Functional genomics: lessons from yeast

Stephen G. Oliver

School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK
(steve.oliver@man.ac.uk)

Functional genomics represents a systematic approach to elucidating the function of the novel genes revealed by complete genome sequences. Such an approach should adopt a hierarchical strategy since this will both limit the number of experiments to be performed and permit a closer and closer approximation to the function of any individual gene to be achieved. Moreover, hierarchical analyses have, in their early stages, tremendous integrative power and functional genomics aims at a comprehensive and integrative view of the workings of living cells. The first draft of the human genome sequence has just been produced, and the complete genome sequences of a number of eukaryotic human pathogens (including the parasitic protozoa *Plasmodium*, *Leishmania*, and *Trypanosoma*) will soon be available. However, the most rapid progress in the elucidation of gene function will initially be made using model organisms. Yeast is an excellent eukaryotic model and at least 40% of single-gene determinants of human heritable diseases find homologues in yeast. We have adopted a systematic approach to the functional analysis of the *Saccharomyces cerevisiae* genome. A number of the approaches for the functional analysis of novel yeast genes are discussed. The different approaches are grouped into four domains: genome, transcriptome, proteome, and metabolome. The utility of genetic, biochemical, and physico-chemical methods for the analysis of these domains is discussed, and the importance of framing precise biological questions, when using these comprehensive analytical methods, is emphasized. Finally, the prospects for elucidating the function of protozoan genes by using the methods pioneered with yeast, and even exploiting *Saccharomyces* itself, as a surrogate, are explored.

Keywords: functional genomics; yeast; *Saccharomyces cerevisiae*; transcriptome; proteome; metabolome

1. INTRODUCTION

The availability of complete genome sequences for a number of organisms places biology in a situation that is almost without parallel in the history of science. It is very rare, in any major field of scientific endeavour, for researchers to be presented with a complete dataset. If we are to take advantage of these complete inventories of genes, we need similarly comprehensive methods of analysis at a number of other levels. These levels are the famous (or infamous) 'omes'. In addition to the genome (the complete set of genes for an organism and its organelles), they comprise the transcriptome (the complete set of mRNA molecules), the proteome (the complete set of proteins), and the metabolome (the complete set of metabolites, the low molecular weight intermediates). The transcriptome, proteome, and metabolome are all context dependent; they vary according to the physiological, developmental, or pathological state of the cell, tissue, organ, or organism. It is this context dependency that makes them so valuable in the analysis of gene function.

These 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 comprehensive way using the massively parallel technique of hybridization-array analysis. However, mRNA molecules are not functional entities within the cell. They are just transmitters of the instructions for synthesizing proteins, and so transcriptome

analyses only approach functionality in an indirect manner. While both proteins and metabolites are functional entities within cells, the analysis of the proteome is very difficult, and most current approaches 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 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 functions of novel genes.

It is to be hoped that the availability of complete genome sequences, and the consequent access to the comprehensive technologies of functional genomic analysis, will greatly accelerate the study of organisms, such as the protozoan parasites, that are not amenable to classical genetic analysis. In applying these technologies, it will be important that investigators of these less tractable organisms learn from the experiences of those working on model organisms. Moreover, it will sometimes be necessary to use the model species as surrogates in the study of the genes and gene products of the more difficult organisms. This paper gives an account of the lessons to be learnt from investigations on the brewers' and bakers' yeast, *Saccharomyces cerevisiae*. Yeast has long been in the vanguard of genomic research. It was the first organism from which

a complete chromosome was sequenced (Oliver *et al.* 1992) and the first eukaryote for which the complete genome sequence was obtained (Goffeau *et al.* 1996). More recently, a collection of mutant yeast strains, each bearing a defined deletion in one of yeast's 5998 potential protein-encoding genes has been constructed (Winzeler *et al.* 1999; euroscarf@em.uni-frankfurt.de).

2. GENOME-LEVEL ANALYSES

The efficiency and accuracy of mitotic recombination in *S. cerevisiae* mean that it is possible to manipulate its genome with exquisite precision. Single ORFs can be deleted from the genome by a process termed PCR-mediated gene replacement (Baudin *et al.* 1993; Wach *et al.* 1994). If novel phenotypes due to gene deletion are to be reliably delineated, then it is essential that the replacement marker itself, be without phenotypic effect. This is true of the *kanMX* replacement cassette (Wach *et al.* 1994) used in the yeast gene deletion programme (Winzeler *et al.* 1999), but is not true of nutritional markers (see Baganz *et al.* 1997). Alternatively, the marker can be removed by exploiting either homologous (Fairhead *et al.* 1996) or heterologous (Güldener *et al.* 1996) site-specific recombination systems. This has the advantage of allowing markers to be recycled and so facilitates the removal of all members of a gene family from the genome (Delneri *et al.* 1999). This is a basic requirement for addressing the (apparent) redundancy within the genome. Inducing site-specific recombination events also allows the directed reconfiguration of the genome by chromosomal translocations (Delneri *et al.* 2000), a process that is proving useful in evolutionary studies.

