Systematic functional analysis of the yeast genome

Stephen G. Oliver, Michael K. Winson, Douglas B. Kell and Frank Baganz

The genome sequence of the yeast Saccharomyces cerevisiae has provided the first complete inventory of the working parts of a eukaryotic cell. The challenge is now to discover what each of the gene products does and how they interact in a living yeast cell. Systematic and comprehensive approaches to elucidation of yeast gene function are discussed and the prospects for the functional genomics of eukaryotic organisms evaluated.

The new field of functional genomics presents yeast researchers, in particular, with new responsibilities and opportunities: the responsibility is to elucidate the function of each and every one of the almost 4000 novel protein-encoding genes discovered by the Saccharomyces cerevisiae Genome Sequencing Project; the opportunity is to determine how all yeast genes, both those that were discovered by classical (function-first) genetics and those that were revealed by the complete genome sequence, interact to allow this simple eukaryotic cell to grow, divide, develop and respond to environmental changes. Thus, functional genomics should not only provide essential information about the role of novel genes, it should also throw new light on the contributions made by the ‘old’ genes.

If this holistic, or fully integrative, view of the yeast cell can be obtained, it should provide an important navigational aid to guide our studies of more complex genomes, such as those of humans, crop plants and farm animals.

Functional genomics requires the development of analytical strategies that are comprehensive and hierarchical. Comprehensive, because we aim to uncover the action and interaction of all of the genes in a given species. Hierarchical, because this daunting task is only possible if we find ways of grouping genes of related function in order to limit the total number of experiments to be performed. Having achieved such a grouping, we can then construct smaller and smaller subgroups, proceeding down the hierarchy of analysis, achieving a closer and closer approximation to the function of each novel gene. At the highest level of this hierarchy, the comprehensive nature of the analytical methods employed requires that special care be taken over the design of experiments. In classical genetics, mutants that show a specific phenotype, such as a requirement for nutrient X or the failure to produce product Y, are isolated. Once the gene defined by the

---

The text includes a list of references at the end, indicating the sources of information used in the article. These references are likely cited within the main body of the text to support the claims and findings presented.
mutation has been identified, it is known without a doubt that it controls the synthesis of X or Y within the target locus. In order to permit gene replacement in any S. cerevisiae strain, including those used in industry, an alternative PCR method was developed in which a gene-disruption cassette with long perfect homology between the short flanking regions and the target locus. To increase the efficiency and accuracy of correct integration, particularly in low efficiency of correct integration, particularly in gene-replacement experiments resulted in a rather low efficiency of correct integration, particularly in strain that do not carry a complete deletion of the marker locus. Large improvements were achieved when kanMX, a completely heterologous dominant resistance gene, was used as a selectable marker. The heterology of the kanMX module means that correct disruption of target ORFs with PCR-generated kanMX casseter containing 35 bp flanking extensions homologous to yeast chromosomal DNA occurs with a very low level of illegitimate integration. Successful targeting of these PCR products is dependent on

Table 1. The current status of the EUROFAN \(^1\) deletion analysis

| Novel open reading frames deleted | 802 |
| Essential open reading frames identified | 99 (12.3%) |
| Deletion cassettes cloned | 692 |
| Cognate genes cloned | 631 |

**FOCUS**

Published; parallel activities are also under way in Canada, large European research network called EUROFAN genes using a systematic and hierarchical approach, a Generation of specific deletion mutants in their hosts. In particular acute when studying the activity of pathogens accurately reflect, both qualitatively and quantitatively, gene products (be they transcripts or proteins) and metabolic intermediates in ways that ensure that they accurately reflect, both qualitatively and quantitatively, their status within the living cell. This problem is particular acute when studying the activity of pathogens in their hosts.

