** 431–442 (2003).
10. Jin, S. & Levine, A. J. The p53 functional circuit. *J. Cell Sci.* **114,** 4139–4140 (2001).
11. Shimizu, H. & Hupp, T. R. Intrasteric regulation of MDM2. *Trends Biochem. Sci.* **28,** 346–349 (2003).
12. Honda, R., Tanaka, H. & Yasuda, H. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.* **420,** 25–27 (1997).
13. Kubbutat, M. H., Jones, S. N. & Vousden, K. H. Regulation of p53 stability by Mdm2. *Nature* **387,** 299–303 (1997).
14. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387,** 296–299 (1997).
15. Leng, R. P. *et al.* Pirh2, a p53-induced ubiquitin-protein ligase, promotes p53 degradation. *Cell* **112,** 779–791 (2003).
16. Thut, C. J., Goodrich, J. A. & Tjian, R. Repression of p53-mediated transcription by MDM2: a dual mechanism. *Genes Dev.* **11,** 1974–1986 (1997).
17. Mirnezami, A. H. *et al.* Hdm2 recruits a hypoxia-sensitive corepressor to negatively regulate p53-dependent transcription. *Curr. Biol.* **13,** 1234–1239 (2003).
18. Allan, L. A. & Fried, M. p53-dependent apoptosis or growth arrest induced by different forms of radiation in U2OS cells: p21WAF1/CIP1 repression in UV induced apoptosis. *Oncogene* **18,** 5403–5412 (1999).
19. Qian, H., Wang, T., Naumovski, L., Lopez, C. D. & Brachmann, R. K. Groups of p53 target genes involved in specific p53 downstream effects cluster into different classes of DNA binding sites. *Oncogene* **21,** 7901–7911 (2002).
20. Doonan, J. & Hunt, T. Cell cycle Why don't plants get cancer? *Nature* **380,** 481–482 (1996).
21. Wertz, I. E. *et al.* Human De-etiolated-1 regulates c-Jun by assembling a CUL4A ubiquitin ligase. *Science* **303,** 1371–1374 (2004).
22. Dornan, D., Shimizu, H., Perkins, N. D. & Hupp, T. R. DNA-dependent acetylation of p53 by the transcription coactivator p300. *J. Biol. Chem.* **278,** 13431–13441 (2003).
23. Dornan, D. & Hupp, T. R. Inhibition of p53-dependent transcription by BOX-I phospho-peptide mimetics that bind to p300. *EMBO Rep.* **2,** 139–144 (2001).
24. Shimizu, H. *et al.* The conformationally flexible S9–S10 linker region in the core domain of p53 contains a novel MDM2 binding site whose mutation increases ubiquitination of p53 *in vivo*. *J. Biol. Chem.* **277,** 28446–28458 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We would like to thank T. Hupp for his contribution of reagents and thought-provoking discussions, G. Lozano for $p53^{-/-}$/$MDM2^{-/-}$ MEFs, S. Benchimol for the Pirh2 antibody and cDNA, J. Chen for H1299 cells, B. Henzel for mass spectrometry support, C. Reed for the generation of COP1 monoclonal antibodies, M. Vasser and P. Ng for oligonucleotide synthesis and purification, A. Waugh for Bioinformatics support, Genentech Protein Engineering and core DNA sequencing facility for support services, and P. Polakis and members of the Dixit lab for advice and encouragement. I.W. is supported by a PSTP fellowship from the University of California, Davis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.M.D (dixit@gene.com).

Integrating high-throughput and computational data elucidates bacterial networks

Markus W. Covert*, Eric M. Knight, Jennifer L. Reed, Markus J. Herrgard & Bernhard O. Palsson

Bioengineering Department, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA

* Present address: Biology Division, California Institute of Technology, 1200 E. California Boulevard, Mail Code 147-75, Pasadena, California 91125, USA