In the worldwide yeast gene deletion programme (Winzeler *et al.* 1999), the pairs of oligonucleotides used to prime the polymerase chain reactions that generated each specific ORF-replacement cassette contained 20 nucleotide sequences that are unique for that particular replacement reaction. Thus each deletion mutant is marked by a pair of molecular 'barcodes' that identify the mutant uniquely (Shoemaker *et al.* 1996), this is a boon in the management of a very large mutant collection (such as EUROSCARF, the strain repository for EUROFAN (Oliver 1996a, 1997)), but also opens up a number of possibilities for the quantitative analysis of phenotype.

Gene deletions create null mutations and are expected to be recessive in diploid cells. However, this is only universally true if phenotype is measured *qualitatively*. If, instead, *quantitative* measures are used, then some genes reveal a phenotype when one member of the homologous pair is deleted. This phenomenon is known as haploinsufficiency; it is due to their simply being too little of the gene product in the cell to permit growth and division to occur at the normal rate. Adherents of the theory of MCA (Kacser & Burns 1973) would say that genes displaying haploinsufficiency exert a high level of control over the pathways in which their products function. Indeed, MCA has been used to explain why most mutations are recessive in their effect (Kacser & Burns 1980). In heterozygous diploids, viability is maintained even if an essential gene is deleted (*ca.* 14% of novel yeast genes proved to be essential for growth). Some of these heterozygous deletants will exhibit growth defects due to haploinsufficiency.

Moreover, yeast strains that are heterozygous deletants for the target protein of a drug or toxin have been shown to exhibit enhanced sensitivity to the inhibitory chemical. Giaever *et al.* (1999) have demonstrated that competition between different heterozygous deletant strains in a mixed population may be monitored by PCR amplification of the oligonucleotide barcodes inserted when the deletion mutation was constructed. In this way, they were able to identify novel targets for known inhibitors. The method can also be used to identify, for the first time, the target of some novel drug or toxin.

Competitions may also be carried out between haploid strains deleted for any one of the 5000 or so non-essential genes in yeast and, if the population of mutant cells is grown in continuous culture, very small quantitative effects on flux (growth rate) may be detected by monitoring the change in proportions of the different mutant types in the population (Baganz *et al.* 1998) via detection of their unique barcodes. The barcoding approach, under the name of 'signature-tagged mutagenesis', has also been exploited to run competitions between different mutants of pathogenic organisms during infections of animal hosts (Holden & Hensel 1998). This is a very powerful technique for assessing the contributions of individual genes to virulence and pathogenicity. Moreover, it may even be possible to design better models of infection by adjusting conditions *in vitro* or *in cellulo*, such that the outcome of competition experiments reflects that found in the animal host.

3. FUNCTIONAL GENOMICS VIA THE METABOLOME: A STRATEGY FOR CHARACTERIZING MUTATIONS WITH A SILENT PHENOTYPE

In spite of the sensitive test provided by growth-rate competition analysis, many yeast deletion mutants show no overt phenotype, in terms of either growth rate or other fluxes. However, measurement of the intracellular concentrations of metabolites (metabolome analysis) can, and should, reveal phenotypes for proteins active in metabolic regulation. Quantification of the change of several metabolite concentrations relative to the concentration change of one selected metabolite can reveal the site of action, in the metabolic network, of a silent gene. In the same way, comprehensive analyses of metabolite concentrations in mutants, providing 'metabolic snapshots', can reveal functions when snapshots from strains deleted for novel genes are compared to those deleted for known genes. We have termed this approach to functional analysis, via comparative metabolomics, FANCY (Oliver 1996b; Teusink *et al.* 1998).