**Generation of specific deletion mutants**

In order to determine the function of novel yeast genes using a systematic and hierarchal approach, a large European research network called EUROFAN (European functional analysis network) has been established. Parallel activities are also under way in Canada, Japan and the USA. The first aim of these projects is the production of a complete set of 6000 single-gene deletion mutants covering all the open reading-frames (ORFs) revealed by the complete S. cerevisiae genome sequence. Such a mutant collection will represent a major resource, not only for the functional analysis of the yeast genome but also for facilitating the analysis of higher genomes by permitting their functional mapping onto that of yeast through transcomplementation experiments.

These deletions are being generated by a PCR-mediated gene-replacement method that relies on the high efficiency and accuracy of mutagenic recombination in S. cerevisiae. This technique was first described for PCR-generated DNA molecules consisting of the S. cerevisiae HIS3 gene flanked by 30–50 bp of DNA homologous to the target locus. However, the use of homologous nutritional markers (e.g. HIS3 or TRP1) in gene-replacement experiments resulted in a rather low efficiency of correct integration, particularly in strains that do not carry a complete deletion of the marker locus. Large improvements were achieved when kanMX, a completely heterologous dominant resistance gene, was used as a selectable marker. The heterology of the kanMX module means that correct disruption of target ORFs with PCR-generated kanMX cassette containing 35 bp flanking extensions homologous to yeast chromosomal DNA occurs with a very low level of illegitimate integration. Successful targeting of these PCR products is dependent on

**The transcriptome**

The quantitative analysis of expression levels of a novel gene under a variety of physiological or developmental conditions is a powerful tool in the elucidation of gene function in comparison with expression patterns of functionally characterized genes. One approach to the efficient analysis of gene expression on a genome-wide scale is the use of SAGE (serial analysis of gene expression) technology, which was employed by Velculescu and co-workers to determine the complete set of yeast genes expressed under a given set of conditions (the transcriptome). In their initial studies, these workers were able to detect 4665 genes (approximately 75% of the predicted protein-encoding ORFs), with most genes being expressed at a low level. Kal has used the SAGE technology to compare yeast gene transcription in cells grown on oleate and cells grown on glucose (PhD thesis, University of Amsterdam, Amsterdam, The Netherlands, 1997). Peroxinsomal activity is induced by growth on oleate and the 35 genes showing the highest oleate-glucose transcript ratio included many known to encode peroxisomal proteins, but 15 novel genes were identified as being likely to be associated with peroxinsomal metabolism.

An alternative approach to SAGE is the use of hybridization-array technologies for the high-throughput analysis of the transcriptome. DeRisi et al. have applied DNA microarrays containing almost every yeast ORF to carry out a comprehensive study of gene expression during the metabolic shift from fermentation (involving the production of ethanol from glucose by glycolysis) to respiration (involving the aerobic utilization of the ethanol once glucose has been exhausted) on a genomic scale. They found marked changes in the global pattern of gene expression in approximately 35% of the ~6000 putative protein-coding genes as glucose was progressively depleted from the medium. About half of these differentially expressed genes were previously uncharacterized. Their response to the catabolic shift thus allowed them to be grouped together at the start of the hierarchy of analysis. The expression profiles observed for genes with known functions provided insights into the cells' response to a changing environment. For example, several classes of genes, such as those involved in the TCA/glyoxylate cycle and carbohydrate storage, were coordinately induced by glucose exhaustion, whereas genes involved in protein synthesis exhibited a decrease in expression.

There are other routes to grouping genes on the basis of similarities in their expression patterns. When previously characterized members of such groups share important similarities in their functions, common regulatory mechanisms may be inferred and the likely function of novel genes in the group may be guessed at. For instance, DeRisi et al. showed that the same DNA
microarrays may be used to identify genes whose expression was affected by deletion of the gene encoding the transcriptional activator gene YAP1. These results demonstrate the feasibility and utility of the DNA-microarray methodology in the analysis of the transcriptome and indicate that this method should enable the classification of all novel genes according to their expression patterns.

