ELSEVIER

Reconstruction of microbial transcriptional regulatory networks

Markus J Herrgård, Markus W Covert and Bernhard Ø Palsson*

Although metabolic networks can be readily reconstructed through comparative genomics, the reconstruction of regulatory networks has been hindered by the relatively low level of evolutionary conservation of their molecular components. Recent developments in experimental techniques have allowed the generation of vast amounts of data related to regulatory networks. This data together with literature-derived knowledge has opened the way for genome-scale reconstruction of transcriptional regulatory networks. Large-scale regulatory network reconstructions can be converted to *in silico* models that allow systematic analysis of network behavior in response to changes in environmental conditions. These models can further be combined with genome-scale metabolic models to build integrated models of cellular function including both metabolism and its regulation.

Addresses
Department of Bioengineering, Bioinformatics Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA
*e-mail: palsson@ucsd.edu

Current Opinion in Biotechnology 2004, **15**:70–77

This review comes from a themed issue on
In silico biotechnology
Edited by Bernhard Palsson

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2003.11.002

Abbreviations
GWLA genome-wide location analysis
TF transcription factor

Introduction

There are three types of intracellular biochemical reaction networks where significant reconstruction efforts are underway: metabolic, transcriptional regulatory and signaling networks. Ultimately all three have to be integrated to generate whole-cell models of microbes and other organisms. Large-scale metabolic networks can be reconstructed relatively easily for any organism that has a published genome sequence and for which a sufficient amount of biochemical and physiological information is available [1]. These metabolic network reconstructions can be converted into mathematical models that can be used to simulate and analyze the behavior of the organism [1]. The models constructed in this fashion have found applications both in studying fundamental aspects of biology, such as evolutionary adaptation [2], and in designing microbial strains for the industrial production of biochemicals [3]. In addition to defining the metabolic interconversions, however, studying the integrated function of a microbial organism also requires a systematic description of the processes that regulate metabolism. In microbial organisms, key cellular processes such as metabolism are regulated at multiple levels, including transcriptional control of mRNA abundance and by a variety of post-transcriptional regulatory mechanisms such as kinetic regulation of enzymatic function. Transcriptional regulation constitutes perhaps the most experimentally tractable of these regulatory mechanisms, as mRNA abundance and DNA binding are easier to measure than, for example, protein abundance and activity. For microbial organisms, the primary role of transcriptional regulation is the response to changes in environmental conditions, such as nutritional status and environmental stresses. Owing to the central role that transcriptional regulation plays in cellular function and the availability of powerful experimental techniques to elucidate regulatory networks, reconstruction of these networks has emerged as a key task in biology [4,5].

Since metabolic network reconstruction primarily based on genome sequence data has been so successful, we discuss here the challenges and opportunities associated with regulatory network reconstruction through a comparison of the two network types. Some of the key differences between regulatory and metabolic networks and their respective reconstruction processes are summarized in Table 1. In addition to discussing progress made in the reconstruction of regulatory networks, we will also review current efforts to derive predictive models based on regulatory network reconstructions and efforts to build integrated models of cellular function that incorporate transcriptional regulation as one component.

Fundamental building blocks and network complexity

Although the fundamental biochemical reaction chemistry is the same for metabolic and regulatory networks, the types of reactions that form the building blocks of these networks are very different. The basic functional element of a regulatory network is the promoter region of a gene or operon, which contains the *cis* regulatory binding sites for the relevant transcription factors (TFs) that regulate the expression of a particular gene. The locations and orientations of these binding sites, as well as the affinity of the TFs to particular variants of the site, determine the expression levels of a gene in response to changes in the active TF concentrations inside the cell. The transcriptional regulatory network is then defined by which

Table 1

Some key differences between regulatory and metabolic networks that affect the network reconstruction process.

