Towards a truly integrative biology through the functional genomics of yeast

Daniela Delneri*§, Francesco L Brancia*†# and Stephen G Oliver*‡

A complete library of mutant *Saccharomyces cerevisiae* strains, each deleted for a single representative of yeast's 6000 protein-encoding genes, has been constructed. This represents a major biological resource for the study of eukaryotic functional genomics. However, yeast is also being used as a test-bed for the development of functional genomic technologies at all levels of analysis, including the transcriptome, proteome and metabolome.

Addresses
*School of Biological Science, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK
†Michael Barber Centre for Mass Spectrometry, UMIST, Sackville Street, Manchester M60 1QD, UK
‡Corresponding author: SG Oliver; e-mail: steve.oliver@man.ac.uk
§e-mail: d.delneri@man.ac.uk
#e-mail: francesco.brancia@stud.umist.ac.uk

Current Opinion in Biotechnology 2001, **12**:87–91

0958-1669/01/$ – see front matter
© 2001 Elsevier Science Ltd. All rights reserved.

Abbreviations
2DGE two-dimensional gel electrophoresis
ICATs isotope-coded affinity tags
MS mass spectrometry
ORF open reading frame

Introduction

The availability of complete genome sequences presents the opportunity of adopting a systems-based approach to biology that will allow the determination of how all the genes in a genome act and interact to produce a functioning organism. Such an approach demands that technologies are developed that allow analyses at the level of mRNAs (transcriptome), proteins (proteome) and low molecular weight intermediates (metabolome) to be carried out in as comprehensive a manner as possible. Moreover, high-throughput methods of generating defined mutants and assessing their phenotypes also are required. The full range of genomic technologies is currently being applied to a small number of model organisms. Of these, the yeast *Saccharomyces cerevisiae* is in the vanguard.

The three levels of functional genomic analysis are both qualitatively and quantitatively distinct. Messenger RNA molecules, the subject of transcriptome analysis, may be studied in a fully comprehensive manner using the massively parallel technique of hybridisation-array analysis. However, mRNA molecules are not functional entities within the cell, but simply transmitters of the instructions for synthesising proteins, and so transcriptome analyses only approach functionality in an indirect manner. While both proteins and metabolites represent true functional entities within cells, the analysis of the proteome is fraught with technical difficulties, and most techniques in current use are neither comprehensive in their scope, nor massively parallel in their execution. The main difficulty with metabolome analysis is conceptual, rather than technical. While it is technically feasible to simultaneously analyse several hundred metabolites at once, the relationship between the metabolome and the genome differs from that of the transcriptome or proteome in that it is indirect. Many genes may be involved in the biosynthesis or degradation of a single metabolite. Thus bioinformatic tools are required that permit prior knowledge of the metabolic impact of known genes to be exploited in elucidating the function of novel genes.

It is bioinformatics that holds the key to functional genomics, since it is clear that investigations at all the levels of 'omic' analysis, and also phenotypic studies, will be required in taking this integrative approach to biology. Thus, massive amounts of data, of qualitatively and quantitatively distinct types, must be integrated, compared and used to construct heuristic models of living systems. In this short review, we will attempt to look at all of the levels of genomic analysis and discuss the prospects for their use in gaining an integrative view of the workings of *S. cerevisiae*.

Genome

A major conceptual and practical problem for the systematic analysis of gene function in eukaryotes is that of genetic redundancy. Gene duplications could have occurred through a series of local events or by complete genome duplication [1,2]. The high level of redundancy, generated by a whole-genome duplication, is thought to have been reduced via deletions and chromosomal rearrangements, while sequence divergence and selection allowed the acquisition of new functions. Thus, much of the redundancy in the yeast genome may be more apparent than real, with identical, or almost identical, gene products fulfilling distinct physiological roles due to differential gene expression or the targeting of similar proteins to different cellular compartments. For example, Delneri *et al.* [3] showed that only one fully functional member of the aryl-alcohol dehydrogenase (*AAD*) gene family in *S. cerevisiae* responds to oxidative stress. Moreover, theoretical studies have indicated that the genetic robustness of this organism does not rely on gene redundancy [4•].

