Inferring network interactions within a cell

Gregory W. Carter

Date received (in revised form): 19th May 2005

Greg Carter
is a postdoctoral fellow in computational biology at the Institute for Systems Biology. A theoretical physicist by training, he is now developing analytical methods and models for analysis of genetic interactions and gene expression.

Abstract

The continuing growth in high-throughput data acquisition has led to a proliferation of network models to represent and analyse biological systems. These networks involve distinct interaction types detected by a combination of methods, ranging from directly observed physical interactions based in biochemistry to interactions inferred from phenotype measurements, genomic expression and comparative genomics. The discovery of interactions increasingly requires a blend of experimental and computational methods. Considering yeast as a model system, recent analytical methods are reviewed here and specific aims are proposed to improve network interaction inference and facilitate predictive biological modelling.

Keywords: *biological interaction, interaction network, genetics, phenotype, computational modelling*

INTRODUCTION

Biological networks are becoming increasingly essential for the analysis and interpretation of large data sets generated by high-throughput technology.[1–3] In general, a network is defined by a set of distinct elements, the nodes, and interactions between the elements, the edges. When applied to biology, the nodes can represent genes, proteins or modular structures which are a combination of many co-dependent units. The edges correspond to biological interactions, connecting the nodes and forming a functional network that governs cellular processing.

Interaction networks are generally dependent on the environment inhabited by the organism under study. This implies many possible network configurations, each facilitating the biological processes most appropriate to external conditions. The data from which networks are generated should therefore be identified with the corresponding experimental environment(s). In practice, this has proved a difficult task owing to limitations in experimental methods and data management.

Methods of inferring biological interactions vary with the interaction type being probed. The interactions fall into three general classes of increasing abstraction: molecular interactions, regulatory interactions and genetic interactions. Although commonly assembled into separate networks, in reality all types are chemical interactions which combine to generate cellular behaviour. Regulatory interactions are, most generally, statements of influences on gene expression, which can correspond to one or more protein–DNA binding interactions or mRNA degradation. Genetic interactions are manifestations of many regulatory and functional interactions at work within a biochemical pathway or network, observed indirectly by comparing phenotype variation between genotypes.

This paper is nevertheless organised into distinct sections reviewing recent advances in the study of each of the three interaction classes, as this separation is not yet anachronistic. The eventual emergence of an integrated systems view of the cell probably first requires a better understanding of individual interaction types, network architectures and environmental specificity, to be followed by the development of predictive models that are limited in scope but highly accurate. Integration could then proceed at the level of these models. Recent works outlined in this review are examples of progress in this regard, with a

Gregory W. Carter,
Institute for Systems Biology,
1441 N. 34th Street,
Seattle, WA 98103, USA

Tel: +1 206 732 1396
Fax: +1 206 732 1299
E-mail: gcarter@systemsbiology.org

concentration on yeast as a model organism for the purposes of brevity and contextual coherence. The final section discusses goals for improving interaction inference and considers implications for modelling.

PROTEIN–PROTEIN AND PROTEIN–METABOLITE INTERACTIONS

The cell is a chemical machine that functions primarily through compartmentalised interactions between proteins, metabolites and nucleic acids. The understanding of these interactions is therefore of utmost importance. Obtaining global, high-confidence interaction data has proved challenging, even in model organisms.

Detection of protein–protein interactions

A prolific genome-wide approach in yeast has been the two-hybrid assay.[4] This method constructs an artificial transcription factor, fusing the DNA-binding domain with the first query protein and a transcriptional-activation domain to the second. When the two query proteins bind, a reporter gene is actively transcribed. Thousands of putative interactions have been discovered and assembled into networks.[5] Further analysis has demonstrated that the network is scale-free[6] and organised in a modular fashion,[7] with individual proteins forming semi-independent interaction clusters.

The primary drawback of the yeast two-hybrid method is an apparently poor reliability.[8] Interactions go undetected when query proteins involved cannot be localised to the nucleus or when fusions interfere with interaction or folding domains. Perhaps more troubling, false-positive interactions occur when query proteins bind in the nuclear environment but either rarely interact in natural locations or interact only under limited, unspecified environmental conditions. The assay is also unable to distinguish stable, functional interactions from weaker, transient interactions sustained only in the two-hybrid system.

