A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations

Léonie M. Raamsdonk[1], Bas Teusink[1,2], David Broadhurst[3], Nianshu Zhang[4], Andrew Hayes[4], Michael C. Walsh[1,5], Jan A. Berden[1], Kevin M. Brindle[6], Douglas B. Kell[3], Jem J. Rowland[7], Hans V. Westerhoff[1,8], Karel van Dam[1], and Stephen G. Oliver[4]*

*[1]Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, University of Amsterdam, Plantage Muidergracht 12, NL-1018 TV Amsterdam, The Netherlands. [2]Current address: TNO-Prevention and Health, Zernikedreef 9, NL-2333 CK Leiden, The Netherlands. [3]Institute of Biological Sciences, Cledwyn Building, University of Wales, Aberystwyth, Aberystwyth SY23 3DD, UK. [4]School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK. [5]Current address: Heineken Technical Services B.V., P.O. Box 510, NL-2380 BB Zoeterwoude, The Netherlands. [6]Department of Biochemistry, University of Cambridge, Old Addenbrooke's Site, 80 Tennis Court Road, Cambridge CB2 1GA, UK. [7]Department of Computer Science, University of Wales, Aberystwyth, Aberystwyth SY23 3DB, UK. [8]Department of Molecular Cell Physiology, BioCentrum Amsterdam, Free University, De Boelelaan 1087, NL-1081 HV, Amsterdam, The Netherlands. *Corresponding author (steve.oliver@man.ac.uk).*

Received 29 June 2000; accepted 24 August 2000

A large proportion of the 6,000 genes present in the genome of *Saccharomyces cerevisiae*, and of those sequenced in other organisms, encode proteins of unknown function. Many of these genes are "silent," that is, they show no overt phenotype, in terms of growth rate or other fluxes, when they are deleted from the genome. We demonstrate how the intracellular concentrations of metabolites can 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 unstudied genes are compared to those deleted for known genes. This approach to functional analysis, using comparative metabolomics, we call FANCY—an abbreviation for functional analysis by co-responses in yeast.

Keywords: functional genomics, metabolome, yeast, *Saccharomyces cerevisiae*, silent mutations, phenotype analysis, metabolic control analysis, co-response analysis

Functional genomics[1,2] seeks to reverse the usual path of genetic analysis and move from DNA sequence to biological function[3], thus revealing the roles of genes discovered by determining the complete genome sequence of an organism. The systematic analysis of gene function is a much more complex and open-ended enterprise than was the systematic genome sequencing that preceded it. Thus, comprehensive methods of analysis are employed at several levels: those of the genome, transcriptome, proteome, and metabolome[4]. This paper reports a conceptual and experimental framework for the elucidation of gene function by analysis of the metabolome. Metabolomics shares two important advantages with proteomics in terms of the elucidation of gene function. Both are context-dependent, that is, the total complement of proteins or metabolites changes according to the physiological, developmental, or pathological state of a cell, tissue, organ, or organism. Moreover, unlike messenger RNA (mRNA) molecules (the subject of transcriptome analysis), proteins and metabolites are functional entities within the cell. A third advantage of metabolomics, which it shares with neither proteomics nor transcriptomics, is also its greatest difficulty. For many organisms, there are far fewer metabolites than genes or gene products. Thus, for the single-celled eukaryote, *S. cerevisiae*, there are fewer than 600 low-molecular-weight intermediates[5], whereas there are ~6,000 protein-encoding genes[6]. However, this apparent simplicity means that there is no direct relationship between metabolite and gene in the way that there is for mRNAs and proteins. To overcome this difficulty, we have proposed that the genes of known function should be exploited in the elucidation of the role of unstudied genes in an approach that we call FANCY, for functional analysis by co-responses in yeast[7]. In this paper, we exemplify the FANCY approach by studies on *S. cerevisiae*, and demonstrate that it may be used to reveal the role of genes that produce no overt phenotype when deleted from the yeast genome.

The approach is based on the idea that the growth rate of a deletant may not be much changed, precisely because the concentrations of intracellular metabolites have altered so as to compensate for the effect of the mutation. By implication, mutants that are silent when scored on the basis of metabolic fluxes (such as growth rate[8–10]) should produce a very obvious effect on metabolite concentrations[7]. Accordingly, metabolome analysis should reveal a phenotype in many previously silent mutants. The deletion or overexpression of a single protein-encoding gene may well cause the concentrations of many metabolites to change simultaneously. In principle, quantifying the relative changes of metabolite concentrations caused by a mutation in a silent gene should make it possible to identify the site of action of that gene's product. In this paper, we develop this principle.