Considerable progress has been made in the analysis of yeast gene function using single open reading frame (ORF) deletion mutants. However, it must be appreciated that the failure of any given mutant to reveal a phenotype may be the consequence of either genetic redundancy or of effective homeostatic controls within the cell. Thus, more studies are

required in which entire gene families are deleted [5,6] or where phenotype is assessed quantitatively, rather than qualitatively [7,8]. The year 2000 saw the establishment of a collection of mutant yeast strains, each bearing a defined deletion in one of yeast's 6138 potential protein-encoding genes (for details, contact euroscarf@em.uni-frankfurt.de). Each deleted ORF is flanked by two 20 bp molecular barcodes that are unique for each deletion [9••]. These allow the parallel analysis of the phenotypes of a large number of deletants to be performed using competitive growth assays. Such analyses assume that the specific ORF replacement is the only genetic change in the deletant. However, the transformation is both mutagenic and recombinogenic, and competitive hybridisation-array studies have shown that deletants may be aneuploid for whole chromosomes or chromosomal segments [10].

Ross-Macdonald *et al.* [11•] have taken a different approach to the generation of a large collection of defined yeast mutants. They used a transposon to mutagenise a yeast clone library in *Escherichia coli*. Individual plasmids were then prepared and used to transform a diploid yeast strain, where each plasmid integrates at its corresponding chromosomal locus, replacing the endogenous copy of the gene. The structure of the transposon allowed them to insert a short haemagglutinin tag within the yeast ORF that may be used for immunolocalisation of the tagged proteins. This permitted more than 300 previously nonannotated ORFs to be identified and the localisation of their protein products to be determined. This collection of mutants has also been used to determine disruption phenotypes for about 8000 strains, using 20 different growth conditions.

Genome-wide expression analysis (see below) has been used to follow adaptive evolution in yeast. The global transcript profiles from mutants selected during aerobic growth in a glucose-limited chemostat were compared with that of their parental strain. Genes involved in glycolysis and the tricarboxylic acid cycle showed alterations in expression in all three independently evolved strains, indicating that increased fitness is acquired by reducing the percentage of glucose which is fermented and increasing that channelled to respiration [12].

Transcriptome

Hybridisation arrays are now used widely to study the effects of cell physiology, development biology, or genetic constitution on the global expression pattern of yeast. For instance, yeast genes regulated directly or indirectly by the transcriptional activator Yap1p have been identified. The Yap1p-binding site was not always found in the promoter region of the target genes, and such genes are presumably under the indirect control of the activator. As ever, it is important to carefully define and regulate cell physiology when carrying out these global analyses. Thus, the *RPI1* gene (which encodes a repressor of the Ras–cAMP pathway) was found to be downregulated by Yap1p during the exponential growth phase, but upregulated in the stationary phase or following oxidative stress [13]. Careful control of cell physiology was employed by Ter Linde *et al.* [14] in their investigation of adaptation to aerobiosis and anaerobiosis in *S. cerevisiae*. About 93% of the ORFs analysed were expressed during both aerobic and anaerobic conditions, but about 140 and 219 genes showed a threefold higher transcription level under anaerobic and aerobic conditions, respectively.

Hughes *et al.* [15•] have recently constructed a 'compendium' of expression profiles corresponding to 300 diverse mutation or chemical treatments. In this way, the authors were able to assign functions to uncharacterised ORFs determining the biochemical and cellular pathways affected, via pattern matching. The compendium also was used to determine a novel target of the drug dyclonine, a sodium-channel blocker.