These DNA microarrays consisted of PCR products produced using primers complementary to the ends of ORFs. An alternative is to use oligonucleotides synthesized in situ on a chip using photolithography. Both these techniques have been reviewed by Hohn et al.21. In a first application of this technique to the yeast transcriptome, Widicka et al.22 compared the transcripts present in yeast cultured on a rich, glucose-containing medium with those of cells grown on a minimal medium containing the same carbon source. Although convincing evidence is provided for the potential of the oligo–chip technology, there is little of biological interest in this particular experiment owing to the large number of variables involved.

The proteome

Another route to the elucidation of gene function is the analysis of all yeast proteins synthesized under a given set of conditions, the so-called proteome.23 The yeast proteome is being defined by two-dimensional gel electrophoresis24 using mass spectrometry to identify the proteins within the spots on the gel25,26. By taking full advantage of the complete genome sequence, it should be possible to determine a single protein fragment of unique mass in order to identify a yeast protein. The use of rapid and accurate mass–spectrometric methods should then allow one to monitor changes in the expression of a novel gene. In the long term, this information will enable the identification of all components of protein complexes (isolated either by biochemical separation or immunoprecipitation) and thus create a map of all the interactions between the yeast proteins, which can be correlated with in situ data obtained with yeast two-hybrid systems. Such a comprehensive two-hybrid approach has already generated a protein-linkage map for the Escherichia coli bacteriophage T7 genome, which encodes 55 proteins.27 In addition, a highly selective two-hybrid procedure adapted for exhaustive screens of the yeast genome is currently being developed by Promont-Racine and co-workers28, because it is known that two-hybrid experiments can sometimes produce false-positive results.

Quantitative phenotypic analysis

One reason for taking a quantitative approach to the analysis of gene function is that there is a growing number of genes found by systematic genome sequencing projects that have homologues of unknown function in a number of species but that had not been discovered by classical molecular genetics. It is possible that these genes have been missed because quantitative, rather than qualitative, data are required to reveal their phenotypic effect. Another reason to consider a quantitative approach to phenotypic analysis is the high level of redundancy that is apparent in the yeast genome. If a particular gene is a member of a paralogous set of identical (or nearly identical) genes then, provided that they are all regulated in a similar manner and their products are targeted to the same cellular location, the contribution that an individual member of the set makes to the phenotype will, necessarily, be some fraction of the whole.

A conceptual and mathematical framework for the quantitative analysis of gene function is provided by metabolic control analysis (MCA) as pioneered by Kacser and Burns29,30, and by Heinrich and Rapoport31. The central mathematical device of MCA is the control coefficient (\(C_j\)), a dimensionless number expressing the relative change in a variable (a flux or metabolite concentration) that is caused by a small change in the activity of an effector (usually, but not necessarily, an enzyme). Flux-control coefficients express the proportional contribution that any given enzyme makes to the flux through a metabolic pathway. The sum of all the flux-control coefficients (\(\sum C_j\)) relevant to a given pathway is 1. Thus, in linear pathways, individual flux-control coefficients will normally have values between 0 (no control) and 1 (complete rate-determining enzyme) but, in branched pathways, values outside this range (i.e. < 0 or > 1) can occur. By contrast, concentration-control coefficients, reflecting the effect of a change in an enzyme’s activity on the concentration of some low molecular weight intermediate, may adopt any value. More detailed reviews of these concepts are provided by Fell32 and Cornish-Bowden33, an explanation for molecular biologists is provided by Tenick et al.34, and an example of large changes in metabolic concentration occurring at a constant flux is given by Mendes et al.35.

‘Top-down’ MCA36 and the related modular method37 have some of the hierarchical features that are required for a systematic approach to the elucidation of gene function.38 This approach aims to divide the metabolic map into a number of large units or modules through, or in, which independent fluxes or metabolite concentrations can be measured. Assignment of a novel gene as having its effect within a particular module then allows that module to be subdivided into a number of individual units that can be tested by further flux or metabolite analyses, to determine in which unit the novel gene under examination has its effect. By moving down this hierarchy of analysis, a closer and closer approximation to the function of the novel gene is obtained.