The FANCY approach to the elucidation of gene function has a firm theoretical basis in MCA (see Teusink *et al.* 1998). The central device of MCA is the control coefficient (see Fell 1996); this measures the fractional change in either the flux through a metabolic pathway, or the concentration of some metabolite, relative to the fractional change in the activity of some effector (such as an enzyme). FCCs add up to 1 (Kacser & Burns 1973) and tend to adopt values between 0 and 1. Thus, if an enzyme completely determines the rate of flux through a given pathway, it will have an FCC of 1. If its action is com-

pletely irrelevant to the pathway, it will have an FCC of 0. In practice, most enzymes have FCC values closer to 0 than 1. By contrast, CCCs sum to zero, such that they may have substantial values, both positive and negative. Changes in the activity of a single enzyme may produce very large changes in the concentration of specific metabolites, often acting to minimize the effect of the change in enzyme activity on flux (see Teusink *et al.* 1998). Thus, in seeking to elucidate the functions of novel genes through metabolomics, either very sensitive measures of flux (since FCC values are on average close to 0), or comprehensive measures of metabolite concentrations (since CCC values may be large but, if a gene is of unknown function, we do not know which metabolite concentrations to measure) are required.

By quantifying the relative change of the intracellular metabolites in wild-type and mutant strains, it is possible to associate mutants that affect a similar domain of metabolism, and distinguish them from mutants affecting other domains and from wild-type cells. It would be possible to employ this method for systematic functional analysis by constructing a database of co-response indicator profiles for a library of strains, carrying single ORF deletions in all kinds of known genes. Such a database could readily be screened for similarities between the metabolic profiles of known and unknown genes. However, in the absence of any independent indication of the probable domain of activity of a novel gene, this would be a very laborious way of pursuing functional genomics via the metabolome.

An attractive alternative is to exploit methods of metabolite extraction and analysis that are comprehensive in their scope and use them to produce metabolic snapshots of strains deleted for single genes of either known or unknown function. In the early stages of such an analysis, it is not necessary to identify which metabolites the mutations affect. Instead, similar snapshots can be grouped together using statistical techniques, such as principal components and discriminant function analysis. Thus, genes of unknown function can be grouped with those of known function, and the concentrations of relevant metabolites determined later in order to localize the lesion via a comparison of co-response coefficients. While this means that high-throughput techniques, such as FTIR spectroscopy, can be used in the early stages of the analysis, it may be more profitable to employ analytical procedures that permit the subsequent assignment of metabolite identities.

Raamsdonk *et al.* (2001) have provided an example of one such technique, high-resolution NMR spectroscopy, although MS or LC-MS may be equally useful. Their study showed that, although a mutation may cause no significant change in growth rate, silent phenotypes can be revealed by significant changes in concentration of the intracellular metabolites. By appropriate analysis of the relative concentrations of specific metabolites, the site of action of the silent mutation in the metabolic network may be revealed. This approach is obviously useful for the rapid identification of new and interesting silent regulatory genes in various parts of the metabolism of various organisms. However, FANCY is also capable of revealing the function of genes that do not participate directly in metabolism or its control. For instance, a gene involved in amino-acid biosynthesis may, upon deletion, affect the

cellular concentration of one or a few amino acids. A gene involved in protein synthesis, by contrast, may affect the concentration of all amino acids.

4. PROTEOME-LEVEL ANALYSES

The proteome should be the most useful level of functional genomic analysis. It is context dependent, proteins are functional entities within the cell, and there is a direct relationship between genome and proteome. However, many proteins may be encoded by a single gene, due to differential splicing of transcripts, and post-translational processing and modification events. Thus although the yeast genome contains only about 6000 ORFs, and the differential splicing of transcripts is almost unknown, the potential size of the proteome is probably 9–10 000 mature protein species (this estimate is based on the data of Perrot *et al.* 1999). This 'protein gap' is likely to be much larger for more complex organisms, and estimates of the total number of mature protein species encoded by the 30–45 000 genes of the human genome range from 60 000 to 150 000 (Claverie 2001; The International Human Genome Sequencing Consortium 2001).

Two-dimensional (2D) gel electrophoresis coupled with peptide-mass fingerprinting (Shevchenko *et al.* 1996; Perrot *et al.* 1999) is potentially a highly parallel way of analysing the proteome. However, its parallelism is absolutely dependent on the reproducibility of the 2D gels, not only from one experiment to another, but also from one laboratory to another. Other problems with this classical route include its low sensitivity and the fact that small proteins, membrane proteins, and proteins with extreme isoelectric points are often missing from the analysis. Moreover, both dimensions of the 2D analysis are denaturing, and this means that valuable information about protein-protein interactions (which can be particularly revealing about function) is lost.