Co-response coefficients[11,12] are defined as the relative steady-state response of two variables (e.g., internal metabolite concentrations) to the change of a system parameter (e.g., a mutation in a

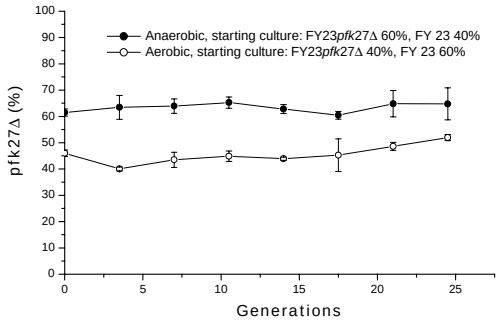

Figure 1. Competition between FY23pfk27Δ and its wild-type parent. Both wild-type and deletant cells were grown to steady state in separate fermenters, under aerobic conditions (dilution rate, 0.1 h^{-1}). Then 40% of the culture volume was exchanged between the two fermenters. One fermenter was set for anaerobic growth (O_2 at 0.6% saturation), the other was left for aerobic growth (O_2 at 100% saturation).

silent gene). Their measurement is central to the FANCY approach. When a (silent) gene encoding a product of unknown function is deleted, the resulting co-response coefficient profile will be similar to that of a strain that is deleted for a (known) gene acting on the same functional domain of the cell[13–15]. Should the gene product act in another part of metabolism, the metabolite concentrations will change in a different way, or will not change at all, resulting in a different co-response coefficient profile. The problem is that, for genes of unknown function, the investigator does not know which metabolites' co-response coefficients should be measured. Thus, comprehensive methods of metabolite analysis are required[5,7,16,17]. In this paper, we show (for a series of known genes, and using both specific and comprehensive methods of analysis) that the measurement of co-response coefficient profiles can reveal the site of action of the product of a silent *S. cerevisiae* gene, even when deletion of that gene has no measurable effect on yeast growth rate.

Results

The test system. We wished to employ an experimental system that would provide a stringent test of the FANCY approach to the elucidation of gene function by metabolome analysis. To this end, we chose to examine the effect of the separate deletion of the yeast *PFK26* and *PFK27* genes, and show that these two genes are silent in terms of growth rate phenotype, but that they do have a metabolome phenotype. Because these metabolome phenotypes can be classified as similar, this will demonstrate that, had the gene function of *PFK27* been unknown, it could have been found on the basis of the gene function of *PFK26*.

PFK26 and *PFK27* encode the same enzyme, 6-phosphofructo-2-kinase (6PF-2-K; EC 2.7.1.105), which catalyzes the conversion of fructose-6-phosphate (F6P) into fructose-2,6-bisphosphate (F2,6bP)[18–21]. This reaction represents a dead-end branch to the Embden–Meyerhof pathway and, although the flux through this branch is not expected to be great, F2,6bP is a strong activator of 6-phosphofructo-1-kinase and an inhibitor of fructose 1,6-bisphosphate-1-phosphohydrolase[22–24]. When both genes are deleted, no 6PF-2-K activity is detectable. Strains in which only one of the two genes has been deleted have been reported to show similar growth rates to wild-type strains on all carbon sources tested[21].

Growth competition experiments fail to reveal phenotypes for the *pfk26* and *pfk27* deletions. Before proceeding with the metabolite analyses, we first checked the impact of single *pfk26* or *pfk27* deletions on growth rate, using a very sensitive technique. Competition experiments in chemostat culture represent a very sensitive approach to the measurement of the effects of single-gene deletions. This method is capable of distinguishing between genes whose deletion has a qualitatively similar, but quantitatively different, phenotypic effect[9,25]. Accordingly, we carried out growth rate competitions between the FY23.pfk26Δ and FY23.pfk27Δ deletion strains and their FY23 parent under both glucose-limited aerobic and glucose-limited anaerobic conditions. The results (shown for FY23.pfk27Δ, only, in Fig. 1) demonstrated that we were unable to detect any selective impact of either deletion under these two growth conditions. Therefore, both *pfk26*Δ and *pfk27*Δ represent exactly the kind of phenotypically silent mutations that we required for a stringent test of the FANCY approach.

FANCY of relevant metabolites can discriminate between different mutant types. As an initial test of the ability of FANCY to reveal the phenotype (and likely site of action) of apparently silent mutations, we analyzed the concentrations of a number of specific metabolites in the parent strain FY23 and five mutant strains. The latter were derived directly from FY23 by the replacement of single open reading frames (ORFs) with the *kan*MX marker, using the short flanking homology (SFH) technique[26]. In each case, the *kan*MX marker was left in the chromosome at the site of the deleted ORF. Because of this, the mutant FY 23.hoΔ was included as a control, since the replacement of the *HO* gene is phenotypically neutral[25,27]. The other four mutant strains were separately deleted for the *PFK26*, *PFK27*, *PET191*, and *COX5a* genes. *PET191* organizes the assembly of the cytochrome oxidase complex, whereas *COX5a* encodes a protein subunit of the same complex. Deletion of *PET191* results in complete respiratory deficiency, whereas *cox5a* deletants retain some 10–15% respiratory activity[28] as a result of the activity of the *COX5b* paralog. Thus FY23.pet191Δ and FY23.cox5aΔ represent deletion mutants that are impaired in a domain of energy metabolism different from that affected by the *pfk26Δ* and *pfk27Δ* mutations. They have qualitatively similar, but quantitatively different, impacts on phenotype[9].

The intracellular concentrations of a number of glycolytic intermediates and products were measured in the mid-exponential phase of such a batch culture of each strain (see Experimental Protocol). Importantly, the pfk26Δ and the pfk27Δ strains were found to have metabolome phenotypes: they exhibited elevated concentrations of F6P and, perhaps, pyruvate (Pyr) (Table 2). Moreover, their ATP/ADP ratios were decreased compared to that of the wild-type parent strain, FY23. Also the two strains that had a reduced respiratory capacity, FY23.pet191Δ and FY23.cox5aΔ, now had phenotypes, showing decreases in ATP/ADP ratio and increases in intracellular pyruvate concentration. Their intracellular levels of G6P had decreased, while those of F6P had increased. In spite of this, all six strains exhibited a similar growth rate on minimal medium with glucose as a carbon source as measured from their batch culture kinetics (Table 2). This illustrates our point that strains without phenotype in terms of flux should have a phenotype in terms of the concentrations of metabolites that are involved in flux homeostasis (cf. ref. 7). As expected, the hoΔ strain did not show any significant differences in the metabolite profile as compared to the wild type.