The flood of high-throughput biological data has led to the expectation that computational (or *in silico*) models can be used to direct biological discovery, enabling biologists to reconcile heterogeneous data types, find inconsistencies and systematically generate hypotheses[1–3]. Such a process is fundamentally iterative, where each iteration involves making model predictions, obtaining experimental data, reconciling the predicted outcomes with experimental ones, and using discrepancies to update the *in silico* model. Here we have reconstructed, on the basis of information derived from literature and databases, the first integrated genome-scale computational model of a transcriptional regulatory and metabolic network. The model accounts for 1,010 genes in *Escherichia coli*, including 104 regulatory genes whose products together with other stimuli regulate the expression of 479 of the 906 genes in the recon-

structed metabolic network. This model is able not only to predict the outcomes of high-throughput growth phenotyping and gene expression experiments, but also to indicate knowledge gaps and identify previously unknown components and interactions in the regulatory and metabolic networks. We find that a systems biology approach that combines genome-scale experimentation and computation can systematically generate hypotheses on the basis of disparate data sources.

We first validated the model, or '*in silico* strain' of *E. coli* (*i*MC1010^{v1}; see ref. 4 for conventions for naming *in silico* strains), against a data set of 13,750 growth phenotypes[5] obtained from the ASAP database[6], and then used this genome-scale model to select

a

Exp/Met/Reg		Percentage
Reg and Met models predict correctly		
+/+/+	6,222	45.3%
–/–/–	2,094	15.2%
Reg model only predicts correctly		
–/+/–	657	4.8%
+/–/+	0	0%
+/n/+	1,350	9.8%
–/n/–	505	3.7%
Met model only predicts correctly		
+/n/–	242	1.8%
–/n/+	153	1.1%
–/–/+	0	0%
+/+/–	257	1.9%
Neither model predicts correctly		
–/+/+	702	5.1%
+/–/–	1,568	11.4%
Met	8,968	65.2%
Reg	10,828	78.7%
Total	13,750	100%

c

	Carbon sources	FA		Nitrogen sources	FA
●	Acetoacetic acid	0.42	●	Adenine	0.42
●	Formic acid	0.05	○	Alanine-Leucine	0.28
○	γ-Amino butyric acid	0.21		Guanosine	0.51
●	Glycine	0.51	◎	L-Lysine	0.44
○	L-Arginine	0.35	◎	L-Methionine	0.42
	L-Glutamic acid	0.30	◎	L-Phenylalanine	0.24
○	L-Ornithine	0.21	●	*N*-Acetyl-D-mannosamine	0.48
○	Putrescine	0.41	●	Putrescine	0.60
◎	Thymidine	0.00	◎	Xanthine	0.04

	Knockout strains	FA		Knockout strains	FA
	Δ*argB*	0.39	◎	Δ*ilvY*	0.45
	Δ*argC*	0.31	◎	Δ*metA*	0.39
	Δ*argD*	0.27		Δ*pgi*	0.30
	Δ*argE*	0.52	◎	Δ*pls*	0.28
	Δ*argG*	0.57	◎	Δ*purD*	0.30
◎	Δ*glgA*	0.25	◎	Δ*purH*	0.41
◎	Δ*glgC*	0.25		Δ*tpiA*	0.50
◎	Δ*ilvD*	0.52			

○ Missing regulatory interactions
◎ Missing metabolic enzymes
● Conflicts with subsequent observations