At the highest level of a top-down MCA approach to yeast gene function, we would wish to define the whole cell as the system. This means that we can define the flux-control coefficient of a particular gene product by manipulating its cellular concentration and measuring the impact of the change on growth rate, to obtain (in effect) a growth-rate-control coefficient. The experimental determination of flux-control coefficients38 demonstrated that they usually have values much closer to 0 than to 1, which indicates that we need a very sensitive method of measuring the change in growth rate resulting from the specific deletion of a novel yeast gene. Competition experiments between mutant and wild-type yeast have been shown...
Colonies of Saccharomyces cerevisiae cells grown with glucose or glycerol as main carbon source were analysed by diffuse reflectance absorbance FTIR (four replicates) in the mid-infrared range (4000–600 cm⁻¹). Principal component analysis (PCA) was performed on the spectral data to reduce dimensionality whilst preserving most of the variance. Discriminant-function analysis (DFA) was then used to separate the samples into groups of replicates. The letters represent the following genetic backgrounds: A, FY23 wild-type; B, FY23 pet191::kanMX4; C, FY23 cox5A::kanMX4; D, FY23 ho::kanMX4; and E, the FY23 wild type. The strains grown on glycerol (A,B) group together but are separate from the 100% respiratory-deficiency mutants FY23 pet191::kanMX4 (C) and FY23 cox5A::kanMX4 (D), and from ho::kanMX4 (B) and the FY23 wild type (A).

An attractive approach to competition analysis, called genetic footprinting, has been developed by Smith and co-workers. In this approach, yeast mutants generated by Ty1 transposon insertions were grown in large populations under different selections using serial batch transfers to extend the period of the competition. The relative proportions of the different mutant strains in the population were monitored by PCR, using a common primer complementary to the Ty sequence and a series of fluorescently labelled primers complementary to the flanking sequences of the genes whose quantitative phenotype was to be assessed. A sequencing machine equipped with GeneScan™ software was used to analyse the PCR products. This method has been used for the quantitative phenotypic analysis of yeast chromosome V (Ref. 38). Data were obtained for 261 (97%) of the predicted protein-encoding genes showing, for approximately 60%, a detectable reduction in fitness in one or more of seven different selection protocols; for the remaining genes, no phenotypic effect was found. Included in the list of phenotypes were novel or unexpected effects for a number of known genes. This result confirms the notion that systematic functional analysis can provide important information that will not only permit the elucidation of the function of novel genes but also provide a more integrated view of the cellular role of previously known genes.

In an approach more suited to application in top-down MCA, because measurements of growth-rate differences are made at steady-state in chemostat cultures, Baganz et al. have carried out competition experiments using specific deletion mutants. Control studies showed both that the replacement marker KanMX (Ref. 10) is phenotypically neutral (in contrast to nutritional markers such as HIS3) and that the deletion of the HO gene by KanMX replacement was also without phenotypic impact. HO was chosen because it has no known role apart from mating-type switching and it has been used as the site of insertion of heterologous genes in brewing yeast without any perceptible effect on the fermentation characteristics of the organism or the quality of the product (Yocum, R., unpublished). Whenever possible, a diploid strain homozygous for the ho deletion is employed, both because HO is inactive in diploids and because the mating of independently derived haploid transformants allows any transformation-induced genetic lesions to be nullified through complementation.