An alternative technique under development for global discovery of protein–protein interactions is *fluorescence resonance energy transfer* (FRET) microscopy.[9] Fluorophores within about 60 Å of one another transfer energy, and tagging different fluorophores to a pair of proteins will generate observable resonant phenomena when the pair binds. The broad applicability of this method is not yet clear, but FRET has been successful in specific trials. High-throughput screens are currently in design. This system potentially has the great advantage of reporting both the localisation and timing of protein interactions, *in vivo* and in response to experimental conditions or perturbations.

Array technology is another candidate for direct detection of protein–protein interactions. Taking DNA microarrays as a model, the aim is to construct a chip onto which an entire proteome is spotted. This *protein array* would facilitate global screens for protein–protein, protein–nucleic acid and protein–metabolite interactions.[10,11] Technical challenges remain, including the synthesis of an entire proteome and difficulties in pinning proteins on a slide, but trials have been successfully performed with 5,000 yeast proteins.[12] While these methods are potentially promising, the *in vitro* nature of the assay will limit its reliability as a measure of the interactions which occur in a cell functioning in a particular environment.

Thus, although protein interactions are the primary means by which many cellular processes occur, they remain difficult to detect.

REGULATORY INTERACTIONS

Substantial progress is being made in inferring interactions that regulate genomic expression. Methods currently in use include direct detection of protein–DNA interactions, computational intragenomic and comparative genomic sequence analysis, and/or probabilistic algorithms that analyse microarray data. Recent combinations of these methods have led to rudimentary maps of genomic

regulation in yeast and signs of global network organisation.

Detection of protein–DNA interactions

The physical detection of protein–DNA interactions is led by chromatin immunoprecipitation combined with array technology, an approach commonly known as the ChIP-chip or genome-wide location analysis.[13] This process isolates and amplifies protein-bound DNA fragments, which are comparatively hybridised with a set of control DNA to a DNA microarray. After initial success in locating binding targets for the yeast transcription factors Gal4 and Ste12,[13] this method has recently been combined with multiple computational methods (discussed below).

Protein binding microarray (PBM) technology is an alternative chip-based method for detection of protein–DNA binding.[14] Simpler in conception than ChIP-chip, this method involves hybridising a putative DNA-binding protein expressed with an epitope tag to an array of double-stranded DNA fragments. The array is washed to remove non-specific protein binding and treated with an antibody specific to the tag. Early trials have identified dozens of previously unknown targets of yeast transcription factors Abf1, Rap1 and Mig1.[4] PBMs allow experiments to be carried out with great rapidity, but as an *in vitro* procedure it is susceptible to reporting interactions that do not occur *in vivo* and it cannot discern condition-specific binding.

Computational search for protein–DNA binding motifs

For organisms with sequenced genomes, computational methods have been developed to precisely identify regulatory binding sites within *cis*-acting promoter regions. The number of software applications that search for transcription factor binding sequences in intergenic regions has proliferated greatly in recent years. While some employ algorithms that search for co-occurrences of given sequence fragments, the majority use probabilistic models to discover functional binding sites amid the noise of the genome. The MEME software suite[15] uses expectation maximisation, and others, such as AlignACE,[16] MDscan[17] and MotifSampler,[18] are based on Gibbs sampling methods to find sequence fragments with a low likelihood of being present in a random genomic sequence.

The generic input for probabilistic algorithms is a set of putatively co-regulated genes (identified through expression analysis, observation of protein–DNA interactions, or some other means) and some portion of their upstream nucleotide sequence in a standard format (usually FastA coding). A set of short sequence fragments (or *motifs*) are returned, usually 6 to 15 bases in length, and scored for high representation and specificity as compared with the entire genome. Some algorithms generate position-specific scoring matrices, which list probabilities for the occurrence of each of the four bases at each of the positions in the motif.

While these algorithms do not directly detect protein–DNA interactions, comparing found motifs with known targets of DNA-binding proteins in the Transfac or similar databases provides strong evidence for an interaction. Like array-based methods, genomic searches suggest what *can* occur in a cell rather than what *does* occur under a given condition.