Table 1. The *S. cerevisiae* strains used in this study

Strain	Genotype	Reference
FY23 (wild-type)	*MATα ura3-52 trp1-Δ63 leu2-Δ1*	Ref. 54
FY23.pfk26Δ	*MATα ura3-52 trp1-Δ63 leu2-Δ1 pfk26Δ::KanMX4*	This study
FY23.pfk27Δ	*MATα ura3-52 trp1-Δ63 leu2-Δ1 pfk27Δ::KanMX4*	This study
FY23.pet191Δ	*MATα ura3-52 trp1-Δ63 leu2-Δ1 pet191Δ::KanMX4*	Ref. 28
FY23.cox5aΔ	*MATα ura3-52 trp1-Δ63 leu2-Δ1 cox5aΔ::KanMX4*	Ref. 28
FY23.ρ^0	*MATα ura3-52 trp1-Δ63 leu2-Δ1* [ρ^0]	Ref. 28
FY 23.hoΔ	*MATα ura3-52 trp1-Δ63 leu2-Δ1 hoΔ::KanMX4*	Ref. 25

Now that we had phenotypes for the previously "silent" strains, the issue was whether we could classify the phenotypes such that they reflected gene function. We decided to perform co-response analysis, that is, to examine whether the measured concentrations of the metabolites varied in the same or in different directions. For the strains that had a metabolome phenotype, we calculated the changes in the logarithm of the metabolite concentrations for each mutant strain versus the wild type. We then took the arctangents of ratios of these relative changes, which correspond to co-response indicators, or the relative directions of the changes in two metabolites (in degrees between –90° and +90°; Table 3). The profiles of the co-response indicators of the pfk26Δ and pfk27Δ strains were similar. This can be appreciated from the similar directions, in degrees of the co-response, between the various metabolites and G6P, or by just reading the signs on the values of their co-response indicators (+ means that the two metabolites increase or decrease together; – means that one metabolite increases whereas the other metabolite decreases). The co-response indicators calculated for the two respiratory-deficient mutants each exhibited a profile that was very different from those of the *pfk*⁻ mutants, but similar to that of each other.

Thus this co-response analysis demonstrated the practical feasibility of the FANCY approach to the elucidation of function. However, it used data on the intracellular concentration of specific metabolites, relevant to the domain of metabolism of the genes under study. For functional genomics, this limitation to a small set of metabolites is a disadvantage because, for a gene of unknown function, it would be impossible to select the correct set of metabolites in advance. It was therefore required to generalize the FANCY approach to examine all, or a large and arbitrary set of, metabolites. This demands the adoption of a comprehensive, rather than specific, method of metabolite analysis. We now show that the FANCY method can be applied to the analysis of changes of the metabolome at large, even without identifying the actual metabolites whose concentrations have changed.

Multivariate FANCY: cell extract analysis clusters genes into functional categories. In this further development of FANCY, we have evaluated three different physical methods for the comprehensive analysis of metabolites within yeast cell extracts. These comprised Fourier-transform infrared spectroscopy (FTIR[5,29,30]), electrospray mass spectrometry (ES-MS [31,32]), and nuclear magnetic resonance (NMR) spectroscopy. High-resolution ^{1}H-NMR spectroscopy has been used previously to analyze metabolite changes in human and animal body fluids and tissue extracts in different disease states and after treatment with drugs and toxins[33]. The technique has the advantage that it is, in principle, capable of detecting any proton-containing metabolite present in the tissue extract provided that it is present above a minimum threshold concentration. The range of metabolites detectable by this method could be extended by changing the method of sample preparation or by detecting other nuclei, for example ^{31}P or natural-abundance ^{13}C. Although we have not assigned the peaks in the spectra obtained from these preparations, this is nevertheless a relatively straightforward procedure.

Metabolites were extracted from mid-exponential phase cultures of FY23 and six mutant strains, grown aerobically on a minimal glucose medium. (The extra strain investigated by this approach was a cytoplasmic petite mutant, FY23.ρ^0. This mutant should be qualitatively and quantitatively identical to the completely respiratory-deficient nuclear petite, FY23.pet191Δ.) The extracts from each strain were then analyzed using ^{1}H-NMR spectroscopy (see Experimental Protocol).