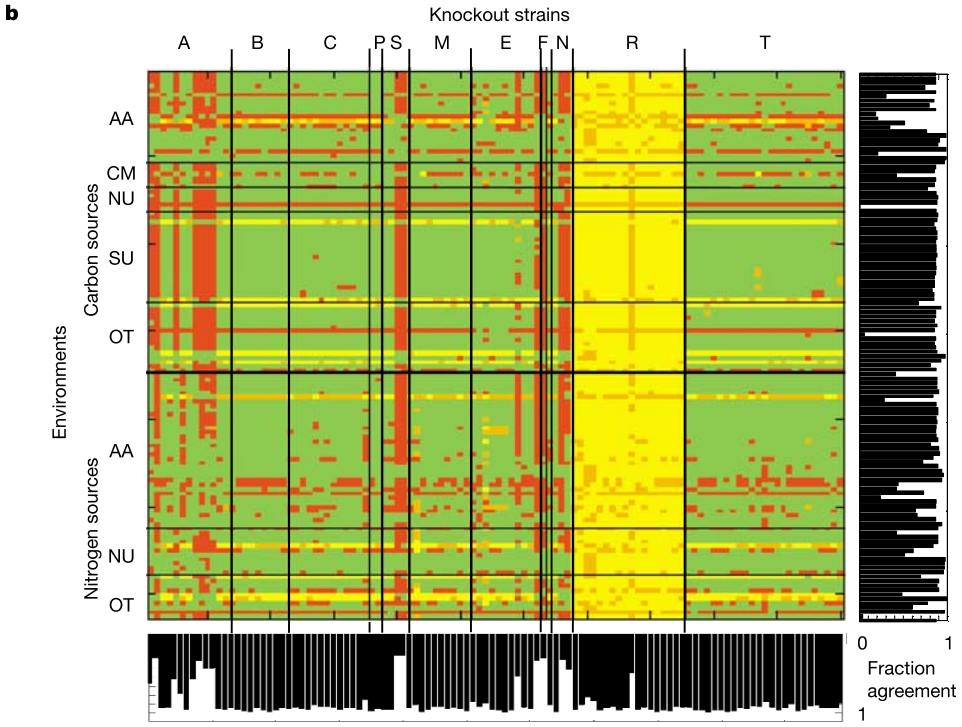

Figure 1 Growth phenotype study. **a**, Comparison of high-throughput phenotyping array data (Exp) with predictions for the *E. coli* network, both considering regulatory constraints (Reg) and ignoring such constraints as a control (Met). Each case is categorized by comparison type (Exp/Met/Reg), and results are listed as '+' (predicted or observed growth), '−' (no growth) or 'n' (for cases involving a regulatory gene knockout not predictable by the Met model). The comparisons are further divided into four subgroups represented by different colours. **b**, Chart showing individual results for each knockout under each environmental condition, with results categorized and coloured as in **a**. The environments involve variation of a carbon or nitrogen source and are further divided into subgroups: AA, amino acid or derivative; CM, central metabolic intermediate; NU, nucleotide or nucleoside; SU, sugar; OT, other. The knockout strains are also divided by functional group: A, amino acid biosynthesis and metabolism; B, biosynthesis of cofactors, prosthetic groups and carriers; C, carbon compound catabolism; P, cell processes (including adaptation and protection); S, cell structure; M, central intermediary metabolism, E; energy metabolism; F, fatty acid and phospholipid metabolism; N, nucleotide biosynthesis and metabolism; R, regulatory function; T, transport and binding proteins; U, unassigned. Each environment and knockout strain is associated with a fraction of agreement (FA) between regulatory model predictions and observed phenotypes, as shown in the bar charts to the right and below. **c**, Table showing all environments or knockout strains for which FA < 0.60. Of these substrates or knockout strains, 18 point to uncharacterized metabolic or regulatory capabilities in this organism, as indicated (see Supplementary Information for a case-by-case analysis).

transcription factors for prospective gene knockout studies. Comparison with the growth phenotypes showed that experimental and computational outcomes agreed in 10,828 (78.7%) of the cases examined, which is roughly the same success rate achieved in previous studies in *E. coli* and yeast that considered only a few hundred phenotypes[7–9]. In addition, 2,512 (18.3%) of the cases were predicted correctly only when regulatory effects were incorporated into the metabolic model (see Supplementary Information for details).

The comparisons in this study identified several substrates and knockout strains whose growth behaviour did not match predictions (Fig. 1). Further investigation of these conditions and strains led to the identification of five environmental conditions in which dominant, as yet uncharacterized, regulatory interactions actively contribute to the observed growth phenotype, and five environmental conditions and eight knockout strains that highlight uncharacterized enzymes or non-canonical pathways that are predicted to be used by the organism (Fig. 1; a detailed analysis of the discrepancies is provided in the Supplementary Information).