Evidence for condition-specificity of discovered protein–DNA interactions is generated when computational methods are combined with genome-wide location (ChIP-chip) data. The power of this union was demonstrated in a recent analysis of the yeast regulatory system.[19] Over two hundred suspected transcription factors, believed to be most of those present in yeast, were tested for DNA-binding through ChIP-chip analysis under varied growth conditions. Sequences encoded by bound probes were fed to a combination of six motif search algorithms and, after a set of significance and cross-species comparison tests, a final set of 65 regulator-and-motif pairs was determined. These and previously known bindings not found in the analysis have been combined into an online regulatory map for *Saccharomyces cerevisiae* of

impressive scale.[20] While this is certainly not the full regulatory network for yeast, it provides a database containing thousands of regulatory interactions with specificity to growth conditions. The authors state some generalities about promoter architectures and global binding behaviours, which is perhaps only the tip of a very rich iceberg of regulatory network organisation.

An alternative approach in the same spirit is based on the MDscan platform.[21] The expanded software, called Motif Regressor, draws co-expressed genes from an input data set and searches their upstream regions for motifs reported with MDscan. Linear regression is applied to the motif-match score and gene expression to remove non-specific candidates and the surviving motifs are grouped into functional sets. The result is a set of inferred regulator–target interactions specific to the growth conditions tested in the expression data set. A set of 15 distinct motifs was determined to be involved in the transcriptional response to amino acid starvation. This, like the above combination of software with chip technology, is an example of successfully integrating computational approaches with experimental data to expand the map of known protein–DNA regulatory interactions in yeast.

Bayesian analysis of microarray data

Beyond physical protein–DNA interactions lies a more abstract set of regulatory interactions generated by probabilistic analysis of gene expression data. These methods, which typically apply Bayesian analysis or machine learning techniques to large data sets,[22] generate networks of interacting gene clusters that form functional modules.[17] These modules are generally composed of one or more DNA-binding proteins and the binding targets. The intermodular interactions suggest a higher organisation of regulatory interactions, although they formally represent probabilistic dependence in expression patterns. Like the motif search algorithms, the use of these techniques is evolving into an integrated approach with physical interaction data[23] or motif search algorithms[24] that link transcription factors to co-regulated genes.

Logic of genetic interactions

Specifically, the Genetic Regulatory Modules (GRAM) algorithm developed by Bar-Joseph and collaborators[23] incorporates DNA binding data and over 500 expression data sets to produce a regulatory network with 106 gene modules regulated by 68 transcription factors. Similarly, the work of Segal and collaborators combines two computational techniques, expression clustering and motif searching, with an expectation maximisation algorithm to detect modules regulated by specific transcription factors.[17,24]

The interactions inferred using these techniques exhibit a higher-order organisation of protein–DNA interactions under which all genes that interact with a set of regulatory protein are gathered into a single biological unit. These gene sets usually show over-representation of functional annotations, strong evidence that the regulatory network is modular in architecture. As this concept becomes increasingly supported by the data, exceptional genes and regulators that do not conform to this pattern become of particular interest.

Recent studies of modular organisation in protein–protein networks demarcate a higher-order architecture of interactivity, both within and between functional modules,[25] which might also characterise regulatory hierarchies. Levels of regulation inherent in such a hierarchy would require methods to disentangle the layers of influence on the expression of every individual gene. Mathematical deconstruction procedures such as *singular value decomposition* (SVD) can isolate independent numeric contributions of genome-wide expression change.[26] Such an analysis of expression response to specific perturbations might identify different levels of the regulatory hierarchy in a pathway or biological subsystem of interest.

GENETIC INTERACTIONS

Genetic interactions are inferred through phenotype observations of allele combinations. More abstract than isolated chemical interactions, they offer a more comprehensive picture of how gene products interact, encompassing every information processing step from genotype to phenotype. Although limited by fewer high-throughput techniques and a lack of measurement standardisation, the field is nevertheless progressing rapidly. Genetic networks have identified functional organisation at the genetic level, revealed pathways of information flow within and between chemical response networks, and have proved to be a powerful tool in the computational modelling of yeast metabolism.