Proteomics

Although the yeast proteome (as it came to be known) has been studied, using two-dimensional gel electrophoresis (2DGE) since the late 1970s, progress in identifying the proteins contained within the spots on such gels has been disappointingly slow [16]. The availability of the complete yeast genome sequence and the development of 'soft' ionisation techniques for mass spectrometry (MS) have done much to speed up spot assignments, but improvements in MS and bionformatic techniques are still required. Recent advances include a method for *de novo* peptide sequencing [17] that improves fragmentation efficiency in post-source decay experiments, the use of guanidination to improve the signal response of carboxy-terminal lysine peptides [18], and the combination of the latter with Edman-type derivatisation [19]. This double-derivatisation approach offers a dual advantage, as both the total number of peptide masses available for database searching is increased, and the search space is reduced due the identification of the carboxy- and amino-terminal amino acids [20].

Yates Jr and co-workers [21] have developed a method for protein separation that is an attractive alternative to 2DGE. Known by the acronym MudPIT (multidimensional protein identification technique), it preserves protein–protein interactions and so facilitates the analysis of multisubunit complexes. MudPIT combines reversed-phase liquid chromatography with either cation-exchange or size-exclusion chromatography prior to analysis with electrospray tandem MS. This method was used to dissect the yeast 80S ribosomal subunit and resulted in the identification of an additional 11 proteins that had not previously been detected using the 2DGE approach.

An alternative approach to the identification of protein–protein interactions was applied in the analysis of the yeast nuclear pore complex (Nup85p), [22]. The complex was isolated using an affinity tag, crosslinked, and the resulting proteins pairs were resolved in sodium dodecyl sulphate-polyacrylamide gel electrophoresis and identified via matrix-assisted laser desorption/ionisation (MALDI)–MS, providing a model of the spatial organisation of the complex.

Gygi *et al.* [23••] combined sequence identification of the components in complex mixtures and accurate relative quantification. It exploits a novel class of chemical reagents, known as isotope-coded affinity tags (ICATs), together with tandem MS. The ICAT reagents contain both a biotin moiety and a thiol group, which is covalently attached to each cysteinyl residue in every protein. These two moieties are joined by a linker that contains either normal hydrogen or its heavy isotope, deuterium. Two protein extracts, obtained from yeast cultures grown on either ethanol or galactose, were compared by treating them with the isotopically light and heavy ICAT reagents, respectively.

A comprehensive analysis of protein–protein interaction in yeast has been undertaken recently by Uetz *et al.* [24•]. They employed two complementary strategies to exploit the yeast two-hybrid system for high-throughput analyses. The array approach (using pairwise crosses) yields more positive interactions, while the high-throughput library approach is more comprehensive, but less productive. Of the 12 'baits' that gave positive interactions with both screens, 48 possible partners were identified by the array approach, against only 14 in the library screen.

The relationship between mRNA and protein expression levels was investigated recently in a genome-wide context. Two groups of scientists made independent analyses comparing, under a given set of physiological conditions, the amount of proteins from a two-dimensional protein gel with the corresponding amount of transcripts calculated from the published serial analysis of gene expression (SAGE) analysis. Futcher *et al.* [25] found a satisfactory correlation between mRNA abundance, protein abundance and codon bias (measured in glucose and ethanol media). On the other hand, Gygi *et al.* [26] found that, for some genes, the mRNA abundance and the corresponding protein levels varied by more than 20-fold. According to this second study, the yeast proteome could not yet be predicted from the simple deduction of the transcript level because of the limits of the current approaches for quantitative analysis of protein levels.

Metabolomics

There are two major approaches to the assignment of gene function via metabolic analyses. One, which provides a direct link to the genome, is to uncover the biochemical reactions catalysed by enzymes encoded by genes of unknown function. Such an approach has been adopted by Martzen and co-workers: they developed a genomic strategy to identify yeast genes specifying biochemical activities by constructing a library of plasmids expressing glutathione *S*-transferase tagged yeast proteins. Using this strategy, they were able to identify proteins with novel biochemical activities [27•]. The problem of such an approach, as with the assignment of function via sequence homology, is that it attributes mechanism, rather than biological function, and is completely context-independent.