We wanted to determine the utility of this model-driven approach in elucidating transcriptional regulatory networks. A previous study, which evaluated the consistency between existing gene expression data sets and the known transcriptional regulatory network of *E. coli*, identified the response to oxygen deprivation as a partially consistent module[10,11]. We therefore targeted this part of the transcriptional regulatory network for further network characterization. Six strains with knockouts of key transcriptional regulators in the oxygen response (Δ*arcA*, Δ*appY*, Δ*fnr*, Δ*oxyR*, Δ*soxS* and the double knockout Δ*arcA*Δ*fnr*) were constructed. The messenger RNA expression profiles of these strains, as well as the

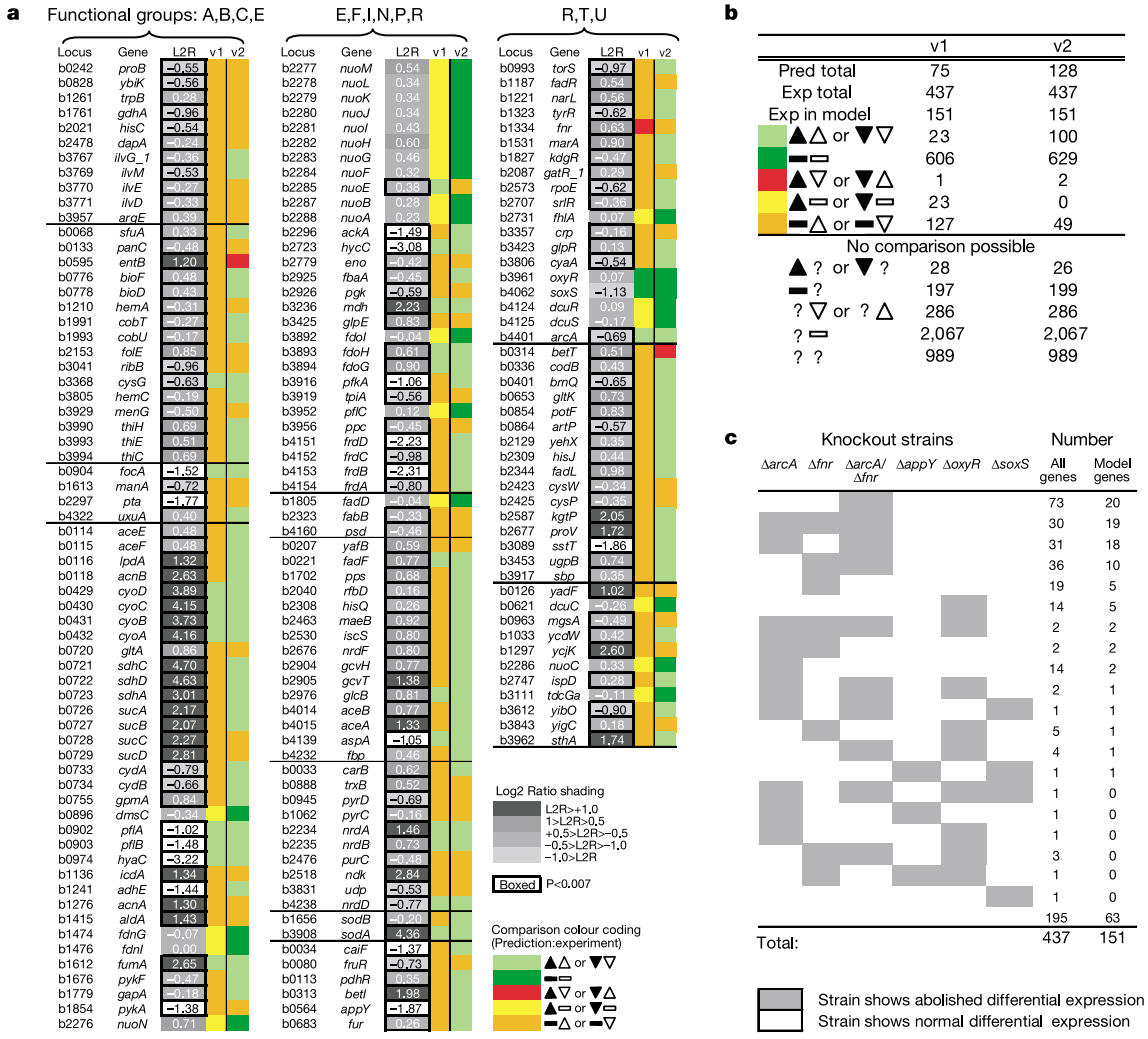

Figure 2 Characterization of the regulatory network related to the aerobic–anaerobic shift. **a**, The locus numbers, gene names and the log2 ratio (L2R) of gene expression (aerobic to anaerobic) are shown for all model genes with either predicted or observed changes in expression (genes were divided into the same functional groups as in Fig. 1). The L2R values are shaded depending on the magnitude of the expression shift, and those enclosed by a box indicate a statistically significant change in expression ($P < 0.007$, FDR < 5%). Comparisons between the experimental data and model predictions are also shown, where v1 (*i*MC1010^{v1}) and v2 (*i*MC1010^{v2}) designate the model used in the predictions. Filled and open symbols indicate model predictions and experimental data, respectively; rectangles indicate no change in gene expression; triangles indicate a change in expression, as well as the direction of change (upregulated or downregulated). **b**, Comparison of the predicted and observed expression changes for the v1 and v2 models. A question mark indicates either that the given gene was not included in the model or that no expression data were obtained for a given shift; other symbols are the same as in **a**. **c**, Systematic perturbation analysis was used to determine the transcription factors responsible for the expression change. The transcription factors knocked out in the six strains are shown on top. Each row indicates a pattern of knockout strains in which differential expression was abolished. The number of genes that show this pattern is indicated on the right. Thus, the first row indicates that for 73 of the 437 genes that showed differential expression in the wild-type strain (or 20 of the 151 genes accounted for by the model), the observed differential expression was abolished only in the Δ*arcA*Δ*fnr* knockout strain.