Phenotype observation has been the method by which genetic interactions are inferred since the beginnings of genetics. The classical concept follows from comparing phenotypes of strains which carry distinct alleles. Controlled experiments with inbred mutant strains of model organisms have become a basic method to better understand the biochemical organisation which underlies the phenotypic effects. A prime example is the effectively complete library of deletion mutant strains for *S. cerevisiae*, which has made the yeast a fruitful laboratory for the study of genetic interactions in eukaryotes. Interaction properties of experimental mutations are of particular interest in the broader context of outbred populations that carry multiple alleles. The genetic basis of human disease susceptibility, in which multigenic interaction effects are often pre-eminent, is naturally the most urgent application.[27,28]

The logic behind genetic interactions is as follows. Consider two genes, A and B, mutations of which yield observable phenotypes. Taking cell growth as an example phenotype, imagine a null mutation of A inhibits growth while a null mutation of B enhances growth. The growth rate of a double-mutant, with null mutations on both A and B, is also inhibited. In this case the A mutant is said to be *epistatic* to the B mutant, as its phenotype does not change in the B-null background. This is one possible genetic interaction which can be inferred from any phenotype that can be measured on a comparative scale.

Systematic interpretation of epistatic interaction was explicitly formulated by Avery and Wasserman[29] and later encoded into software.[30] Taking sex determination and apoptosis in *Caenorhabditis elegans* as models, these authors explicated common assumptions for biological information flow and delineated a set of rules to predict the relative positions of the two test genes in a sequence of events that determines the observed phenotype. In the above example, in which A is epistatic to B, the interpreted model places A downstream of B. Thus, gene A's absence overrides effects from the B-null allele.

Epistasis is one of many possible modes of genetic interaction. Another familiar interaction which implies interpretation is *synthesis*. In this case, single mutants of A and B have no effect on the phenotype, but the double-mutation does. This is consistent with gene products of A and B inhabiting parallel positions in pathways responsible for the phenotype or being members of a complex able to withstand the loss of one component but not two.

A readily observed phenotype for screens of synthetic interactions is cell fitness in yeast. Specifically, in many cases the double-null mutation of two genes with viable single deletions is lethal (or severely limits cell growth). These synthetic sick or lethal (SSL) genetic interactions are being mapped on a large scale by Boone and collaborators[28] using synthetic genetic array (SGA) high-throughput technology. Hundreds of yeast genes have been tested pairwise for SSL, and the resulting network is one of thousands of nodes and tens of thousands of interactions. Similar studies have been carried out on a smaller scale in fruitfly and worm genetics. It has been found that functionally related genes often exhibit

SSL interactions, and the immense scale of the network allows for statistical analysis of higher-order structure in the network.

Interestingly, trials that searched for SSL in triple-mutants found relatively few cases where the growth defect could not be accounted for by double-mutant effects. Although too few triple-mutants were tested for statistical certainty, interactions involving two proteins seem to dominate the biomolecular dynamics that control cellular growth.

Further analysis has demonstrated that integration of other genetic interactions (eg epistasis) or other types of biological interactions (eg protein–DNA binding) can further reveal functional architecture within this dense network.[31] Ultimately, discovery of repeated substructures, perhaps analogous to amplifiers and filters in electrical circuitry, can signify specific biological mechanisms or functions.

It has been noted[32] that the large number of interactions in the SSL network might ensure robustness, but this presents a substantial hurdle for an easy understanding multigenic effects in human disease. The large number of interactions per gene – Tong *et al.*[28] found an average of 34 – would produce insufficiently specific information in genome-wide association studies. Instead, a subset of candidate genes must be targeted.

Other commonly observed interactions are suppression, in which one mutant suppresses the phenotype of another in the double-mutant, and additivity, in which phenotypes of both single mutants combine to yield a third phenotype in the double-mutant. Other, more exotic interactions are also possible, depending on the results of phenotype measurements. The systematics of these relations between single-mutants, double-mutants and wild-type measurements form a finite set of possibilities that can be formally classified.