For these reasons, there has been much recent effort to develop techniques that permit the direct identification of the members of protein complexes from peptide mass data. The different approaches have included the use of chemical derivatization to obtain a limited amount of amino-acid sequence information (Summerfield *et al.* 1997; Brancia *et al.* 2000, 2001; Sullivan *et al.* 2001), the use of ion-exchange or size-exclusion chromatography either before (Butt *et al.* 2001) or after (Link *et al.* 1999) tryptic digestion of proteins, and improvements in techniques for the purification of protein complexes by the use of affinity tags (ICATs; Gygi *et al.* 1999a) or a dual-antibody precipitation technique (TAPS; Rappsilber *et al.* 2000).

Valuable as these biochemical analyses of protein interactions are, it is wise to validate them by monitoring such interactions in the living cell. The yeast two-hybrid system (Cagney *et al.* 2000) can be used for this purpose, and can monitor interactions, not just between yeast proteins, but also between heterologous proteins expressed in yeast. However, it should be remembered that the two-hybrid system only monitors *potential* interactions, and not real ones. Moreover, these interactions must be able to occur within the yeast nucleus, no matter in what cellular compartment the two proteins normally find themselves. Although a comprehensive analysis of all potential

protein–protein interactions in yeast has been made (Uetz *et al.* 2000), it revealed a disappointingly small number of interactions. This study monitored interactions between complete yeast proteins, whereas the two-hybrid analysis of interactions between proteins specified by the genome of the bacterium *Helicobacter pylori* examined the interaction between protein domains (Rain *et al.* 2001). Although it was not a comprehensive analysis, the domain approach uncovered far more protein–protein interactions than did the whole-protein yeast study. Thus, as ever, proteome-level analysis requires a lot of hard work, and does not yield its answers easily.

5. TRANSCRIPTOME-LEVEL ANALYSES

The analysis of gene expression at the RNA level through the use of hybridization-array technology is exactly the kind of comprehensive, massively parallel, high-throughput approach that the availability of complete genome sequences demands. The transcriptome is context dependent and is directly related to the genome, but its main drawback remains that it is proteins and not mRNA molecules that are the functional products of the coding sequences within the genome. Therefore, it is very important to define the correlation between the transcriptome and proteome. The two published studies comparing mRNA and protein abundance in yeast (Futcher *et al.* 1999; Gygi *et al.* 1999b) reach opposite conclusions about how well they are correlated. However, in neither case were these two variables measured in the same experiment. Thus, there is an urgent need to compare transcript and protein levels in the same population of yeast cells. However, such comparisons are fraught with problems, mainly due to the nature of the data on protein levels provided by 2D gel electrophoresis. Many spots on 2D gels contain more than one protein, in which case it is impossible (using mass spectral analysis) to determine their relative proportions. On the other hand, there are also single coding sequences that are represented by many protein spots. In this case, the amounts of protein in all of the spots relevant to a given coding sequence may be summed. However, until the proteins contained within every spot on the 2D gel have been identified, it has to be accepted that the estimate of the amount of protein expressed from a given coding sequence may be an underestimate.

The major problem of transcriptome analysis, other than its indirect nature, is its very comprehensiveness. We are used to studying just a small segment of an organism's biological activity in any given experiment. In the case of transcript analyses, this has meant examining the expression of only a few genes at a time, using Northern hybridization. A comparison, for more than 1000 yeast ORFs, of Northern and hybridization-array data (Brown *et al.* 2001) has pointed to the fact that it is essential to only compare different sets of expression data when they address biologically equivalent situations. In addition, such comparisons were shown to be enormously sensitive to differences in data normalization regimes. This emphasizes that it is essential for researchers to have access to each other's raw transcriptome data, so that they may apply the data normalization procedure that is most appropriate to the biological problem they are posing.

Therefore, the current drive for international standards for data deposited in the public repositories (Brazma *et al.* 2000) is very welcome.

The problems of analysing transcriptome data in an appropriate manner are secondary to those of designing adequately controlled experiments to allow biologically valid conclusions to be drawn from these large and comprehensive datasets. Recent experiments in my own laboratory provide a graphic example of such a problem. The impact of a mutation in a gene encoding a transcription factor on the transcriptome profile of yeast was examined by hybridization-array analysis. Messenger RNA was extracted from mid-exponential-phase batch cultures of mutant and wild-type cells, each grown under exactly the same conditions. The two RNA extracts were converted into cDNAs, each labelled with a different fluorescence dye, and then combined and hybridized to microarrays of PCR products for all 6000 ORFs. It was found that approximately 1000 ORFs were downregulated ≥ 2 -fold in mutant cells and fewer than 10 ORFs were upregulated to a similar extent. However, the downregulated genes included many encoding ribosomal proteins, and it appeared probable that the effect was a consequence of the severe reduction in growth rate that the mutation causes. To test this, the experiment was repeated using chemostat culture instead of batch culture. This allowed control of the growth rate via the dilution rate. Accordingly, mutant and wild-type cultures were grown in identical glucose-limited chemostats at the same dilution rate, which set a growth rate well within the capabilities of both strains. Now, the hybridization-array data showed that 50–100 genes were upregulated ≥ 2 -fold and only 10–40 genes were downregulated to a similar extent. The upregulated ORFs all belonged to yeast transposons. Thus the very specific effect of this mutation in a transcription-factor gene was masked until the general growth rate effect was controlled for. It is possible to do such stringently controlled experiments for yeast, but will be difficult for more complex organisms. It is to be hoped that a number of common 'syndromes' can be recognized, such as the growth-rate response, which can be subtracted from the transcriptome data during bioinformatic analysis.