The power of the NMR approach, as compared to the enzymatic analysis approach of Table 3, is that (for a limited investment in time) it measures changes in many more (in principle, arbitrary)

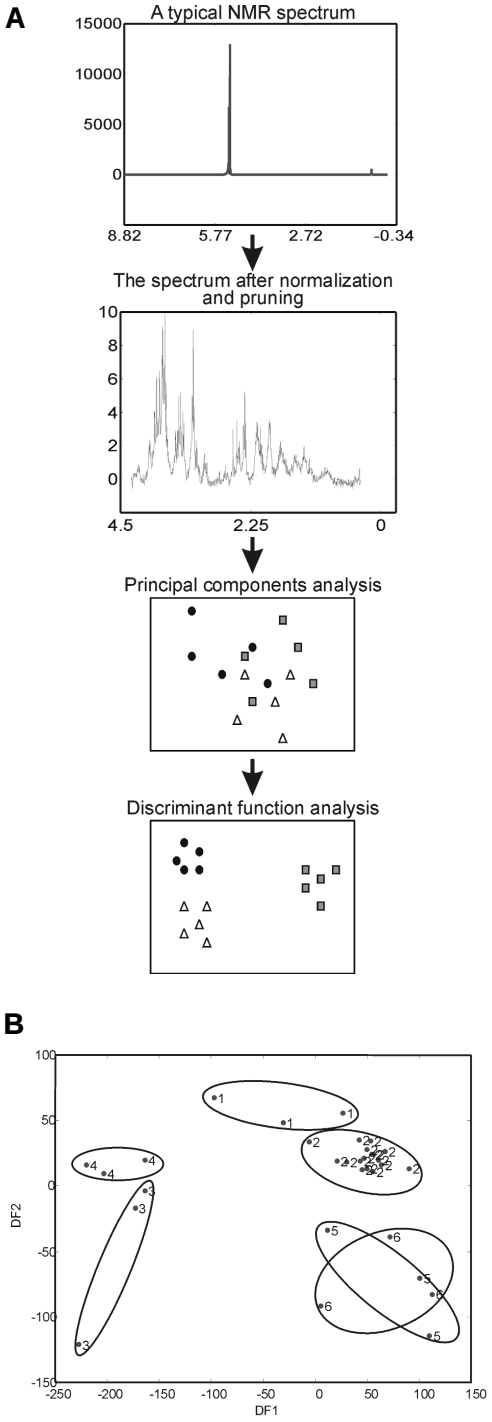

Figure 2. Cluster analysis of NMR spectra from cell extracts. (A) Flowchart to show chemometric approach used to cluster the NMR data. Step 1: The region around the most prominent NMR peak due to water (4.4 – 5.5 p.p.m.) is removed and the internal standard (0 p.p.m.) used to normalize each ordinate (thus allowing quantitative comparison of spectra) before it, too, is removed from the spectrum. The region beyond 5.5 p.p.m. (the aromatic region) is also removed. The resulting reduced data set describes the subspectral region between 0 p.p.m. and 4.4 p.p.m. (i.e., 1,300 variables). Step 2: PCA transforms the original set of variables to a new set of orthogonal variables called principal components (PCs). Step 3: DFA has "a priori" information based on spectral replicates and uses this to minimize within-group variance and maximize between-group variance. (B) DFA plot based on the first eight PC projections from the NMR spectral data. The numbers represent the NMR spectra of extracts of the following strains: (1) FY23.cox5aΔ; (2) FY23.hoΔ; (3) FY23.ρ^0; (4) FY23.pet191Δ; (5) FY23.pfk26Δ; (6) FY23.pfk27Δ.

metabolite concentrations. This discriminatory potential was optimized further by examining which (combinations of) metabolite concentrations were most diagnostic of the differences between different strains and which merely reflected experimental error (differences between experiments with the same strain). Figure 2A illustrates the principal components analysis (PCA) and discriminant function analysis (DFA) methods that were implemented for this purpose. The two most discriminatory combinations of metabolite concentrations are used as the axes in Figure 2B. Mutants with identical phenotypes should cluster in this plot. Mutants with qualitatively different phenotypes should be clearly displaced from each other.

The results shown in Figure 2B demonstrate the power of this approach. Three main clusters are delineated. One contains the two completely respiratory-deficient mutants, FY23.pet191Δ and FY23.ρ^0, whereas a second contains all extracts obtained from the control strain, FY 23.hoΔ. The wild type has previously been shown to be very similar to this[5]. The third cluster groups spectra from FY23.pfk26Δ and FY23.pfk27Δ together and clearly separates them from the control strain's cluster. In terms of the DF1 component, spectra from the partially respiratory-deficient mutant, FY23.cox5aΔ, occupy positions intermediate between those of the control strain cluster and that of the completely respiratory-deficient mutants. The fact that multivariate FANCY has greater resolution than the FANCY employing data on a limited number (six) of specific metabolites (Table 2) becomes clear from the fact that, in terms of the DF2 component, the FY23.cox5aΔ mutant is distinguished from the FY23.pet191Δ and FY23.ρ^0 mutants.

The co-response indicators of Table 3 correct for differences due to various *extents* of the same phenotype (i.e., for phenotypes being only quantitatively different). For the multivariate FANCY, this corresponds to focusing on the *direction* in which a mutant is displaced from the origin in Figure 2B. For the FY23.pfk26Δ and FY23.pfk27Δ, that direction is, indeed, the same (approximately –50°) and quite different from the direction in which the respiratory-deficient mutants were displaced. The analysis of these NMR spectra, therefore, has clearly demonstrated FANCY's ability to group genes of related biological activity and even to discriminate appropriately between mutations that have qualitatively similar, but quantitatively different, effects on phenotype.