Modes of genetic interaction

To this end, consider a genotype X, and its associated phenotype observation P_X. The phenotype could be a numerical measurement or any other observation that can be clearly compared across mutant genotypes (eg slow *v.* standard *v.* fast growth). If genotypes are labelled by mutant alleles, a set of four phenotype observations can be assembled which defines a genetic interaction: P_A and P_B for the A and B mutant alleles, P_{AB} for the AB double mutant, and P_{WT} for the wild type. The relationship between these four measurements defines a genetic interaction; for example, $P_{AB} = P_A < P_{WT} < P_B$ described an epistatic interaction while $P_{AB} < P_{WT} = P_A = P_B$ represents synthesis. It is immediately clear that some of these interactions are symmetric under an exchange of genes A and B (eg synthetic) while others are not (eg epistatic). There are in fact 75 distinct inequalities that can be constructed from four phenotypes.

These inequalities were recently catalogued and classified into nine distinct interaction modes by Drees *et al.*[33] Many of these modes are familiar: epistatic, suppressive, conditional, additive and non-interaction. Three are not: asynthetic, single-nonmonotonic, and double-nonmonotonic. Asynthesis occurs when mutants A, B and AB exhibit the same deviant phenotype. In single-nonmonotonic interaction, a mutant gene shows opposite effects in the *WT* background and the other mutant background (eg $P_{WT} < P_A$ and $P_{AB} < P_B$). In double-nonmonotonic interaction, both mutant genes show opposite effects.

All of the nine interaction modes were detected in a yeast network constructed from photoassays of invasive growth on agar plates by a set of single and double mutant strains. Perturbations were chosen for relevance to the yeast filamentous response, a cell differentiation process which occurs under nitrogen starvation on solid media. Specific alleles were found to interact in a particular mode with neighbour genes of coherent biological function, leading to hypotheses on regulatory and pathway organisation. Large-scale patterns of mutual information were extracted from the data

set, which potentially allow for predictions of the interactions of one mutation from the effects of another it systematically follows. A picture arises of the biological network as an information conduit, with information flow determined by dense network architecture. When combined with more specific physical interaction and gene expression data to specify the direction of information flow, the systems-level network will enable precise predictions of system behaviour under perturbation.

Prediction of genetic interactions

To study genetic interactions computationally, recent work by Segrè *et al.* proposed an entirely model-based method for enabling the discovery of genetic interactions.[34] Using the theoretical assumption of flux balance analysis (FBA) devised by Famili, Förster and collaborators,[35] the metabolic network of yeast was simulated with multiple genetic backgrounds. Taking predicted fitness as a phenotype, double and single mutations were compared to obtain a spectrum of genetic interactions. The distribution is clearly tri-modal, with peaks corresponding to three interaction types: alleviating, aggravating and non-interacting. The first type roughly corresponds to synthetic lethality and the second to epistasis in the discussion above.

When functional annotations of each gene are considered, it is immediately noticeable that genes of similar function overwhelmingly interact in the same manner, a property the authors call 'monochromaticity' from the colour-mapping on their diagrams. Furthermore, interactions between two sets of genes with different functions tend to exhibit interactions of the same type, which suggests a modularity in the organisation of the genetic interactions. When laid out in terms of monochromatic clusters, the network transforms into one of coherent interactions between modules of functional genes.

While these results are specific to the computational model's general design, they intriguingly suggest a series of experimental screens to detect this monochromaticity. The procedure also alludes to a possible future of biological modelling, in which theoretical predictions from quantitative models drive experimental discovery.

Finally, we note that while the availability of mutant strains makes yeast a common choice for constructing genetic interaction networks, recent advances in RNAi technology[36] might soon provide a technique for the systematic study of genetic interactions in *C. elegans*[37] and *Drosophila melanogaster*.[38] The literature also contains a large amount of phenotypic data for flies which is currently being curated.[39] Continued progress in chemical genetics, the use of small organic molecules to inhibit molecular activity, may someday provide a high-throughput method for global analysis in mammalian cells.[40] Genetic interactions inferred from diverse organisms could be curated into a set of networks for cross-species comparison of commonly observed phenotypes.

IMPLICATIONS FOR BIOLOGICAL MODELLING

This review has concentrated on the inference of interactions that compose biological networks, broadly categorised as protein–protein, regulatory or genetic. Substantial advances have been made, many in the past year, and the potential of existing technology alone is enough to expect continued progress in the near future. Although this review concentrated on yeast as a model organism, similar advances have been made in the study of bacteria, archaea and other eukaryotes (for a few of the many examples see Davidson *et al.*,[41] Weston *et al.*,[42] Heyl and Schmulling,[43] Giot *et al.*,[44] Bock and Gough[45] and Li *et al.*[46]).