Network feature	Metabolic networks	Regulatory networks
Structure	Hard stoichiometry	Qualitative statements
Evolutionary conservation	Enzyme sequences highly conserved across species	Limited conservation of *cis* regulatory sites between closely related species
Malleability	Fixed structure in terms of the substrates that a particular enzyme can process	Adjustable structure, because of the possibility that mutations in the *cis* regulatory sites change binding specificity
Level of biochemical characterization	Fairly complete understanding of most subsystems in microbial organisms	Most subnetworks have not been well characterized even in microbial model organisms
Modeling approaches	Quantitative constraint-based models can be constructed at the genome-scale	Quantitative models can be currently constructed only on a small scale; qualitative discrete network models can be used to study large networks
Role of noise	Relatively small, because of high concentrations of metabolites involved in most reactions	Possibly significant in determining both structural features of the network and the overall response of the network to a stimulus

TFs bind to which promoters and what the integrated effect of all these TFs is on the expression of all the genes [6]. It has been demonstrated that the known organization of promoter regions in bacteria allows the implementation of a wide class of regulatory logic functions within a single promoter [7], so that even a single 'node' in the regulatory network can be relatively complex. At the basic level the mechanisms of transcriptional regulation are the same for prokaryotes and eukaryotes, but eukaryotic organisms add an additional level of complexity to the regulatory network in the form of chromatin-modifying enzymes and other co-regulators that are typically recruited to promoters by specific TFs [8].

Estimating the scope of a metabolic network reconstruction task for a given organism can be done relatively easily, by estimating the number of genes with potential metabolic function present in the genome on the basis of sequence similarity. For regulatory networks the number of TFs cannot be simply used to estimate the complexity of the network, owing to the fact that TFs can have multiple target genes and can often act in synergistic combinations. However, the relative fraction of TF coding genes tends to be higher for organisms that encounter more varied environmental conditions during their lifetime [9], indicating that there are limits to the complexity that can be achieved with a fixed number of TFs. Information on well-studied organisms can be used to evaluate the level of complexity of transcriptional regulatory networks in terms of the number of components (e.g. TFs and target genes) and regulatory interactions (Table 2). *Escherichia coli* has been predicted to have 314 TFs [10] and on the basis of the primary literature 1468 regulatory interactions have been identified [11••]. In *Saccharomyces cerevisiae*, there are 141 verified TFs [12] and large-scale *in vivo* protein–DNA binding screens indicate that there are at least 4000 regulatory interactions [13••]. For both *E. coli* and yeast these numbers of regulatory interactions are most likely to be underestimates, but they give an indication of the order of magnitude of the regulatory network reconstruction task.

Evolutionary conservation

From the viewpoint of network reconstruction perhaps the most significant difference between metabolic and regulatory networks is the degree of evolutionary conservation of the molecular components that form these networks. There is a high degree of sequence similarity between metabolic enzymes in different organisms, which allows functions to be assigned to open reading frames on the basis of sequence comparison between genomes. For regulatory networks, the corresponding

Table 2

Examples of reconstructed regulatory network structures in *E. coli* and *S. cerevisiae*.

	E. coli core metabolic[a]	*E. coli* full metabolic[b]	*E. coli* database[c]	*S. cerevisiae* core metabolic[d]	*S. cerevisiae* database[e]	*S. cerevisiae* GWLA[f]
Regulatory genes	16	104	123	55	109	106
Target genes	43	451	762[g]	168	418	2343
Regulatory interactions	–	–	1468[g]	258	945	3985[h]
Regulated reactions	46	555	–	117	–	–

[a][28•]; [b]MW Covert, *et al*., unpublished; [c][11••]; [d]MJ Herrgard, BO Palsson, unpublished; [e][48•]; [f][13••]. [g]Counting each gene in an operon separately. [h]Regulatory interactions with $P < 0.001$. Core metabolic refers to the regulatory network controlling core metabolism in an organism. Full metabolic refers to the regulatory network controlling all metabolic processes in an organism.

concept would be the conservation of either TF coding sequences or DNA-binding sites on promoter regions. Sequence similarity between TF coding genes is in itself of limited use for reconstructing the actual regulatory network structures, beyond being able to predict which genes in a genome code for TFs. This limitation exists because TFs typically utilize a small set of conserved DNA-binding domains [14] and knowing the sequence or structure of the domain does not generally allow prediction of the corresponding genomic binding sites.

TF binding *cis* regulatory sites are usually short (5–25 nucleotides) and degenerate, and small changes in these sites can lead to major changes in TF binding affinity [15]. For these reasons detecting regulatory sites by sequence comparison (commonly referred to as phylogenetic footprinting) requires access to genome sequences of multiple closely related species [16••,17••]. Even with a sufficient amount of comparative sequence data available, the complex structures of promoter regions might prevent the reliable identification of binding sites [15]. With more genomes being sequenced, and the development of more powerful computational approaches for regulatory motif finding [5,18,19], comparative approaches for reconstructing regulatory networks will probably become feasible in the future. However, deriving regulatory network structures in a semi-automatic fashion for less well-characterized organisms through comparative approaches will be significantly more challenging than the corresponding task for metabolic networks.