In their experiments, Baganz and co-workers compared test strains bearing KanMX replacements of specific ORFs with an isogenic standard strain that carries a KanMX replacement of HO. This allows the accurate quantitation of the impact of a specific single-gene deletion on growth rate. Moreover, the deletion mutant can subsequently be used for more specific phenotypic analyses. The proportions of mutant and standard-strain cells in the population were measured by quantitative PCR, using either densitometry or the ABI GeneScan™ system. Both methods have a rather limited potential for multiplexing, although DNA extracts from cultures can be aliquoted so that multiple PCR reactions may be carried out to compare the proportions of the standard strain and any number of test deletants. However, the best approach for increasing the number of mutants that may be analysed against the standard strain in a single competition experiment has been presented by the development of a system to generate deletion mutants that are labelled with unique oligonucleotide tags, called molecular bar codes, that can be detected by hybridization to a high-density oligonucleotide array. This method should greatly accelerate the systematic analysis of the deletion mutants on a genome-wide scale.
The metabolome

The second type of data required for the MCA approach is the measurement of the change in the relative concentrations of metabolites as a result of the deletion or overexpression of a gene. In contrast to the determination of flux-control coefficients, where very sensitive and discriminatory analytical tools are required, the determination of effective metabolite-concentration-control coefficients requires analyses that are very comprehensive in their scope, because (for any particular novel gene) it is not known which metabolites will have their levels altered as a result of the gene’s deletion or overexpression (or even in which direction such alterations will occur). Tsuchink et al. proposed an analytical strategy termed FANCY (functional analysis by co-responses in yeast) that is based on theoretical work that we will discuss in detail and takes advantage of the fact that the number of metabolites is an order of magnitude lower than the number of genes, and that the functions of about 40% of these genes are already known.

Tibbetts and co-workers have shown that some enzymes can be grouped into monofunctional units, in which any perturbation of the enzymes inside the unit will always produce the same co-responses outside the unit, regardless of which enzyme was affected and the magnitude of the perturbation. This property can be exploited for functional analysis by identifying mono-functional units all over the metabolic map of yeast. For example, when two deletion mutants, have the same co-responses, they will affect the same monofunctional unit. If the unit in which one of the genes causes an effect is known, that of the other can be inferred and the origin of the changes in metabolites thereby located. Such an approach requires a fast and reliable method of measuring the concentrations of as many metabolites as possible to produce a metabolic snapshot of each deletion mutant. Oliver and co-workers are developing a two-stage strategy to obtain such data. In the first phase, deletants of novel genes are grouped using techniques such as tandem mass spectrometry, and in the second phase, using techniques such as infrared (FTIR) spectroscopy to the analysis of mutants showing quantitative differences in respiratory activity is shown in Fig. 1. This information then determines the type of metabolic snapshot to be taken in the second phase, using techniques such as tandem mass spectrometry. The combination of the metabolic profile of a mutant together with sophisticated chemometrics should allow the target site of a novel gene product (or, indeed, inhibitor) to be located on the metabolic map of yeast. These measurements are often best made at steady state, which can be achieved using chemostat cultures.

Conclusion

It is evident that a wide range of experimental approaches are being developed for use in S. cerevisiae that will allow functional genomics to build up an integrative view of the workings of a simple eukaryotic cell. This should enable a deeper understanding of more complex eukaryotes, both by the identification of orthologous genes in different species and also by the expression of foreign coding sequences in yeast for complementation or two-hybrid analyses. However, many of these techniques are sufficiently general that, once they have been tried and tested in the experimentally tractable yeast system, they should be directly applicable to the study of the functional genomics of higher organisms.

Acknowledgments

Work on yeast genome analysis in SGO’s laboratory has been supported by the European Commission (both the Yeast Genome Sequencing Network and EUROFAN), the BBSRC, the Wellcome Trust, Pfizer Central Research, Applied Biosystems, Amersham International and Zeneca. FB is grateful to Pfizer Central Research for the provision of a postgraduate research studentship. Work on the analysis of metabolism in DBK’s laboratory is supported by the BBSRC, GlaxoWellcome and Bruker.

References

Focus

40 Haber, J. E. (1992) Trends Gastroenterol. 2, 446–452