With proliferate discovery of biological interactions and networks, it is important to note that data assembly is merely an early step towards creating a systems approach to biology which can accurately predict responses to perturbations, both

on a cellular and organism-wide level. This goal will ultimately require some combination of two elements: (1) cohesive integration of disparate data types, which in the extreme case would constitute a comprehensive map of every microscopic cellular process and the conditions under which it occurs; and (2) inference of universal organisational principles that can be applied across systems and organisms. A manageable and flexible modelling strategy will almost certainly result from a hybrid of these two elements (if the second in fact exists in nature).

Advances in high-throughput data acquisition and graphical representation is facilitating the first requirement. General organisational principles, the second necessity, are being gradually uncovered through novel network analyses. These include the discovery of basic promoter architectures and condition-dependences of regulators,[19] repeated topological motifs,[47] modularity[48,49] and information flow[50] in interaction networks, and higher-order organisation of modular network structure.[25]

Requirements for advanced network inference

Progress in key areas will be of great importance for inferring the interactions that make this program feasible. These include the following:

- A method to probe protein–protein and similar molecular interactions with high coverage and reliability, preferably informing on spatial and temporal localisation as well as specifying dependence on environmental conditions. These constitute the vast majority of biological activity and modelling the dynamics of cellular response will rely on a clear view of interactome map and its organisation.

- Community-wide consensus on a set of standard phenotype assays and ontologies with associated measurement scales. This would greatly facilitate comparative study of genetic interaction networks in order to verify basic network features and help elucidate specialised functional mechanisms. On the whole, phenotypic assays remain narrowly targeted and, as a result, comparisons of genetic interaction networks for different phenotypes in a given species or similar phenotypes across diverse species remain extremely limited.

- High-throughput phenotype assays, preferably quantitative in nature, used in conjunction with libraries of systematic genetic perturbations in multiple organisms. The ensuing large sets of genetic interaction data would be central in understanding the structure of regulatory mechanisms.

- Progress in computational studies of existing data in order to determine fundamental properties upon which biological models can be built. Consider the construction of the flux-balance analysis model used in predicting genetic interactions, as discussed above. This required an exhaustive amalgamation of literature on yeast metabolism;[35] could a similar model someday be developed from a small set of rules for metabolic networks? Perhaps, as outlined in Harbison *et al.*,[19] a complete yeast regulatory map will suggest basic organisation of the transcriptional network; could this architecture be generalised and applied to predict transcriptional response and genomic expression? Additional knowledge of both the prevalence and character of network modularity would also benefit the development of basic modelling strategies.

Models developed under such constraints would furnish reliable and testable predictions for the behaviour of cells and organisms.

Acknowledgments
The author thanks Tim Galitski for contributions to the manuscript.