Although the individual TFs and binding sites may not be evolutionary conserved, there is evidence that certain patterns of network structure (commonly described as 'motifs') such as feed-forward loops have been preferentially selected for in evolution [11••,20]. The existence of a small class of over-represented network structural motifs was first established in *E. coli* on the basis of a literature-derived regulatory network structure [11••]. Similar motifs were also found in the experimentally established yeast regulatory network structure [13••]. In addition to similarity at the network 'motif' level, the actual physiological responses to environmental changes executed by the regulatory networks can be quite similar between species. The observation that regulatory network motifs have arisen by convergent evolution rather than by duplication of an ancestral motif [21•] appears to support this notion of convergent responses executed by different molecular level network elements.

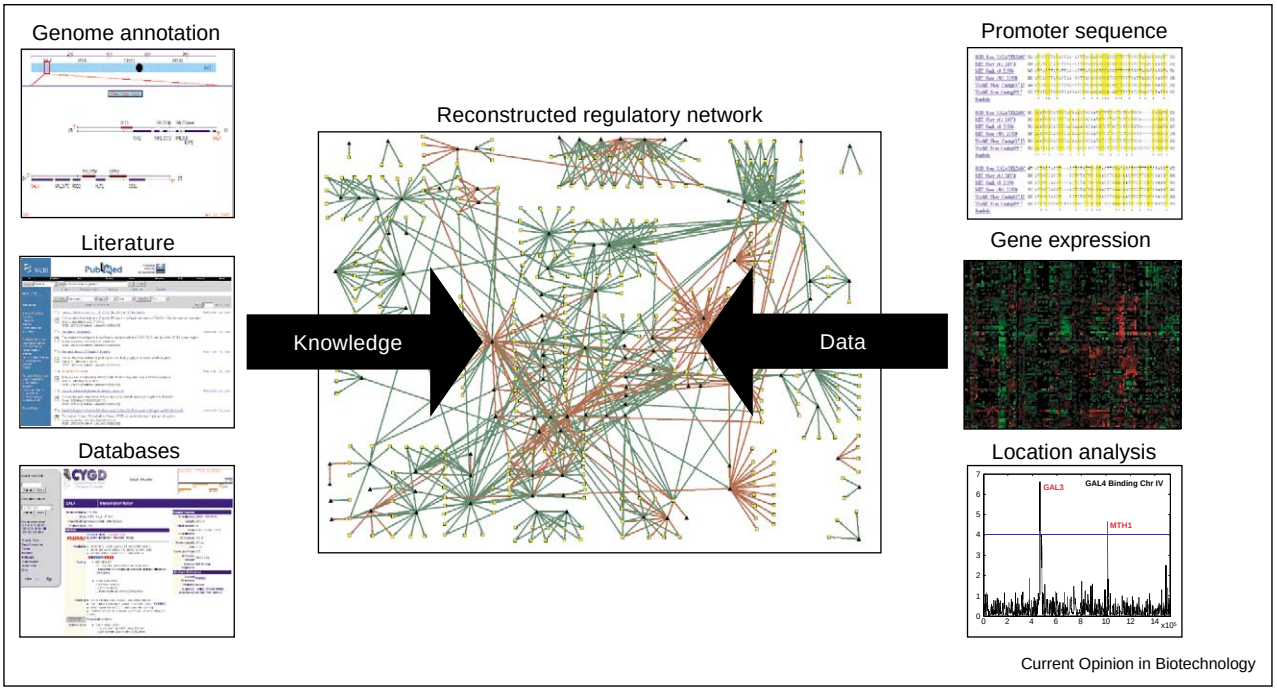
Databases and experimental data

Owing to the ease with which sequence similarity searching can be used to establish hypothetical enzymatic functions, many well-curated metabolic network databases exist [22–24]. These databases form the basis of any more detailed metabolic network reconstruction or modeling task. For regulatory networks similar comprehensive databases covering genome-scale regulatory networks in multiple organisms do not currently exist. For individual organisms, however, such network databases containing experimentally verified regulatory interactions have been established, the most prominent one being RegulonDB for *E. coli* [25]. There are also general databases for individual organisms, such as the Yeast Proteome Database (YPD) for yeast [12], that contain significant amounts of regulatory information. In addition to databases describing regulatory network structures, there are comprehensive databases that specialize in describing TF-binding sites, such as SCPD [26] for yeast and the general TF-binding site database TRANSFAC [27]. Although these databases contain valuable information for regulatory network reconstruction, they are not very complete and for the most part lack information about the synergistic effects between TFs acting on one gene. Nevertheless, these databases and primary research literature can be utilized to reconstruct regulatory networks for well-characterized organisms such as *E. coli* [11••,28•].