wild-type strain, were measured in aerobic and anaerobic glucose minimal medium conditions. The data were analysed[12] in the context of *i*MC1010^{v1} predictions to identify new interactions in the regulatory network (Fig. 2).

Expression profiling of the wild-type strain identified 437 genes that experienced a significant change in transcription in response to oxygen deprivation (*t*-test, multiple testing corrected to give a false discovery rate (FDR) of less than 5%); of these, 151 genes were included in *i*MC1010^{v1}. Computationally, 75 genes were predicted by *i*MC1010^{v1} to show differential expression in response to oxygen deprivation. These 75 genes could be divided into three categories: 23 agreed with measured expression changes; 24 had a predicted expression change that was either not found to be statistically significant in the experimental data (23/24) or in a direction opposite to that of the experimental data (1/24); and for 28 genes there were no expression data available (transcript abundance was determined to be 'absent' for two or more of the replicates). Thus, of the 47 (=23 + 24) differentially expressed genes that could be compared between the model computation and experiment, 23 (or 49% accuracy) agreed. Considering the overall number of genes in the model for which there were experimental data, the overlap (23) between the sets of predicted (47) and experimentally detected (151) differentially expressed genes is significant in comparison to a model that would randomly predict expression changes ($P < 0.005$ on the basis of a cumulative binomial distribution). There were 151 genes that were differentially expressed and included in the model; however, with only 23 (or 15% coverage) correctly computed, there is much room for expanding the transcriptional regulatory network in *i*MC1010^{v1} on the basis of the experimental data (Fig. 3).

To understand which transcription factors are involved in regulating these differentially expressed genes after oxygen deprivation, we compared the gene expression data for the wild-type and each knockout strain separately. Using two-way analysis of variance (ANOVA), we could determine whether the differential expression was significantly altered in the knockout strain as compared with the wild type. A large portion of the expression changes observed for the wild-type strain were not significantly affected in any of the knockout strains (195/437 or 44.6% of genes overall, 63/151 or 41.7% of genes in the model, FDR < 5%), suggesting that none of the five transcription factors studied here regulates the expression of these genes or that combinatorial interactions between multiple transcription factors are involved in regulation. For the remainder of the genes, differential expression was abolished in one or more of the knockout strains (Fig. 2c).