References

1. Aitchison, J. D. and Galitski, T. (2003), 'Inventories to insights', *J. Cell Biol.*, Vol. 161, pp. 465–469.
2. Bader, G. D., Heilbut A., Andrews, B. *et al.* (2003), 'Functional genomics and proteomics: Charting a multidimensional map of the yeast cell', *Trends Cell Biol.*, Vol. 13, pp. 344–356.
3. Galitski, T. (2004), 'Molecular networks in model systems', *Annu. Rev. Genomics Human Genet.*, Vol. 5, pp. 177–187.
4. Fields, S. and Song, O. (1989), 'A novel genetic system to detect protein–protein interactions', *Nature*, Vol. 340, pp. 245–246.
5. Schwikowski, B., Uetz, P. and Fields, S. (2000), 'A network of protein–protein interactions in yeast', *Nat. Biotechnol.*, Vol 18, pp. 1257–1261.
6. Jeong, H., Tombor, B., Albert, R. *et al.* (2000), 'The large-scale organization of metabolic networks', *Nature*, Vol. 407, pp. 651–654.
7. Rives, A. W. and Galitski, T. (2003), 'Modular organization of cellular networks', *Proc. Natl Acad. Sci. USA*, Vol. 100, pp. 1128–1133.
8. von Mering, C., Krause, R., Snel, B. *et al.* (2002), 'Comparative assessment of large-scale data sets of protein–protein interactions', *Nature*, Vol. 417, pp. 399–403.
9. Wouters, F. S., Verveer, P. J. and Bastiaens, P. I. H. (2001), 'Imaging biochemistry inside cells', *Trends Cell. Biol.*, Vol. 11, pp. 203–211.
10. Schweitzer, B., Predki, P. and Snyder, M. (2003), 'Microarrays to characterize protein interactions on a whole-proteome scale', *Proteomics*, Vol. 3, pp. 2190–2199.
11. Smith, M. G., Jona, G., Ptacek, J. *et al.* (2005), 'Global analysis of protein function using protein microarrays', *Mech. Ageing Dev.*, Vol. 126, pp. 171–175.
12. Zhu, H., Bilgin, M., Bangham, R. *et al.* (2001), 'Global analysis of protein activities using proteome chips', *Science*, Vol. 14, pp. 2101–2105.
13. Ren, B., Robert, F., Wyrick, J. J. *et al.* (2000), 'Genome-wide location and function of DNA binding proteins', *Science*, Vol. 290, pp. 2306–2309.
14. Mukherjee, S., Berger, M. F., Jona, G. *et al.* (2004), 'Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays', *Nat. Genet.*, Vol. 36, pp. 1331–1339.
15. Bailey, T. L. and Elkan, C. (1994), 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers', in 'Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology', AAAI Press, Menlo Park, CA, pp. 28–36.
16. Roth, F. R., Hughes, J. D., Estep, P. E. and Church, G. M. (1998), 'Finding DNA regulatory motifs within unaligned non-coding sequences clustered by whole-genome mRNA quantitation', *Nat. Biotechnol.*, Vol. 16, pp. 939–945.
17. Segal, E., Shapira, M., Regev, A. *et al.* (2003), 'Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data', *Nat. Genet.*, Vol. 34, p. 166.
18. Thijs, G., Lescot, M., Marchal, K. *et al.* (2001), 'A higher order background model improves the detection of regulatory elements by Gibbs sampling', *Bioinformatics*, Vol. 17, pp. 1113–1122.
19. Harbison, C. T., Gordon, D. B., Lee, T. I. *et al.* (2004), 'Transcriptional regulatory code of a eukaryotic genome', *Nature*, Vol. 431, pp. 99–104.
20. Fraenkel Lab website (URL: http://jura.wi.mit.edu/fraenkel/regulatory_map), maintained by Gordon, B. (2004).
21. Conlon, E. M., Liu, X. S., Lieb, J. D. and Liu, J. S. (2003), 'Integrating regulatory motif discovery and genome-wide expression analysis', *Proc. Natl Acad. Sci. USA*, Vol. 100, pp. 3339–3344.
22. Hartemink, A. J., Gifford, D. K., Jaakkola, T. S. and Young, R. A. (2001), 'Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks', in 'Proceedings of the 6th Pacific Symposium on Biocomputing', 3rd–7th January, Hawaii, pp. 422–433.
23. Bar-Joseph, Z., Gerber, G. K., Lee, T. I. *et al.* (2003), 'Computational discovery of gene modules and regulatory networks', *Nat. Biotechnol.*, Vol. 21, pp. 1337–1342.
24. Segal, E., Yelensky, R. and Koller, D. (2003), 'Genome-wide discovery of transcriptional modules from DNA sequence and gene expression', *Bioinformatics*, Vol. 19 (Suppl 1), pp. i273–182.
25. Han, J. D., Bertin, N., Hao, T. *et al.* (2004), 'Evidence for dynamically organized modularity in the yeast protein–protein interaction network', *Nature*, Vol. 430, pp. 88–93.
26. Alter, O., Brown, P. O. and Botstein, D. (2000), 'Singular value decomposition for genome-wide expression data processing and modeling', *Proc. Natl Acad Sci USA*, Vol. 97, pp. 10101–10106.
27. Hartman IV, J. L., Garvik, B. and Hartwell, L.

(2001), 'Principles for the buffering of genetic variation', *Science*, Vol. 291, pp. 1001–1004.