The major advantage that regulatory network reconstruction has over metabolic network reconstruction is the availability of high-throughput experimental data that is directly relevant to the network structure. For metabolic processes the only widely available data source is the genome sequence and its annotation — techniques for measuring relevant metabolic quantities such as metabolic fluxes and metabolite levels are still not commonly used and have not been fully scaled to the whole-genome level [29,30]. By contrast, the two primary data types useful for the regulatory network reconstruction task — genome-wide mRNA expression and location analysis data — are widely available.

Gene expression data can be readily generated for well-studied microbial organisms using several standard technologies [31]. Advances in statistical data analysis allow both significant changes in gene expression under different conditions to be established [32,33] and hypotheses about regulatory interactions or co-regulated gene modules to be derived directly from the data [34,35•,36•]. In particular, gene expression changes in response to the deletion of regulatory genes have been productively used to obtain sets of potential target genes for many regulatory proteins (e.g. in yeast and *E. coli* [37–39]). Genome-wide location analysis [40,41] (GWLA) that allows the direct detection of genomic target sites for DNA-binding proteins such as TFs promises to lead to an even more significant improvement in our ability to reconstruct regulatory network structures than gene expression profiling. So far, GWLA has been most extensively applied in yeast, where it has been used to map the target genes of 106 TFs under one set of conditions [13••]. In principle, the technique can be readily extended to other organisms [42]. GWLA has also been used to study the stimulus-dependent binding of TFs

Figure 1



Combining knowledge-based and data-based regulatory network reconstruction strategies. Regulatory networks can be reconstructed by collecting individual regulatory interactions from relevant databases and the primary literature (knowledge). Alternatively, networks can be derived directly from high-throughput experimental data and promoter sequence analysis through various data-mining methods. The combination of these strategies is expected to be the most productive way to achieve large-scale network reconstruction.

[43••], opening up the possibility of using this technique to map combinatorial interactions between TFs on a genome-wide scale.

The combination of expression profiling with GWLA as well as promoter sequence motif analysis has allowed the generation of hypothetical regulatory network structures using a variety of data integration methods [13••,35•, 44,45]. Deriving full regulatory network structures solely based on experimental data appears to be challenging, however, owing to the large quantities of high-quality data that would be required for such a task. One alternative to this purely data-driven approach would be to utilize well-curated regulatory network structures derived from databases and primary literature as a starting point for expanding the network on the basis of high-throughput data (Figure 1). For such an approach to succeed, one first needs to evaluate how well current known regulatory network structures agree with high-throughput datasets. This type of analysis has been performed for yeast and *E. coli* [46•–48•]. These studies have allowed the definition of network subcomponents and network structural motif types that are well supported by gene expression data and thus are good targets for data-driven model expansion. In the future, such combinations of knowledge-driven and data-driven regulatory network reconstruction strategies may allow the acceleration of network reconstruction in well-studied organisms.

Both knowledge-driven and data-driven network reconstruction strategies have so far been primarily applied to the two best-characterized microbial organisms, *E. coli* and *S. cerevisiae*. Existing network reconstructions for these organisms are summarized in Table 2, which lists the numbers of regulatory and target genes as well as the numbers of regulatory interactions for each reconstruction.

From network reconstructions to mathematical models

Regulatory network reconstruction can be achieved at different levels of detail depending on the intended application of the resulting network model. While different *in silico* modeling approaches have been extensively reviewed elsewhere [49], we will discuss these approaches here from the viewpoint of network reconstruction. The representation of network structure is another major difference between regulatory and metabolic network reconstruction: the latter are naturally described through the reaction stoichiometry, whereas for the former there is no single widely accepted description. Clearly, the more detailed descriptions of regulatory networks require increasingly large amounts of

parameters and larger quantities of data that cannot be easily obtained experimentally [50].