The ANOVA-based identification of transcription factors that influence differential expression of specific genes enabled us systematically to rewrite, relax or remove various regulatory rules in the model to resolve the discrepancies between *i*MC1010^{v1} and the experimentally determined wild-type differential gene expression. For many (81) of the genes, a regulatory rule already existed and had to be reconciled with our new data to accommodate the newly determined transcription factor dependencies. For genes where none of the knockouts abolished differential expression, we simply based a new regulatory rule on the presence of oxygen rather than a transcription factor (39 genes). By contrast, for genes where a change in expression was predicted but not observed, we removed oxygen dependency from the existing regulatory rule (23 genes). In addition, for 12 genes the predicted expression changes agreed with the observed expression in the wild type, but our knockout perturbation analysis indicated that the transcription factors involved in the regulation differed from previously reported data and the model needed to be changed (all new regulatory rules are detailed in the Supplementary Information).

The updated model (*i*MC1010^{v2}) was used to recalculate all of the predictions for both the aerobic and anaerobic expression data and the high-throughput phenotyping arrays. Note that *i*MC1010^{v2} accounts for the same genes as *i*MC1010^{v1} but has different regulatory interactions among the gene products and oxygen as an environmental variable. We found agreement between model predictions and the gene expression data to be substantially higher using the *i*MC1010^{v2} model, as expected (Fig. 2c). Specifically, 100 of the 151 expression changes were correctly computed with *i*MC1010^{v2}, and the number of false-positive predictions (Fig. 2, yellow boxes) was reduced to zero. In resolving many of the cases of unpredicted differential expression (Fig. 2, orange boxes), we found that implementation of the ANOVA-derived rule resulted in the inability of the wild-type or knockout *in silico* strain to grow aerobically or anaerobically on glucose, or under other conditions where growth had been previously established (for example, wild-type and knockout strain average growth rate under aerobic conditions, 0.68 ± 0.04 per hour; anaerobic, 0.43 ± 0.07 per hour). Such cases may be thought of as an 'overfit' of the microarray data. Accordingly, we relaxed the regulatory rule for these genes (42 in total) to allow for a correct phenotype prediction. Comparisons for the high-throughput phenotyping data revealed very little difference from Fig. 1 (only 11 out of 13,750 cases were affected; see Supplementary Information).

The iterative modification of the regulatory rules led to three main observations. First, some of the results of the knockout perturbation analysis are complex enough to make boolean rule formulation difficult. For example, the interplay of Fnr and ArcA can lead to complex behaviours where the expression change observed in wild type is abolished in the Δ*arcA* or the Δ*fnr* strains, but not in the Δ*arcA*Δ*fnr* strain. Such complex interplay among transcription factors can lead to specialized expression changes, as observed in the *cydAB* response to anaerobic, microaerobic and aerobic conditions[13,14].

Second, in revising regulatory rules for transcription factors we found that whereas in some cases, such as *arcA*, expression of a regulatory protein correlates positively with its activity, in several

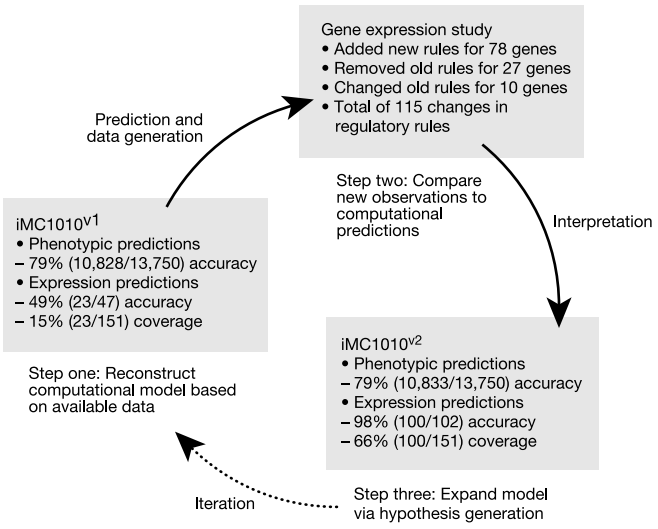

Figure 3 Biological network elucidation by a model-centric approach. Metabolic and regulatory networks may be expanded by using high-throughput phenotyping and gene expression data coupled with the predictions of a computational model. If model predictions are consistent with experimental observations, the network is adequately characterized. If not, the model identifies a knowledge gap and may be used to update, validate and generate hypotheses about organism function. Accuracy refers to the percentage of model predictions that agree with experimental data; coverage indicates the percentage of experimental changes predicted correctly by the model.