28. Tong, A. H. Y., Lesage, G., Bader, G. D. *et al.* (2004), 'Global mapping of the yeast genetic interaction network', *Science*, Vol. 303, pp. 808–813.
29. Avery, L. and Wasserman, S. (1992), 'Ordering gene function: The interpretation of epistasis in regulatory hierarchies', *Trends Genet.*, Vol. 8, pp. 312–316.
30. Zupan, B., Demsar, J., Bratko, I. *et al.* (2003), 'GenePath: A system for automated construction of genetic networks from mutant data', *Bioinformatics*, Vol. 19, pp. 383–389.
31. Wong, S. L., Zhang, L. V., Tong, A. H. *et al.* (2004), 'Combining biological networks to predict genetic interactions', *Proc. Natl Acad. Sci. USA*, Vol. 101, pp. 15682–15687.
32. Hartwell, L. (2004), 'Genetics. Robust interactions', *Science*, Vol. 303, pp. 807–808.
33. Drees, B. L., Thorsson, V., Carter, G. W. *et al.* (2004), 'Phenotype and the interaction of genetic perturbations', *Genome Biol.*, Vol. 6, p. R38.
34. Segrè, D., DeLuna, A., Church, G. M. and Kishony, R. (2005), 'Modular epistasis in yeast metabolism', *Nat. Genet.*, Vol. 37, pp. 77–83.
35. Förster, J., Famili, I., Fu, P. *et al.* (2003), 'Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network', *Genome Res.*, Vol. 13, pp. 244–253.
36. Fire, A., Xu, S., Montgomery, M. K. *et al.* (1998), 'Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*', *Nature*, Vol. 391, pp. 806–811.
37. Kamath, R. S., Fraser, A. G., Dong, Y. *et al.* (2003), 'Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi', *Nature*, Vol. 421, pp. 231–237.
38. Boutros, M., Kiger, A. A., Armknecht, S. *et al.* (2004), 'Genome-wide RNAi analysis of growth and viability in *Drosophila* cells', *Science*, Vol. 303, pp. 832–835.
39. Drysdale, R. (2001), 'Phenotypic data in FlyBase', *Brief. Bioinformatics*, Vol. 2, pp. 68–80.
40. Stockwell, B. R. (2004), 'Exploring biology with small organic molecules', *Nature*, Vol. 432, pp. 846–854.
41. Davidson, E. H., Rast, J. P., Oliveri, P. *et al.* (2002), 'A genomic regulatory network for development', *Science*, Vol. 295, pp. 1669–1678.
42. Weston, A. D., Baliga, N. S., Bonneau, R. and Hood, L. (2003), 'Systems approaches applied to the study of *Saccharomyces cerevisiae* and *Halobacterium* sp.', *Cold Spring Harbor Symp. Quant. Biol.*, Vol. 68, pp. 345–357.
43. Heyl, A. and Schmulling, T. (2003), 'Cytokinin signal perception and transduction', *Curr. Opin. Plant Biol.*, Vol. 6, pp. 480–488.
44. Giot, L., Bader, J. S., Brouwer, C. *et al.* (2003), 'A protein interaction map of *Drosophila melanogaster*', *Science,* Vol. 302, pp. 1727–1736.
45. Bock, J. R. and Gough, D. A. (2003), 'Whole-proteome interaction mining', *Bioinformatics*, Vol. 19, pp. 125–134.
46. Li, S., Armstrong, C. M., Bertin, N. *et al.* (2004), 'A map of the interactome network of the metazoan *C. elegans*', *Science*, Vol. 303, pp. 540–543.
47. Milo, R., Shen-Orr, S., Itzkovitz, S. *et al.* (2002), 'Network motifs: Simple building blocks of complex networks', *Science*, Vol. 298, pp. 824–827.
48. Prinz, S., Avila-Campillo, I., Aldridge, C. *et al.* (2004), 'Control of yeast filamentous-form growth by modules in an integrated molecular network', *Genome Res.*, Vol. 14, pp. 380–390.
49. Ravasz, E., Somera, A. L., Mongru, D. A. *et al.* (2002), 'Hierarchical organization of modularity in metabolic networks', *Science*, Vol. 297, pp. 1551–1555.
50. Ideker, T., Ozier, O., Schwikowski, B. and Siegel, A. F. (2002), 'Discovering regulatory and signalling circuits in molecular interaction networks', *Bioinformatics*, Vol. 18 (Suppl 1), pp. S233–240.