The first class of modeling approaches are primarily intended to describe the structural features of regulatory networks and do not accurately predict gene expression levels in response to changes in regulator activity. Directed graphs with TFs and target genes as nodes and regulatory interactions as edges are commonly used to visualize regulatory networks and to analyze their structural properties [11••,51]. Most methods for reconstructing regulatory networks based on gene expression and/or GWLA data describe the regulatory network as a directed graph [13••]. These graphs cannot represent important synergistic interactions between TFs and they do not allow simulation of model behavior or effective integration of regulatory networks with models of other cellular processes. However, the graph-based models of regulatory networks can also be used as a basis for building more quantitative models through measuring the regulatory strengths for different regulatory interactions experimentally [52].

The second class of modeling approaches primarily focuses on the prediction of gene expression levels at the expense of the scale of regulatory network subcomponents that can be modeled. Linear differential equations or linear models relating TF and target expression levels are the simplest of these approaches. This type of linear model was utilized in a recent study of the SOS response system in *E. coli*, in which experimental gene expression data was used to directly reconstruct a model for a small regulatory network without any prior knowledge of the network structure [53••]. Modeling approaches that go beyond linear models, such as nonlinear kinetic models and stochastic models, are necessary to understand the full dynamic and stochastic behavior of regulatory networks [54,55]. The most promising approach to understanding the nonlinear and stochastic behavior of regulatory networks has been studying both *in vivo* and *in silico* small engineered regulatory networks representing prototypical network components such as switches, autoregulatory loops, cascades and oscillators [56]. These studies have greatly improved our understanding of the dynamics of regulatory networks as well as the role of noise in these networks. The large number of parameters required to reconstruct nonlinear and stochastic models, however, have limited their use to small networks, and it is currently not clear whether individual models of such subnetworks can be systematically combined to form large-scale network models.

Towards integrated models of cellular function

As the role of transcriptional regulation is to modulate other cellular processes, integrating the reconstructed regulatory networks with models of these other processes is central to understanding regulatory network function in the context of the whole organism. Currently, there are major challenges for achieving this integration, relating both to obtaining the relevant data and to the modeling frameworks that are able to support the required large-scale integration. Possibilities for a suitable modeling framework include discrete network models, such as Boolean and Bayesian networks that allow key combinatorial interactions between regulators acting on the same gene to be represented either deterministically or probabilistically [57,58]. These qualitative network models allow the network behavior to be simulated, unlike graph-based models, but require significantly fewer parameters than linear or non-linear kinetic models. Boolean network models can also be readily integrated with, for example, whole-genome scale metabolic models to formulate integrated models of cellular function [59].

So far, this kind of integrated model has been formulated for the core metabolism in *E. coli* based on information in the databases and research literature [28•]. This model has also been recently expanded to a genome-scale model (MW Covert, *et al.*, unpublished), representing the first large-scale integrated model of multiple cellular functions in a microbial organism (Table 2). The major advantage with such integrated models is that even when the modeling of the regulatory network function is done at the qualitative level, the integrated regulatory/metabolic model can be used to quantitatively predict phenotypes such as growth rates. These predictions can then be experimentally verified by determining the phenotype of knockout strains of regulatory or metabolic genes. Furthermore, comparisons between model predictions and experimental data can be used to improve the model systematically. These types of integrated model are a powerful way to bring together multiple types of high-throughput data (e.g. gene expression and phenotyping) and to interpret these datasets, as discrepancies between model predictions and experimental data can point to specific inconsistencies in the current reconstructed regulatory network model.

Conclusions

The reconstruction of large-scale biochemical networks is necessary for understanding integrated network properties and for building quantitative predictive models of these networks. With the increased availability of high-throughput experimental data, the reconstruction of transcriptional regulatory networks is becoming feasible at the genome-scale for any organism that has been sufficiently well characterized and for which suitable data are available. Specific properties of regulatory networks, such as the lack of direct evolutionary conservation of *cis* regulatory sites, make both reconstruction and modeling of these networks more challenging than the corresponding tasks for metabolic networks. New developments in both experimental techniques and

computational reconstruction and modeling methods are still needed to improve our ability to reconstruct quantitative models of regulatory network. Nevertheless, the first large-scale regulatory network models that have been constructed show that we are already in a position to build these models and to utilize them to construct integrated models of microbial cells.

Update

A novel method for extracting co-regulated gene modules based on integrated analysis of GWLA and gene expression data has been developed recently [60•]. In comparison to utilizing GWLA data alone, the integrated approach is shown to provide improved sensitivity for identifying regulatory interactions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO: **Metabolic modeling of microbial strains *in silico***. *Trends Biochem Sci* 2001, **26**:179-186.