An alternative approach is to study the change in the cell's metabolic profile (or metabolome) which results from the deletion or overexpression of a given gene, and to assign function by comparing the metabolome change that result from the deletion of unknown genes with those that occur due to similar manipulations of known genes. This approach has been applied to *S. cerevisiae* and has been termed FANCY (for functional analysis by co-responses in yeast) [28•]. It involves a comprehensive analysis of cellular metabolites using either MS or nuclear magnetic resonance spectroscopy, combined with sophisticated chemometric analysis of the resulting spectra. The approach is able to reveal a phenotype for gene deletions that have no measurable effect on cell growth and cluster together metabolome profiles resulting from the deletion of genes affecting similar domains of metabolism. Yeast has a rather a limited range of metabolites, but a similar approach has been successfully employed with the plant, *Arabidopsis thaliana* [29•], which produces a much more complex range of metabolites.

Bioinformatics

Thousands of data points are accumulated in a single genome-wide expression experiment. In this context, the role of bioinformatics becomes essential in order to make biological sense out of the data and to assign functions to uncharacterised coding regions. An obvious approach to assigning gene function from transcriptome data is to group together genes with similar expression profiles [30]. Such methods analyse patterns of gene activity in an 'unsupervised' fashion. That is, without recourse to a training set of data relating to genes of well-known function and regulatory pattern [31]. Recently, Brown *et al.* [32] introduced a method for the identification of functionally related yeast proteins based on the theory of support vector machines, which represent a supervised learning technique that exploits prior knowledge of gene function to identify clusters. An alternate approach that exploits genetic programming has been published recently and appeared to be even more successful in that it proved able to learn the class of helix-turn-helix proteins, which include the transcription factors [33].

A different approach to the cluster analysis has been developed by Marcotte *et al.* [34••]. They grouped proteins by 'experimental data', 'related metabolic function', 'related phylogenetic profiles', 'rosetta stone method' (which links individual proteins whose homologues, in other organisms, are combined into a single multifunctional complex), and 'correlated mRNA expression'. Using these methods, they found a total of 93,000 pairwise links between functionally related yeast proteins, allowing the assignment of a general function to more than half the uncharacterised ORFs. Such an approach is an excellent example of the benefits of integrating data from all levels of functional genomic analysis. These integrative approaches require database structures that have sufficient breadth and flexibility to allow complex queries to be made over the qualitatively and quantitatively different datasets represented by the 'omes'. An appropriate object data model has been constructed recently that permits the integration of genome, transcriptome and proteome data for yeast [35].

Conclusions

The fact that the genome sequences of even well-characterised organisms contained about 40% of genes whose function had neither been established nor predicted was a shock to many biologists. The challenge was to assign a function to each of these novel genes. However, the comprehensive methods of analysis that are used to pursue these assignments have revealed that our view of biological function is rather one-dimensional. The hope now is that the analysis of genome, transcriptome, proteome and metabolome, as well as the phenotype, will allow a much more integrative view of biology at the level of the whole organism. In the early stages of such a process, single-celled organisms offer the advantages of simplicity combined with genetic and physiological malleability. The experimental methods and theoretical framework established using an organism such as yeast should provide a firm foundation for an integrative biology of human beings and their domestic plants and animals.

Acknowledgements

Functional genomics research in the authors' laboratory is supported by the Biotechnology and Biological Sciences Research Council, the Wellcome Trust, and the European Commission. F Brancia thanks the BBSRC and Nycomed-Amersham for a CASE Studentship.

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. Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome.** *Nature* 1997, **387**:708-713.

2. Seoighe C, Wolfe KH: **Updated map of duplicated regions in the yeast genome.** *Gene* 1999, **238**:253-261.

3. Delneri D, Gardner DCJ, Oliver SG: **Analysis of the seven-member *AAD* gene set demonstrates that genetic redundancy in yeast may be more apparent than real.** *Genetics* 1999, **153**:1591-1600.

4. • Wagner A: **Robustness against mutations in genetic networks of yeast.** *Nat Genet* 2000, **24**:355-361.