Discussion

The use of metabolome data in the systematic analysis of gene function has the twin advantages that metabolites are functional cellular entities that vary with the physiological context and also (for many organisms, including yeast) that the number of metabolites is far fewer than the number of genes or gene products. However, in contrast to the case for transcripts and proteins, there is no direct relationship between metabolites and genes. For this reason, metabolomics (more than any of the other levels of analysis in functional genomics) requires that we exploit our knowledge of experimentally characterized genes in the elucidation of the function of unstudied genes. This may be achieved by comparing the change in the cell's metabolite profile that is produced by deleting a gene of unknown function with a library of such profiles generated by individually deleting genes of known function.

The FANCY approach to the elucidation of gene function has a firm theoretical basis in metabolic control analysis (MCA; see ref. 7). The central device of MCA is the control coefficient[34–39]; this measures the fractional change in either the flux through a pathway, or the concentration of some metabolite, relative to the fractional change in the activity of some effector (usually, but not necessarily, an enzyme). Flux control coefficients (FCCs) add up to 1 (refs 34, 39) 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 completely irrelevant to the pathway, it will have an FCC of 0. In practice, most enzymes have FCC values closer to 0 than 1. In contrast, concentration control coefficients (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 earlier comments and ref. 7). Thus, in seeking to elucidate the functions of unstudied genes through metabolomics, we must have either very sensitive measures of flux (because FCC values are on average close to 0), or comprehensive measures of metabolite concentrations (because CCC values may be large but, if a gene is of unknown function, we do not know which metabolite concentrations to measure).

In a top-down approach to functional analysis via the metabolome, growth rate competition experiments provide a sensitive means of measuring changes in flux[9,10,40]. Competition experiments in chemostat culture failed to reveal any growth rate differences between the *pfk26* and *pfk27* deletants and their wild-type parent (Fig. 1). Thus both mutants had completely silent phenotypes. However, by comparing the glycolytic metabolite profiles in these mutants with those of the wild-type strain and of two respiratory mutants, we detected a "phenotype" at the level of metabolite concentrations. Furthermore, by quantifying the relative change of the intracellular metabolites in wild-type and mutant strains, we were able to associate the glycolytic and respiratory mutants in pairs, distinguishable from one another and from the wild type. Thus, had *PFK27* been a gene of unknown function, we would have identified the part of the metabolic network on which its gene product impacts, that is, on the same part as the gene product of *PFK26*. The present approach, relying as it does solely on steady-state metabolic snapshots, consequently differs noticeably from those requiring a significantly larger number of time-dependent data points (e.g., ref. 41).

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 co-response indicator profiles of known and

Table 2. Internal metabolite concentrations[a]

Strain	Growth rate (h^{-1})	G6P	F6P	ATP	Pyr	ADP	AMP	ATP/ ADP
Wild type	0.31	2.05 ±0.11	0.40 ±0.03	2.80 ±0.32	3.39 ±0.54	0.42 ±0.11	0.20 ±0.02	6.67
hoΔ	0.31	2.22 ±0.10	0.39 ±0.03	2.71 ±0.07	4.35 ±0.40	0.43 ±0.06	0.20 ±0.03	6.30
pfk26Δ	0.30	2.24 ±0.06	0.57 ±0.04	2.45 ±0.13	5.45 ±1.01	0.71±0.10	0.17 ±0.00	3.45
pfk27Δ	0.30	2.71 ±0.02	0.67 ±0.02	2.35 ±0.02	4.56 ±0.38	0.67 ±0.07	0.15 ±0.02	3.51
cox5aΔ	0.30	1.81 ±0.08	0.53 ±0.09	1.70 ±0.17	5.51 ±1.91	0.76 ±0.14	0.16 ±0.01	2.24
pet191Δ	0.31	1.88 ±0.11	0.46 ±0.07	1.75 ±0.06	5.97 ±1.04	0.84 ±0.09	0.15 ±0.03	2.08

[a]Given in mM, assuming that 3.75 ml cytosol is equivalent to 1 g of total protein. Each metabolite concentration is the average of the values obtained from three independently grown cultures (± s.e.m.).

Table 3. Co-response indicators[a]

Strain	F6P/G6P	Pyr/G6P	ATP/G6P	ADP/G6P	AMP/G6P	[ATP/ADP]/G6P
pfk26Δ	+80°	+80°	–60°	+80°	–60°	–80°
pfk27Δ	+60°	+50°	–30°	+60°	–50°	–70°
cox5aΔ	–70°	–80°	+80°	–80°	–60°	+80°
pet191Δ	–60°	–80°	+80°	–80°	–70°	+90°

[a]The responses, for each of the four mutants, of various metabolite concentrations (relative to the response of G6P) are given as co-response indicators in units of degrees (in between –90° and +90°). These were defined as indicated in the Experimental Protocol, and calculated from the data in Table 2.

unknown genes. Indeed, simply the sign (or quadrant) of the indicators, rather than their absolute values, could be compared (see Table 3). Nevertheless, in the absence of any independent indication of the likely domain of activity of an unstudied gene, this would be a rather ponderous 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 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 by 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. We have provided an example of one such technique, high-resolution NMR spectroscopy (Fig. 2), although MS or LC-MS may be equally useful.

This study shows 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. Moreover, in combination with co-response coefficient profiles, the site of action of the silent mutation in the metabolic network may be revealed. This approach should obviously be useful for the rapid identification of new and interesting silent regulatory genes in various parts of the metabolism of various organisms. However, the FANCY approach 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, in contrast, may affect the concentration of all amino acids. Having demonstrated the feasibility and analytical power of the FANCY approach on a difficult test system, we are now exploiting the method in a high-throughput screen of single-ORF deletion mutants of yeast[5,42].