2. Ibarra RU, Edwards JS, Palsson BO: ***Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth**. *Nature* 2002, **420**:186-189.

3. Burgard AP, Maranas CD: **An optimization based framework for inferring and testing hypothesized metabolic objective functions**. *Biotechnol Bioeng* 2003, **84**:647-657.

4. Wyrick JJ, Young RA: **Deciphering gene expression regulatory networks**. *Curr Opin Genet Dev* 2002, **12**:130-136.

5. Stormo GD, Tan K: **Mining genome databases to identify and understand new gene regulatory systems**. *Curr Opin Microbiol* 2002, **5**:149-153.

6. Ptashne M, Gann A: *Genes and Signals*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2002.

7. Buchler NE, Gerland U, Hwa T: **On schemes of combinatorial transcription logic**. *Proc Natl Acad Sci USA* 2003, **100**:5136-5141.

8. Struhl K: **Fundamentally different logic of gene regulation in eukaryotes and prokaryotes**. *Cell* 1999, **98**:1-4.

9. Cases I, de Lorenzo V, Ouzounis CA: **Transcription regulation and environmental adaptation in bacteria**. *Trends Microbiol* 2003, **11**:248-253.

10. Perez-Rueda E, Collado-Vides J: **The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12**. *Nucleic Acids Res* 2000, **28**:1838-1847.

11. •• Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of *Escherichia coli***. *Nat Genet* 2002, **31**:64-68.
The authors reconstruct the transcriptional regulatory network in *E. coli* based on the RegulonDB database as well as the primary literature. They develop algorithms to systematically discover commonly occurring network structural motifs in the reconstructed network. Three such motifs were found and the role of these motifs in network function was further elucidated.

12. Csank C, Costanzo MC, Hirschman J, Hodges P, Kranz JE, Mangan M, O'Neill K, Robertson LS, Skrzypek MS, Brooks J *et al.*: **Three yeast proteome databases: YPD, PombePD, and CalPD (MycoPathPD)**. *Methods Enzymol* 2002, **350**:347-373.

13. •• Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I *et al.*: **Transcriptional regulatory networks in *Saccharomyces cerevisiae***. *Science* 2002, **298**:799-804.
This article describes a comprehensive application of the GWLA technique to mapping the regulatory targets of 106 TFs in yeast under one particular condition (rich medium). The network built from these data is used to discover commonly occurring network motifs, some of which are found to be the same as the ones present in the *E. coli* network [11••]. The authors also develop methods to automatically reconstruct regulatory networks based on location analysis and gene expression data and apply these methods to the cell-cycle regulatory network. Furthermore, the authors find that different cellular functions are highly interconnected at the level of transcriptional regulation.

14. Madan Babu M, Teichmann SA: **Evolution of transcription factors and the gene regulatory network in *Escherichia coli***. *Nucleic Acids Res* 2003, **31**:1234-1244.

15. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: **The evolution of transcriptional regulation in eukaryotes**. *Mol Biol Evol* 2003, **20**:1377-1419.

16. •• Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M: **Finding functional features in Saccharomyces genomes by phylogenetic footprinting**. *Science* 2003, **301**:71-76.
See annotation for [17••].

17. •• Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements**. *Nature* 2003, **423**:241-254.
These two papers describe the sequencing of multiple yeast species closely related to *S. cerevisiae* and the application of comparative sequence analysis to finding *cis* regulatory sites in genomes. Putative functions for these motifs were also suggested based on the utilization of GWLA and gene expression data. These studies demonstrate the utility of multiple closely related genome sequences for identifying conserved non-coding regions of the genome and indicate that comparative genomics approaches may be feasible for regulatory network reconstruction.

18. Qin ZS, McCue LA, Thompson W, Mayerhofer L, Lawrence CE, Liu JS: **Identification of co-regulated genes through Bayesian clustering of predicted regulatory binding sites**. *Nat Biotechnol* 2003, **21**:435-439.

19. Chiang DY, Moses AM, Kellis M, Lander ES, Eisen MB: **Phylogenetically and spatially conserved word pairs associated with gene-expression changes in yeasts**. *Genome Biol* 2003, **4**:R43.