Using functional genomics data from *S. cerevisiae*, the author demonstrates that robustness against mutation is mainly due to interactions of unrelated genes rather than gene redundancy in itself.

5. Delneri D, Tomlin GC, Wixon JL, Hutter A, Sefton M, Louis EJ, Oliver SG: **Exploring redundancy in the yeast genome: an improved strategy for use of the cre-loxP system.** *Gene* 2000, **252**:127-135.

6. Llorente B, Fairhead C, Dujon B: **Genetic redundancy and gene fusion in the genome of the baker's yeast *Saccharomyces cerevisiae*: functional characterization of a three-member gene family involved in the thiamine biosynthetic pathway.** *Mol Microbiol* 1999, **32**:1140-1152.

7. Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, Davis RW: **Genomic profiling of drug sensitivities via induced haploinsufficiency.** *Nat Genet* 1999, **21**:278-283.

8. Baganz F, Hayes A, Farquhar R, Butler PR, Gardner DCJ, Oliver SG: **Quantitative analysis of yeast gene function using competition experiments in continuous culture.** *Yeast* 1998, **14**:1417-1427.

9. •• Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H *et al.*: **Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis.** *Science* 1999, **285**:901-906.

Using a PCR-mediated gene disruption strategy, the authors constructed ~6900 *S. cerevisiae* strains with a precise deletion of more than one-third of the open reading frames in the genome. In addition, two molecular barcodes were introduced into each deletion strain to allow a large number of deletion strains to be analysed in parallel in competitive growth assays.

10. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ *et al.*: **Widespread aneuploidy revealed by DNA microarray expression profiling.** *Nat Genet* 2000, **25**:333-337.

11. • Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung KH, Sheehan A, Symoniatis D, Umansky L *et al.*: **Large-scale analysis of the yeast genome by transposon tagging and gene disruption.** *Nature* 1999, **402**:413-418.

The authors used a transposon-based method for the genome-wide analysis of disruption phenotypes, gene expression and protein localisation. A yeast genomic DNA library was mutagenised in *E. coli* with a mini-transposon (mTn) containing the *URA3* marker and *lacZ* reporter gene flanked by *loxP* sites, and three copies of haemagglutinin. This collection of mutants has been used to determine disruption phenotypes for about 8,000 strains using 20 different growth conditions. Moreover, 300 previously nonannotated open reading frames were identified and analysed by indirect immunofluorescence.

12. Ferea TL, Botstein D, Brown PO, Rosenzweig RF: **Systematic changes in gene expression patterns following adaptive evolution in yeast.** *Proc Natl Acad Sci USA* 1999, **96**:9721-9726.

13. Dumond H, Danielou N, Pinto M, Bolotin-Fukuhara M: **A large-scale study of Yap1p-dependent genes in normal aerobic and H_2O_2-stress conditions: the role of Yap1p in cell proliferation control in yeast.** *Mol Microbiol* 2000, **36**:830-845.

14. ter Linde JJ, Liang H, Davis RW, Steensma HY, van Dijken JP, Pronk JT: **Genome-wide transcriptional analysis of aerobic and anaerobic chemostat cultures of *Saccharomyces cerevisiae*.** *J Bacteriol* 1999, **181**:7409-7413.

15 • 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.

Using a 'compendium database' of expression profiles from 300 different mutations and chemical treatments in *S. cerevisiae*, the authors were able to identify the cellular pathways affected via pattern matching.

16. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.** *Proc Natl Acad Sci USA* 2000, **97**:9390-9395.

17. Keough T, Youngquist RS, Lacey MP: **A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry.** *Proc Natl Acad Sci USA* 1999, **96**:7131-7136.

18. Brancia FL, Oliver SG, Gaskell SJ: **Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides.** *Rapid Commun Mass Spectr* 2000, **14**:2070-2073.

19. Summerfield SG, Bolgar MS, Gaskell SJ: **Promotion and stabilization of b(1) ions in peptide phenylthiocarbamoyl derivatives: Analogies with condensed-phase chemistry.** *J Mass Spectrom* 1997, **32**:225-231.