6. YEAST AS A SURROGATE IN THE FUNCTIONAL ANALYSIS OF OTHER ORGANISMS' GENOMES

Although many of the techniques in functional genomics that were pioneered in yeast are readily transferable to other organisms, it remains the case that there is probably no other eukaryote that is as readily manipulated, both genetically and physiologically, as *S. cerevisiae*. While this remains the case, it would be foolish not to exploit yeast directly in the functional analysis of genes from other, less malleable, eukaryotes. For instance, the fidelity and efficiency of homologous mitotic recombination in yeast may be exploited to manipulate genes from other organisms. Homologous recombination in yeast can be used to reconstruct intact human protein-coding sequences from incomplete cDNA clones in the publicly available libraries. Moreover, it should be possible to perform gene replacement or deletion, or to exchange promoters and other regulatory sequences, on genomic clones propagated in *S. cerevisiae* on high-capacity cloning vectors, such as

YACs. The manipulated clones can then be reintegrated at the correct locus within the donor genome by exploiting the extensive sequence homology flanking the gene of interest that YAC clones provide.

Yeast may also be used as a surrogate to speed the experimental demonstration of the function of genes from large, or poorly defined, eukaryotic genomes by complementing deletion mutants of the increasingly well-defined *S. cerevisiae* genome. There have been a number of examples of the functional complementation of lesions in the yeast genome using cDNA clones from totally unrelated eukaryotic microorganisms, including the parasitic protozoan, *Plasmodium falciparum* (Volkman *et al.* 1995; Wooden *et al.* 1997). There is a great disparity between the G + C content of the *P. falciparum* genome (*ca.* 20%) and that of *S. cerevisiae* (42%), and this often leads to the failure of transcomplementation of yeast mutations due to the premature termination of the transcript from the *P. falciparum* coding sequence. The situation with regard to transcomplementation using human cDNA clones is more encouraging and there are quite a number of examples already in the literature (Schild *et al.* 1990; Plon *et al.* 1993; Thon *et al.* 1993; Glerum & Tzagoloff 1994). In all of these cases, the function of both the plasmodial or human gene, and that of its yeast counterpart, were already known. However, in the case of the yeast ORF YDL120w and the human protein frataxin (the product of the gene that determines the heritable neurodegenerative disease, Friedreich's ataxia), the function of neither protein had been elucidated. However, the application of the systematic and hierarchical approach to functional analysis taken by EUROFAN (Oliver 1996a, 1997; Oliver *et al.* 1998) allowed Foury & Cazzalini (1997) to demonstrate that a yeast strain bearing a deletion of YDL120w exhibited a mitochondrial dysfunction, which they located to a fault in iron transport. A similar phenotype was found in cells from patients suffering from Friedreich's ataxia, and transcomplementation of the yeast lesion with a cDNA copy of the human gene was demonstrated (Koutnikova *et al.* 1997; Rotig *et al.* 1997). At the level of amino-acid sequence similarity between frataxin and Ydl120p, it would be expected that 42% of single-gene determinants of human heritable diseases would be capable of transcomplementation in yeast (Foury 1997).

It may be that the proportion of human coding sequences capable of complementing yeast mutations is even larger than *ca.* 40%. There are two reasons for this. First, there are many examples of proteins of dissimilar structure having similar (or even identical) biochemical functions. Second, the draft sequence of the human genome (The International Human Genome Sequencing Consortium 2001; Venter *et al.* 2001) has shown that human beings only have five to six times as many protein-encoding genes as yeast (Claverie 2001). Much of this expansion will be within classes of proteins, such as protein kinases and phosphatases (The International Human Genome Sequencing Consortium 2001) that are well represented in the microbial eukaryote. This implies that it should be possible to divide human gene families into functional groupings according to which yeast mutations they are able to complement or suppress, and how well (or under what conditions) they are able to do this. Thus, transcomplementation of yeast mutations could be a

powerful tool in the functional analysis of the human genome. It should be an even more powerful aid to the functional analysis of other microbial eukaryotes, including fungal and protozoal pathogens of plants and animals.