Experimental protocol

Yeast strains. The yeast strains used in this work are all direct derivatives of strain FY23 (Table 1). FY23.ρ^0, a mtDNA^0 cytoplasmic petite mutant was generated by ethidium bromide mutagenesis[43]. All deletion derivatives were generated by PCR-mediated gene replacement using the *kan*MX cassette[26]. Strains FY23.pfk26Δ and FY23.pfk27Δ are deleted for the *PFK26* and *PFK27* genes, respectively. Strains FY23.pet191Δ and FY23.cox5aΔ are nuclear petites exhibiting no and 15% respiratory activity, respectively[28]. Finally, strain FY23.hoΔ, which is deleted for the *HO* gene, was used as a control, because this deletion has no measurable phenotypic effect[25].

Batch cultures. Yeast cells were pregrown at 30°C in 20 ml 2% (wt/vol) glucose, 0.5% (wt/vol) ammonium sulfate, 0.17% (wt/vol) yeast nitrogen base without amino acids (Difco, Detroit, MI) and 100 mM potassium phthalate at pH 5.0, supplemented with required nutrients (40 mg/L uracil, 40 mg/L L-tryptophan, 60 mg/L L-leucine). The next day, the cells were diluted in 100 ml fresh medium and grown to an OD_{600} of 1.

Competition experiments in chemostat culture. Competitions between FY23.pfk26Δ and FY23 pfk27Δ and their parent strain, FY23, were carried out in aerobic and anaerobic, glucose-limited chemostat cultures, exactly as described previously[9].

Metabolite extraction and analysis. A 6 ml sample of culture was injected into 24 ml of 60% (vol/vol) methanol at –40°C. The pellet was washed, spun down at –20°C, and extracted with buffered boiling ethanol[44]. Intracellular metabolite concentrations in the resulting cell-free extract were determined enzymatically[45] using a Cobas Bio Autoanalyser (Roche, Basel, Switzerland). The relative response of two metabolites to an infinitesimal modulation of any system parameter can be expressed in terms of the co-response coefficient, Ω (ref. 11). Here, we consider the analog for the finite modulation due to a mutation:

$$o_{mutation}^{X:G6P} \equiv \frac{\ln(X_M) - \ln(X_W)}{\ln(G6P_M) - \ln(G6P_W)} \approx \Omega_{mutation}^{X:G6P} \qquad (1)$$

X refers to the concentration of any metabolite, *G6P* to the concentration of glucose-6-phosphate, and subscripts *M* and *W* to mutant and wild-type cells, respectively. This co-response coefficient has the disadvantage that it becomes disproprtionately large if the mutation has little effect on the G6P concentration. We therefore discuss the co-response in terms of the co-response indicator, θ (ref. 46), which we define as the arctangent of the co-response coefficient:

$$\theta_{mutation}^{X:G6P} = \arctan(o_{mutation}^{X:G6P}) \qquad (2)$$

This corresponds to the direction of the co-response in terms of an angle, θ, between –90° and +90°. If θ = +45°, *X* and *G6P* respond identically and positively to the mutation. If θ = –45°, *X* and *G6P* respond equally but in opposite directions, etc.

For comprehensive analysis of cellular metabolites, NMR spectrometry was used. An aliquot (20 ml) of a mid-exponential phase culture was transferred into a 50 ml Falcon tube, containing 20 ml of HPLC-grade methanol at –40 to –50°C in an ethanol/dry ice bath. Extracts were made by adding 2 ml of hot ethanol to the pellet after a cold spin (see above) and, after temporary storage at –80°C, were dried under vacuum. The dried extracts were dissolved in 0.6 ml of 0.1 M potassium phosphate buffer, pH 7.0, in D_2O, containing 1 mM 3 (trimethylsilyl) tetradeutero sodium propionate (TSP), quickly vortexed, and then centrifuged at 14,000 r.p.m. for 10 min to remove material that had not dissolved. ^{1}H spectra were acquired at 400 MHz into 40,000 data points using a 90 s pulse, an acquisition time of 4 s, and a sweep width of 5 KHz. The overall pulse repetition time was 5 s. The samples were spun at 16 Hz and maintained at 30°C during data acquisition. The spectra were the sum of 128 transients. The summed free induction decays (FIDs) were multiplied by an exponential 1 Hz line-broadening before Fourier transformation into 8,000 data points. Peak intensities were output from an 8.84 p.p.m. window, which started at –0.52 p.p.m. and contained 2,897 data points. The chemical shift scale was referenced to the signal from TSP at 0.0 p.p.m.

Chemometrics. The region around the most prominent NMR peak due to water (4.4–5.5 p.p.m.) was removed and the internal standard used to normalize each ordinate (thus allowing quantitative comparison of spectra) before it, too, was removed from the spectrum. Finally, it was known that the region beyond 5.5 p.p.m. (the aromatic region) is of little interest to this study, so this was removed as well. The resulting reduced data set describes the subspectral region between 0 p.p.m. and 4.4 p.p.m. (i.e. 1,300 variables). Each column of the data set was then normalized to unit variance. Following this preprocessing, a multistage projection method[47] was employed to cluster similar NMR spectra.