cases, including *fnr*, *betI* and *fur* among others, the mRNA level of a regulatory gene is reduced when the protein is in fact activated. For example, under anaerobic conditions when Fnr is known to be active[11], its expression is significantly reduced. Such behaviour, underscored by similar observations of mRNA transcript levels and corresponding protein product abundance in yeast[15], suggests that the identification of regulatory networks, and transcription factors involved in regulation in particular, will not be accomplished by the determination of co-regulated gene sets alone.

Third, many of these gene expression changes involve complex interactions and indirect effects. Transcription factors may be affected, for example, by the presence of fermentation by-products or the build up of internal metabolites. Such effects would be extremely difficult to identify or account for without a computational model.

In summary, we find that the reconciliation of high-throughput data sets with genome-scale computational model predictions enables systematic and effective identification of new components and interactions in microbial biological networks. Our study illustrates only the first round of an iterative model building strategy where an initial model based on literature-derived information (iMC1010^{v1}) is used to design informative experiments and then updated on the basis of the new experimental data obtained (iMC1010^{v2}). Another round of perturbation experiments will lead to iMC1010^{v3}, and so on. We expect that after an effort of some years and many iterations of this process, regulatory network elucidation for *E. coli* will be essentially complete. □

Methods

Computational model

We constructed the model of the *E. coli* metabolic and regulatory network by identifying network components, their functions and interactions from the primary literature[4,9,16]. Many approaches have been developed to analyse large-scale metabolic[17–22] and transcriptional regulatory[23–25] networks. Growth and gene expression simulations were done by regulated flux-balance analysis, which combines linear optimization to determine a growth-optimized metabolic flux distribution with logic statements to simulate the effects of regulatory processes over time. The whole model construction and simulation process has been described elsewhere in detail[26].

Strains and culture

The parent strain for knockout strains in this study was K-12 MG1655 (ref. 27), and all deletion strains were generated as described[28]. Growth experiments for the gene expression study were done on M9 glucose medium (2 g l^{-1}) under aerobic and anaerobic conditions, as described[17]. The growth data contained in the ASAP database were obtained by using high-throughput phenotype arrays (Biolog)[5]. In some cases (where the viability of a particular environment was unclear from the phenotype array data), we cross-validated the ASAP phenotyping data by culturing the wild-type strain under the given conditions in our laboratory (see Supplementary Information).

Gene expression profiling and analysis

All gene expression measurements were done at least in triplicate. Samples were stabilized by using RNAProtect bacterial reagent (Qiagen), and total RNA was isolated from exponentially growing cells using a RNeasy mini kit (Qiagen) in accordance with the manufacturer's protocols (see http://www1.qiagen.com). The RNA (10 μg) was then used as the template for complementary DNA synthesis, the product of which was fragmented, labelled and hybridized to an *E. coli* Antisense Genome Array (Affymetrix), which was washed and scanned to obtain an image in accordance with the manufacturer's protocols (see http://www.affymetrix.com). The image files were processed and expression values were normalized using dChip software[29]. We used quantitative real-time polymerase chain reaction with reverse transcription (RT–PCR) to validate expression changes for selected genes. The statistical significance of expression changes for each gene and each strain between aerobic and anaerobic conditions was determined by a *t*-test (log-transformed data, equal variance).