To test this contention, we have developed a system in which yeast coding sequences are placed under the control of a downregulatable promoter whose expression may be switched off by adding a repressor substance that is not, itself, a metabolite in yeast. To do this, we used the *tetR/tetO* regulatory system developed by Herrero and his colleagues (Gari *et al.* 1997; Belli *et al.* 1998). This is a similar system to that constructed previously for mammalian cells (Gossen & Bujard 1992; Gossen *et al.* 1993). Tetracycline (or, rather, doxycycline) is not a metabolite in yeast and Belli *et al.* (1998) have produced a cassette system (based on the classic positively selectable marker *kanMX*; Wach *et al.* 1994) that allows the one-step substitution of any yeast promoter with a *tetO* promoter that is either positively or negatively regulatable by doxycycline. For our *transcomplementation* system (N. Zhang, P. Gisham & S. G. Oliver, unpublished data), we are constructing a panel of yeast strains, each of which contains an essential yeast gene placed under the negative control of doxycycline. These strains are then transformed with a library of cDNA clones from a heterologous source whose expression is under the control of the *MET3* promoter (Mountain *et al.* 1991). At high doxycycline concentrations, the essential yeast gene will be completely switched off and, hence, full complementation of the lesion by the foreign coding sequence will be demanded. However, in cases where only partial complementation may be achieved, there should be an intermediate concentration of the drug where the expression of the essential yeast gene is downregulated to a sufficient extent that the upregulation of the nearest equivalent foreign gene allows wild-type levels of growth. This 'balanced partial complementation' can be recognized by comparing the growth of untransformed and transformed yeast strains at a range of doxycycline concentrations in plates containing a concentration of methionine that permits an intermediate level of expression of the foreign cDNA, or no methionine, which allows overexpression of the foreign cDNA. Positive clones are verified by demonstrating a complete absence of complementation in the presence of high concentrations of methionine. In trials of this system, we have isolated five human cDNAs that fully complement three essential yeast genes. These include a human cDNA encoding a novel ubiquitin-conjugating enzyme.

7. CONCLUSIONS AND PROSPECTS

The many successes of yeast genomics and functional genomics provide both a conceptual and methodological framework that should be readily exploitable in the study of other microbial eukaryotes, including the parasitic protozoa. The yeast research community is very large, which, despite the competitive nature of research on this important model eukaryote, has always been able to initiate collaborative projects to produce data or research tools for the common good. While other, smaller, research communities would do well to emulate this method, they cannot afford some of the indulgences of the yeast researchers. There needs to be a common agreement on

strains and media, there needs to be an agreed set of specific barcodes for mutants from the very start of any systematic gene knock-out or knock-down programme, and (most of all) there needs to be a single database with a structure capable of integrating the quantitatively and qualitatively diverse data from the different levels of analysis in order to reveal biological meaning. It might be argued that with an organism as experimentally malleable and intensively studied as *S. cerevisiae*, functional genomics has only made a difference to the rate of yeast research, revealing nothing that would not have eventually been discovered by more traditional approaches. However, I think this ignores the tremendous integrative power of functional genomics and its potential for providing a holistic view of cell function. Whatever the rights or wrongs of these arguments, I think it is beyond dispute that functional genomics will make a major qualitative and quantitative difference to our understanding of less tractable organisms, such as the parasitic protozoa. It is to be hoped that the molecular parasitologists will learn the 'lessons from yeast', and I look forward with keen anticipation to observing their progress in the years ahead.

The author is grateful to all members of his research group and his collaborators (past and present) for their contributions. Research on functional genomics in the author's own laboratory has been supported by the BBSRC, the European Commission, the Wellcome Trust, the EPSRC, and Pfizer, AstraZeneca, and Nycomed-Amersham.