The aim of the multivariate analysis was to observe the natural groupings of the strains in terms of their NMR spectra, and to establish the extent to which the patterns so revealed are concordant with the physiological properties of the strains. The multistage projection method detailed in Figure 2 is a robust and reliable way of analyzing spectroscopic data[48]. Before employing

any multivariate analysis, each column of the data set is normalized to unit variance. This is done to eliminate bias, in subsequent analysis, toward any column that contains either large absolute values or large variances[49].

The first stage of this analysis involved the reduction of the dimensionality of the NMR data by PCA (refs 50,51), a method to redefine each spectrum as a linear combination of several subspectra. The spectrum for each strain analyzed is divided into a similar set of subspectra. Thus, the original data are represented as a series of vectors in *n*-dimensional space, each of which is unique for a particular strain.

The influence of each of the original variables on the new principal components (PC) is determined on the basis of the maximum variance criterion. Thus, the first PC is considered to lie in the direction describing maximum variance in the original data. Each subsequent PC lies in an orthogonal direction of maximum variance that has not been considered by the former components. The number of PCs computed for a given data set is up to the analyst. However, usually as many PCs are calculated as are needed to explain a preset percentage of the total variance in the original data (the number of PCs is always less than or equal to the number of original variables).

The second stage of the data analysis involved using DFA (refs 52,53) to separate the samples further into groups of replicates using the PCs calculated in stage 1 as the source data. DFA is a supervised projection method (whereas PCA is unsupervised). A priori information (as to which samples are replicates of one another) is used to produce measures of within-group variation and between-group variation. This information is then used to define discriminant functions that optimally separate the a priori groups; that is, a plane is defined within *n*-dimensional space onto which all of the vectors are projected. This plane is defined such that it gives the maximum separation between the different groups, and so gives a diagram in which data from similar strains are clustered together and well separated from those produced by strains having different metabolic characteristics.

Acknowledgments

This work was supported by EC contracts, within the frame of the EUROFAN program, to S.G.O., K.v.D., and H.V.W., and by a grant from the UK's Biotechnology and Biological Sciences Research Council to S.G.O. and D.B.K. We would like to thank Cathy Day for her superb technical assistance, and Barbara Bakker and Johann Rohwer for stimulating discussions.