20. Brancia FL, Butt A, Benyon RJ, Hubbard SJ, Gaskell SJ, Oliver SG: **A combination of chemical derivatisation and improved bioinformatic tools optimises protein identification for proteomics.** *Electrophoresis* 2001, in press.

21. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* 1999, **17**:676-682.

22. Rappsilber J, Siniossoglou S, Hurt EC, Mann M: **A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry.** *Anal Chem* 2000, **72**:267-275.

23 •• Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.

This method is based on a class of new chemical reagents termed isotope-coded affinity tags and tandem mass spectrometry. The relative quantification exploits stable isotope dilution techniques and enabled comparison of protein expression in the yeast *S. cerevisiae*, using either ethanol or galactose as a carbon source.

24 • Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **Protein–protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.

A genomic scale analysis of protein–protein interactions in yeast was undertaken by the two-hybrid method, with more than 2500 interactions described.

25. Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI: **A sampling of the yeast proteome.** *Mol Cell Biol* 1999, **19**:7357-7368.

26. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.

27 Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S,
• Grayhack EJ, Phizicky EM: **A biochemical genomics approach for identifying genes by the activity of their products.** *Science* 1999, **286**:1153-1155.

The authors developed a genomic strategy to identify yeast genes specifying biochemical activities: a set of ~6144 strains carrying a plasmid expressing different open reading frames (ORFs) fused with a glutathione *S*-transferase (GST) tag were generated. The GST-ORFs were purified in 64 pools of 96 fusions each. Every pool was assayed for a particular activity, and to determine, within the positive pool, the source strain responsible for that activity. GST-ORFs from each of the eight rows and 12 columns of strains from the microtitre plate were tested. Using this strategy, they identified a cyclic phosphodiesterase, and Appr-1′-p-processing activity and a cytochrome c methyltransferase.

28. Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A,
• Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ *et al.*: **Functional genomics via the metabolome: a FANCY approach to characterising mutations with a silent phenotype.** *Nat Biotechnol* 2001, **19**:45-50.

The authors studied the change in the yeast metabolome as result of gene deletion. Clustering together metabolome profiles resulting from the deletion of genes affecting similar domains of metabolism, the authors were able to reveal a phenotype for gene deletions that has no measurable effect on yeast cell growth, and to assign the gene to a functional domain.

29 Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L:
• **Metabolite profiling for plant functional genomics.** *Nat Biotechnol* 2000, **18**:1157-1161.

Using gas chromatography/mass spectrometry, the authors were able to quantify more than 300 compounds from *Arabidopsis thaliana* and to assign a chemical structure to about half of them. They compared four genotypes, showing that each of them possess a distinct metabolic profile.

30. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.

31. Kell DB, King RD: **On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning.** *Trends Biotechnol* 2000, **18**:93-98.

32. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr., Haussler D: **Knowledge-based analysis of microarray gene expression data by using support vector machines.** *Proc Natl Acad Sci USA* 2000, **97**:262-267.

33. Gilbert RJ, Rowland JJ, Kell DB. **Genomic computing: explanatory modelling for functional genomics.** In *Proceedings of the Genetic and Evolutionary Computation Conference*. Edited by Whitely D, Goldberg D, Cantú-Paz E, Spector L, Permee I, Beyer HG. San Francisco: Morgan Kaufmann; 2000:551-557.

34. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D: **A**
•• **combined algorithm for genome-wide prediction of protein function.** *Nature* 1999, **402**:83-86.

The authors developed a new method to find protein–protein interactions. They linked the proteins by 'correlated evolution', 'correlated mRNA expression' and 'patterns of domain evolution', finding a total of 93,000 pairwise links between functionally related yeast proteins, and assigning a general function to more than half uncharacterised open reading frames.

35. Paton NW, Khan SA, Hayes A, Moussouni F, Brass A, Eilbeck K, Goble CA, Hubbard SJ, Oliver SG: **Conceptual modelling of genomic information.** *Bioinformatics* 2000, **16**:548-557.