20. Wolf DM, Arkin AP: **Motifs, modules and games in bacteria**. *Curr Opin Microbiol* 2003, **6**:125-134.

21. • Conant GC, Wagner A: **Convergent evolution of gene circuits**. *Nat Genet* 2003, **34**:264-266.
The authors evaluate the role of network motif duplication in the evolution of regulatory networks by studying reconstructed regulatory networks in *E. coli* and yeast. They find that separate instances of a particular network motif in the networks have not evolved by duplication of the genes participating in the motif, pointing towards convergent evolution as the origin of these motifs. The results indicate that the network motifs have an optimal design to perform their function.

22. Kanehisa M, Goto S, Kawashima S, Nakaya A: **The KEGG databases at GenomeNet**. *Nucleic Acids Res* 2002, **30**:42-46.

23. Schomburg I, Chang A, Schomburg D: **BRENDA, enzyme data and metabolic information**. *Nucleic Acids Res* 2002, **30**:47-49.

24. Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, Selkov E Jr, Liolios K, Joukov V, Kaznadzey D, Anderson I *et al.*: **The ERGO genome analysis and discovery system**. *Nucleic Acids Res* 2003, **31**:164-171.

25. Salgado H, Santos-Zavaleta A, Gama-Castro S, Millán-Zárate D, Díaz-Peredo E, Sánchez-Solano F, Pérez-Rueda E, Bonavides-Martínez C, Collado-Vides J: **RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12**. *Nucleic Acids Res* 2001, **29**:72-74.

26. Zhu J, Zhang MQ: **SCPD: a promoter database of the yeast *Saccharomyces cerevisiae***. *Bioinformatics* 1999, **15**:607-611.

27. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV *et al.*: **TRANSFAC: transcriptional regulation, from patterns to profiles**. *Nucleic Acids Res* 2003, **31**:374-378.

28. Covert MW, Palsson BO: **Transcriptional regulation in constraints-based metabolic models of *Escherichia coli***. *J Biol Chem* 2002, **277**:28058-28064.
This article describes the reconstruction of a transcriptionally regulated model of *E. coli* core metabolism and its application to predicting and interpreting gene deletion phenotypes as well as dynamic metabolic shifts. The integrated model was found to improve the predictive capability of constraint-based models of biochemical networks. The modeling approach presented in this paper presents a scalable framework to build genome-scale constraint-based models of integrated regulatory/metabolic network function.

29. Stitt M, Fernie AR: **From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon–nitrogen interactions**. *Curr Opin Biotechnol* 2003, **14**:136-144.

30. Sanford K, Soucaille P, Whited G, Chotani G: **Genomics to fluxomics and physiomics — pathway engineering**. *Curr Opin Microbiol* 2002, **5**:318-322.

31. Holloway AJ, van Laar RK, Tothill RW, Bowtell DD: **Options available – from start to finish – for obtaining data from DNA microarrays II**. *Nat Genet* 2002, **32(Suppl)**:481-489.

32. Quackenbush J: **Microarray data normalization and transformation**. *Nat Genet* 2002, **32(Suppl)**:496-501.

33. Churchill GA: **Fundamentals of experimental design for cDNA microarrays**. *Nat Genet* 2002, **32(Suppl)**:490-495.

34. Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N: **Revealing modular organization in the yeast transcriptional network**. *Nat Genet* 2002, **31**:370-377.

35. • Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: **Module networks: identifying regulatory modules and their condition- specific regulators from gene expression data**. *Nat Genet* 2003, **34**:166-176.
See annotation for [36•].

36. • Wang W, Cherry JM, Botstein D, Li H: **A systematic approach to reconstructing transcription networks in *Saccharomyces cerevisiae***. *Proc Natl Acad Sci USA* 2002, **99**:16893-16898.
These two articles introduce statistical methods for reconstructing regulatory network modules (defined as genes co-regulated by one or more TFs) based on gene expression and promoter sequence data. Both approaches also allow identification of the conditions under which particular TFs regulate the activity of particular genes. This represents an important advance over previous approaches for data-driven network reconstruction, which typically assumed that the same set of TFs would regulate each gene under all conditions.

37. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD *et al.*: **Functional discovery via a compendium of expression profiles**. *Cell* 2000, **102**:109-126.

38. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network**. *Science* 2001, **292**:929-934.

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

40. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E *et al.*: **Genome-wide location and function of DNA binding proteins**. *Science* 2000, **290**:2306-2309.

41. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO: **Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF**. *Nature* 2001, **409**:533-538.