1. Hieter, P. & Boguski, M. Functional genomics: It's all how you read it. *Science* **278**, 601–602 (1997).
2. Brent R. Genomic biology. *Cell* **10**, 169–183 (2000).
3. Oliver, S.G. From DNA sequence to biological function. *Nature* **379**, 597–600 (1996).
4. Oliver, S. Guilt-by-association goes global. *Nature* **403**, 601–603 (2000).
5. Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378 (1998).
6. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–7 (1996).
7. Teusink, B., Baganz, F., Westerhoff, H.V. & Oliver, S.G. Metabolic control analysis as a tool in the elucidation of the function of novel genes. *Methods Microbiol.* **26**, 297–336 (1998).
8. Smith, V., Chou, K.N., Lashkari, D., Botstein, D. & Brown, P.O. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274**, 2069–2074 (1996).
9. Baganz F. *et al.* Quantitative analysis of yeast gene function using competition experiments in continuous culture. *Yeast* **14**, 1417–1427 (1998).
10. Giaever, G. *et al.* Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* **21**, 278–283 (1999).
11. Hofmeyr, J.H., Cornish-Bowden, A. & Rohwer, J.M. Taking enzyme kinetics out of control; putting control into regulation. *Eur. J. Biochem.* **212**, 833–837 (1993).
12. Hofmeyr, J.H. & Cornish-Bowden, A. Co-response analysis: a new experimental strategy for metabolic control analysis. *J. Theoret. Biol.* **182**, 371–380 (1996).
13. Kholodenko, B.N., Schuster, S., Rohwer, J.M., Cascante, M. & Westerhoff, H.V. Composite control of cell function: metabolic pathways behaving as single control units. *FEBS Lett.* **368**, 1–4 (1995).
14. Rohwer, J.M. Interaction of functional units in metabolism (Ph.D. Thesis). (University of Amsterdam, Amsterdam; 1997).
15. Rohwer, J.M., Schuster, S. & Westerhoff, H.V. How to recognize monofunctional units in a metabolic system. *J. Theoret. Biol.* **179**, 213–228 (1996).
16. Oliver, S.G. Yeast as a navigational aid in genome analysis. *Microbiology* **143**, 1483–1487 (1997).
17. Kell, D.B. & Mendes, P. Snapshots of systems: metabolic control analysis and biotechnology in the post-genomic era. In *Technological and medical implications of metabolic control analysis.* (eds Cornish-Bowden, A. & Cárdenas, M.L.) 3–25 (Kluwer Academic Publishers, Dordrecht; 2000).
18. Kretschmer, M. & Fraenkel, D.G. Yeast 6-phosphofructo-2-kinase: sequence and mutant. *Biochemistry* **30**, 10663–10672 (1991).
19. Kretschmer, M., Tempst, P. & Fraenkel, D.G. Identification and cloning of yeast phosphofructokinase 2. *Eur. J. Biochem.* **197**, 367–372 (1991).
20. Paravicini, G. & Kretschmer, M. The yeast FBP26 gene codes for a fructose-2,6-bisphosphatase. *Biochemistry* **31**, 7126–7133 (1992).
21. Boles, E., Goehlmann, W.H. & Zimmermann, F.K. Cloning of a second gene encoding 6-phosphofructo-2-kinase in yeast, and its characterization of mutant strains without fructose-2,6-bisphosphate. *Mol. Microbiol.* **20**, 65–76 (1996).
22. Rousseau, G.G. & Hue, L. Mammalian 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase: a bifunctional enzyme that controls glycolysis. *Prog. Nucleic Acids Res. Mol. Biol.* **45**, 99–127 (1993).
23. Van Schaftingen, E. Fructose 2,6-bisphosphate. *Adv. Enzymol. Rel. Areas Mol. Biol.* **59**, 315–395 (1987).
24. Van Schaftingen, E., Lederer, B., Bartrons, R. & Hers, H.G. A kinetic study of pyrophosphate: fructose-6-phosphate phosphotransferase from potato tubers. Application to a microassay of fructose 2,6-bisphosphate. *Eur. J. Biochem*. **129**, 191–195 (1982).
25. Baganz, F., Hayes, A., Marren, D., Gardner, D.C.J. & Oliver, S.G. Suitability of replacement markers for functional analysis studies in *Saccharomyces cerevisiae*. *Yeast* **13**, 1563–1573 (1997).
26. Wach, A, Brachat, A, Pöhlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808 (1994).
27. Yocum, R. Genetic engineering of industrial yeasts. *Proc. Bio Expo* **86**, 17 (1986).
28. Hutter, A. & Oliver, S.G. Ethanol production using nuclear petite yeast mutants. *Appl. Microbiol. Biotechnol*. **49**, 511–516 (1998).
29. Griffiths, P.R. & de Haseth, J.A. *Fourier transform infrared spectrometry*. (Wiley, New York; 1986).
30. Winson, M.K. *et al.* Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC). A hyperspectral FT-IR-based approach to rapid screening for metabolite overproduction. *Analyt. Chim. Acta* **348**, 273–282 (1997).
31. Cole, R.B. *Electrospray ionization mass spectrometry: fundamentals, instrumentation and applications.* (Wiley, New York; 1997)
32. Gaskell, S.J. Electrospray: principles and practice. *J. Mass Spec.* **32**, 677–688 (1997).
33. Lindon, J.C., Nicholson, J.K. & Wilson, I.D. Direct coupling of chromatographic separations to NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **29**, 1–49 (1996).
34. Kacser, H. & Burns, J.A. The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104 (1973).
35. Burns, J.A. *et al.* Control analysis of metabolic systems. *Trends Biochem. Sci*. **10**, 16 (1985).
36. Kell, D.B. & Westerhoff, H.V. Metabolic control theory—its role in microbiology and biotechnology. *FEMS Microbiol. Rev.* **39**, 305–320 (1986).
37. Fell, D.A. *Understanding the control of metabolism.* (Portland Press, London; 1996)
38. Cornish-Bowden, A. & Cardenas, M.L. From genome to cellular phenotype—a role for metabolic flux analysis? *Nat. Biotechnol.* **18**, 267–268 (2000).
39. Heinrich, R. & Rapoport, T.A. Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. *Acta Biol. Med. Ger.* **31**, 479–494 (1973).
40. Oliver, S.G. Redundancy reveals drugs in action. *Nat. Genet*. **21**, 245–246 (1999).
41. Arkin, A., Shen, P.D. & Ross, J. A test case of correlation metric construction of a reaction pathway from measurements. *Science* **277**, 1275–1279 (1997).
42. Winzeler, E. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
43. Slonimski, P.P., Perrodin, G. & Croft, J.H. Ethidium bromide induced mutation of yeast mitochondria: complete transformation of cells into respiratory deficient non-chromosomal petites. *Biochem. Biophys. Res. Commun.* **30**, 232–239 (1968).
44. Gonzalez, B., Francois, J. & Renaud, M. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* **13**, 1347–1355 (1997).
45. Bergmeyer, H.U. *Methods of enzymatic analysis*. (Verlag Chemiel, Basel; 1974)
46. Cornish-Bowden, A. & Hofmeyr, J.H. Determination of control coefficients in intact metabolic systems. *Biochem. J.* **298**, 367–375 (1994).
47. Goodacre, R. *et al.* On mass spectrometer instrument standardization and interlaboratory calibration transfer using neural networks. *Analyt. Chim. Acta* **348**, 511–532 (1997).
48. Goodacre, R. *et al.* Rapid identification of urinary tract infection bacteria using hyperspectral, whole organism fingerprinting and artificial neural networks. *Microbiology* **144**, 1157–1170 (1998).
49. Martens, H. & Næs, T. *Multivariate calibration.* (Wiley, Chichester; 1989).
50. Causton, D.R. *A biologist's advanced mathematics*. (London, Allen & Unwin; 1987).
51. Jolliffe, I.T. *Principal components analysis.* (Springer-Verlag, New York; 1986)
52. MacFie, H.J.H., Gutteridge, C.S. & Norris, J.R. Use of canonical variates in differentiation of bacteria by pyrolysis gas–liquid chromatography. *J. Gen. Microbiol.* **104**, 67–74 (1978).
53. Windig,W., Haverkamp, J., & Kistemaker, P.G. Interpretation of sets of pyrolysis mass spectra by discriminant-analysis and graphical rotation. *Analyt. Chem.* **55**, 81–88 (1983).
54. Winston, F., Dollard, C. & Ricupero-Hovasse, S.L. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* **11**, 53–55 (1995).