For each deletion strain, we used a two-way ANOVA (strain as the first factor and aerobic or anaerobic condition as the second factor) to determine whether the differential expression observed in the wild-type strain was significantly altered in the deletion strain by determining the statistical significance of the strain–condition interaction effect. For both the *t*-test and the ANOVA analysis, correction for multiple testing was done by using the Benjamini–Hochberg false discovery rate procedure[30], which determines the *P*-value cut-off for each test separately by estimating the FDR resulting from using a particular *P*-value cut-off. The false discovery rate refers to the fraction of true null tests out of all the tests called significant and an FDR of 5% was used for all tests. All gene expression data and the relevant information (such as the MIAME checklist) are provided in the Supplementary Information.

Received 22 November 2003; accepted 1 March 2004; doi:10.1038/nature02456.

1. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2,** 343–372 (2001).
2. Palsson, B. O. The challenges of *in silico* biology. *Nature Biotechnol.* **18,** 1147–1150 (2000).
3. Kitano, H. Systems biology: a brief overview. *Science* **295,** 1662–1664 (2002).
4. Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4,** R54.1–R54.12 (2003).
5. Bochner, B. R. New technologies to assess genotype–phenotype relationships. *Nature Rev. Genet.* **4,** 309–314 (2003).
6. Glasner, J. D. *et al.* ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* **31,** 147–151 (2003).
7. Forster, J., Famili, I., Palsson, B. O. & Nielsen, J. Large-scale evaluation of *in silico* gene knockouts in *Saccharomyces cerevisiae*. *Omics* **7,** 193–202 (2003).
8. Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97,** 5528–5533 (2000).
9. Covert, M. W. & Palsson, B. O. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277,** 28058–28064 (2002).
10. Herrgard, M. J., Covert, M. W. & Palsson, B. O. Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res.* **13,** 2423–2434 (2003).
11. Salmon, K. *et al.* Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J. Biol. Chem.* **278,** 29837–29855 (2003).
12. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292,** 929–934 (2001).
13. Compan, I. & Touati, D. Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Mol. Microbiol.* **11,** 955–964 (1994).
14. Cotter, P. A., Melville, S. B., Albrecht, J. A. & Gunsalus, R. P. Aerobic regulation of cytochrome *d* oxidase (*cydAB*) operon expression in *Escherichia coli*: roles of Fnr and ArcA in repression and activation. *Mol. Microbiol.* **25,** 605–615 (1997).
15. Griffin, T. J. *et al.* Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **1,** 323–333 (2002).
16. Reed, J. L. & Palsson, B. O. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J. Bacteriol.* **185,** 2692–2699 (2003).
17. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19,** 125–130 (2001).
18. Price, N. D., Papin, J. A., Schilling, C. H. & Palsson, B. O. Genome-scale microbial *in silico* models: the constraints-based approach. *Trends. Biotechnol.* **21,** 162–169 (2003).
19. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99,** 15112–15117 (2002).
20. Burgard, A. P. & Maranas, C. D. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* **74,** 364–375 (2001).
21. Burgard, A. P., Vaidyaraman, S. & Maranas, C. D. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.* **17,** 791–797 (2001).
22. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407,** 651–654 (2000).
23. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31,** 64–68 (2002).
24. Gutierrez-Rios, R. M. *et al.* Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles. *Genome Res.* **13,** 2435–2443 (2003).
25. Bar-Joseph, Z. *et al.* Computational discovery of gene modules and regulatory networks. *Nature Biotechnol.* **21,** 1337–1342 (2003).
26. Covert, M. W., Schilling, C. H. & Palsson, B. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213,** 73–88 (2001).
27. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1474 (1997).
28. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
29. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98,** 31–36 (2001).
30. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank K. Stadsklev and A. Fleming for assistance with computation; Z. Zhang and A. Raghunathan for experimental assistance; the Perna and Blattner laboratories for access to the high-throughput phenotyping data in the ASAP database; and the NIH for funding and support. M.W.C. and B.O.P. designed the project and were involved in all phases of the study; E.M.K. carried out experiments; J.L.R. reconstructed the model, ran simulations and did the phenotyping analysis; M.J.H. did the statistical analysis of the gene expression data.

Competing interests statement The authors declare competing financial interests: details accompany the paper on **www.nature.com/nature**

Correspondence and requests for materials should be addressed to B.O.P. (palsson@ucsd.edu). The gene expression data are available online in GEO (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE1121.