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

43. •• Zeitlinger J, Simon I, Harbison CT, Hannett NM, Volkert TL, Fink GR, Young RA: **Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling**. *Cell* 2003, **113**:395-404.
The authors illustrate how GWLA experiments can be used to map the condition-dependent targets of TFs utilizing yeast Ste12p and Tec1p TFs as examples. They found that Ste12p binds a different set of target genes depending on the environmental signals it receives and that the target selection is influenced by concurrent binding of Tec1p. The approach presented in this paper could be used to determine combinatorial interactions between TFs by doing GWLA experiments for one TF in a strain where another TF with potential synergistic effect with the first TF is deleted.

44. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA: **Combining location and expression data for principled discovery of genetic regulatory network models.** *Pac Symp Biocomput* 2002:437-449.

45. Liu XS, Brutlag DL, Liu JS: **An algorithm for finding protein DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments**. *Nat Biotechnol* 2002, **20**:835-839.

46. • Herrgard MJ, Covert MW, Palsson BO: **Reconciling gene expression data with genome-scale regulatory network structures**. *Genome Res* 2003, **13**:2423-2434.
These papers describe the first comprehensive efforts to evaluate the agreement between known regulatory network structures reconstructed based on database and literature information and large-scale gene expression data sets. The results can be used to evaluate both the completeness of the reconstructed network structures and the utility of the publicly available expression data collections for expanding these networks.

47. • Gutierrez-Rios RM, Rosenblueth DA, Loza JA, Huerta AM, Glasner JD, Blattner FR, Collado-Vides J: **Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles**. *Genome Res* 2003, **13**:2435-2443.
See annotation for [46•].

48. • Yu H, Luscombe NM, Qian J, Gerstein M: **Genomic analysis of gene expression relationships in transcriptional regulatory networks**. *Trends Genet* 2003, **19**:422-427.
See annotation for [46•].

49. de Jong H: **Modeling and simulation of genetic regulatory systems: a literature review**. *J Comput Biol* 2002, **9**:67-103.

50. Bolouri H, Davidson EH: **Modeling transcriptional regulatory networks**. *Bioessays* 2002, **24**:1118-1129.

51. Guelzim N, Bottani S, Bourgine P, Kepes F: **Topological and causal structure of the yeast transcriptional regulatory network**. *Nat Genet* 2002, **31**:60-63.

52. Ronen M, Rosenberg R, Shraiman BI, Alon U: **Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics**. *Proc Natl Acad Sci USA* 2002, **99**:10555-10560.

53. •• Gardner TS, di Bernardo D, Lorenz D, Collins JJ: **Inferring genetic networks and identifying compound mode of action via expression profiling**. *Science* 2003, **301**:102-105.
The authors present a scalable method to identify linear models of regulatory networks from gene expression (quantitative PCR) data and an application to the SOS regulatory network in *E. coli*. There are some differences between the network structure identified from data and the known SOS network structure, most likely because of the relatively high noise level in the experiments. Despite differences in the network structure, the model was found to have good predictive power for predicting the primary target of a drug action based on gene expression data.

54. Tyson JJ, Chen KC, Novak B: **Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell**. *Curr Opin Cell Biol* 2003, **15**:221-231.

55. Rao CV, Wolf DM, Arkin AP: **Control, exploitation and tolerance of intracellular noise**. *Nature* 2002, **420**:231-237.

56. Hasty J, McMillen D, Collins JJ: **Engineered gene circuits**. *Nature* 2002, **420**:224-230.

57. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C *et al.*: **A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo**. *Dev Biol* 2002, **246**:162-190.

58. Pe'er D, Regev A, Elidan G, Friedman N: **Inferring subnetworks from perturbed expression profiles**. *Bioinformatics* 2001, **17**:S215-S224.

59. Covert MW, Schilling CH, Palsson B: **Regulation of gene expression in flux balance models of metabolism**. *J Theor Biol* 2001, **213**:73-88.

60. • Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK: **Computational discovery of gene modules and regulatory networks**. *Nat Biotechnol* 2003, **21**:1337-1342.
This article describes a new method that uses gene expression profiles to iteratively expand a core set of co-regulated genes identified by GWLA. The method is applied to both the previously published yeast GWLA dataset [13••] and to a new dataset for 14 TFs in rapamycin-treated cells. Application of the proposed method to these datasets reveals a number of unexpected connections between different regulatory subnetworks.