REFERENCES

- Baganz, F., Hayes, A., Marren, D., Gardner, D. C. J. & Oliver, S. G. 1997 Suitability of replacement markers for functional analysis studies in *Saccharomyces cerevisiae*. *Yeast* **13**, 1563–1573.
- Baganz, F., Hayes, A., Farquhar, R., Butler, P. R., Gardner, D. C. J. & Oliver, S. G. 1998 Quantitative analysis of yeast gene function using competition experiments in continuous culture. *Yeast* **14**, 1417–1427.
- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. 1993 A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**, 3329–3330.
- Belli, G. E., Piedrafita, L., Aldea, N. & Herrero, E. 1998 An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Res.* **26**, 942–947.
- Brancia, F. L., Oliver, S. G. & Gaskell, S. J. 2000 Improved MALDI mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. *Rapid Commun. Mass Spectr.* **14**, 2070–2073.
- Brancia, F. L., Butt, A., Beynon, R. J., Hubbard, S. J., Gaskell, S. J. & Oliver, S. G. 2001 A combination of chemical derivatisation and improved bioinformatic tools optimises protein identification for proteomics. *Electrophoresis* **22**, 552–559.
- Brazma, A., Robinson, A., Cameron, G. & Ashburner, M. 2000 One-stop shop for microarray data—is a universal, public DNA-microarray database a realistic goal? *Nature* **403**, 699–700.
- Brown, A. J. P. (and 21 others) 2001 Transcript analysis of 1003 novel yeast genes using high-throughput northern hybridisations. *EMBO J.* **20**, 3177–3186.
- Butt, A., Davison, M. D., Gaskell, S. J., Oliver, S. G. & Beynon, R. J. 2001 Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis. *Proteomics* **1**, 42 053.
- Cagney, G., Uetz, P. & Fields, S. 2000 High-throughput screening for protein–protein interactions using two-hybrid assay. *Meth. Enzymol.* **328**, 3–14.
- Claverie, J.-M. 2001 What if there are only 30,000 human genes? *Science* **291**, 1255–1257.
- Delneri, D., Gardner, D. C. J. & Oliver, S. G. 1999 Analysis of the seven-member AAD gene set demonstrates that genetic redundancy in yeast may be more apparent than real. *Genetics* **153**, 1591–1600.
- Delneri, D., Tomlin, G. C., Wixon, J. L., Hutter, A., Sefton, M., Louis, E. J. & Oliver, S. G. 2000 Exploring redundancy in the yeast genome: an improved strategy for use of the cre/loxP system. *Gene* **252**, 127–135.
- Fairhead, C., Llorente, B., Denis, F., Soler, M. & Dujon, B. 1996 New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. *Yeast* **12**, 1439–1458.
- Fell, D. A. 1996 *Understanding the control of metabolism*. London: Portland Press.
- Foury, F. 1997 Human genetic diseases: a cross-talk between man and yeast. *Gene* **195**, 1–10.
- Foury, F. & Cazzalini, O. 1997 Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria. *FEBS Lett.* **411**, 373–377.
- Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. 1999 A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**, 7357–7368.
- Gari, E., Piedrafita, L., Aldea, M. & Herrero, E. 1997 A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* **13**, 837–848.
- Giaever, G., Shoemaker, D. D., Jones, T. W., Liang, H., Winzler, E. A., Astromoff, A. & Davis, R. W. 1999 Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nature Genet.* **21**, 278–283.
- Glerum, D. M. & Tzagoloff, A. 1994 Isolation of a human cDNA for heme A farnesyltransferase by functional complementation of a yeast *cox10* mutant. *Proc. Natl Acad. Sci. USA* **91**, 8452–8456.
- Goffeau, A. (and 15 others) 1996 Life with 6000 genes. *Science* **274**, 546–567.
- Gossen, M. & Bujard, H. 1992 Tight control of gene-expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl Acad. Sci. USA* **89**, 5547–5551.
- Gossen, M., Bonin, A. L. & Bujard, H. 1993 Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. *Trends Biochem. Sci.* **18**, 471–475.
- Güldener, U., Heck, S., Fiedler, T., Beinhauer, J. & Hegemann, J. H. 1996 A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519–2524.
- Gygi, S. P., Rist, B., Gerber, S. A., Turece, F., Gelb, M. H. & Aebersold, R. 1999a Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999.
- Gygi, S. P., Rochon, Y., Franz, B. R. & Aebersold, R. 1999b Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720–1730.
- Holden, D. W. & Hensel, M. 1998 Signature-tagged mutagenesis. *Meth. Microbiol.* **27**, 359–370.
- Kacser, H. & Burns, J. A. 1973 The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104.
- Kacser, H. & Burns, J. A. 1980 The molecular basis of dominance. *Genetics* **97**, 639–666.
- Koutnikova, H., Campuzano, V., Foury, F., Dolle, P., Cazzalini, O. & Koenig, M. 1997 Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. *Nature Genet.* **16**, 345–351.

- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. & Yates III, J. R. 1999 Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**, 676–682.
- Mountain, H. A., Bystrom, A. S., Larsen, J. T. & Korch, C. 1991 Four major transcriptional responses in the methionine/threonine biosynthetic pathway of *Saccharomyces cerevisiae*. *Yeast* **7**, 781–803.
- Oliver, S. G. (and 146 others) 1992 The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38–46.
- Oliver, S. G. 1996a A network approach to the systematic analysis of yeast gene function. *Trends Genet.* **12**, 241–242.
- Oliver, S. G. 1996b From DNA sequence to biological function—the new *Voyage of the Beagle*. *Biochem. Soc. Trans.* **24**, 291–292.
- Oliver, S. G. 1997 EUROFAN's analysis of gene function enters its second phase. *Genome Digest* **4**, 4.
- Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. 1998 Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378.
- Perrot, M., Sagliocco, F., Mini, T., Monribot, C., Schneider, U., Shevchenko, A., Mann, M., Jenö, P. & Boucherie, H. 1999 Two-dimensional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis* **20**, 2280–2298.
- Plon, S. E., Leppig, K. A., Do, H. N. & Groudine, M. 1993 Cloning of the human homolog of the *cdc34* cell cycle gene by complementation in yeast. *Proc. Natl Acad. Sci. USA* **90**, 10 484–10 488.
- Raamsdonk, L. M. (and 12 others) 2001 Functional genomics via the metabolome: a strategy for characterising mutations with a silent phenotype. *Nature Biotechnol.* **19**, 45–50.
- Rain, J. C. (and 12 others) 2001 The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211–215.
- Rappsilber, J., Siniosoglou, S., Hurt, E. C. & Mann, M. 2000 A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Analyt. Chem.* **72**, 267–275.
- Rotig, A., deLonlay, P., Chretien, D., Foury, F., Koenig, M., Sidi, D., Munnich, A. & Rustin, P. 1997 Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich's ataxia. *Nature Genet.* **17**, 215–217.
- Schild, D., Brake, A. J., Kiefer, M. C., Young, D. & Barr, P. J. 1990 Cloning of three human multifunctional *de novo* purine biosynthetic genes by functional complementation of yeast mutations. *Proc. Natl Acad. Sci. USA* **87**, 2916–2920.
- Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. & Mann, M. 1996 Linking genome and proteome by mass-spectrometry: large-scale identification of yeast proteins from two-dimensional gels. *Proc. Natl Acad. Sci. USA* **93**, 14 440–14 445.
- Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. 1996 Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **4**, 450–456.
- Sullivan, A. G., Brancia, F. L., Tyldesly, R., Bateman, R., Sidhu, K., Hubbard, S. J., Oliver, S. G. & Gaskell, S. J. 2001 The exploitation of selective cleavage of singly protonated peptide ions adjacent to aspartic acid residues using a quadrupole orthogonal time-of-flight mass spectrometer equipped with a MALDI source. *Int. J. Mass Spectr.* **210**, 665–676.
- Summerfield, S. G., Bolgar, M. S. & Gaskell, S. J. 1997 Promotion and stabilization of b(1) ions in peptide phenylthiocarbamoyl derivatives: analogies with condensed-phase chemistry. *J. Mass Spec.* **32**, 225–231.
- The International Human Genome Sequencing Consortium 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
- Teusink, B., Baganz, F., Westerhoff, H. V. & Oliver, S. G. 1998 Metabolic control analysis as a tool in the elucidation of the function of novel genes. *Meth. Microbiol.* **26**, 297–336.
- Thon, V. J., Khalil, M. & Cannon, J. F. 1993 Isolation of human glycogen-branching enzyme cDNAs by screening complementation in yeast. *J. Biol. Chem.* **268**, 7509–7513.
- Uetz, P. (and 20 others) 2000 Protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627.
- Venter, J. C. (and 271 others) 2001 The sequence of the human genome. *Science* **291**, 1304–1351.
- Volkman, S. K., Cowman, A. F. & Wirth, D. F. 1995 Functional complementation of the *ste6* gene of *Saccharomyces cerevisiae* with the *pfmdr1* gene of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **92**, 8921–8925.
- Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. 1994 New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808.
- Wooden, J. M., Hartwell, L. H., Vasquez, B. & Sibley, C. H. 1997 Analysis in yeast of antimalaria drugs that target the dihydrofolate reductase of *Plasmodium falciparum*. *Mol. Biochem. Parasitol* **85**, 25–40.
- Winzeler, E. A. (and 51 others) 1999 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906.