

Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in *Saccharomyces cerevisiae**[§]

Timothy J. Griffin[‡], Steven P. Gygi[§], Trey Ideker[¶], Beate Rist^{||}, Jimmy Eng, Leroy Hood, and Ruedi Aebersold^{**}

Using an integrated genomic and proteomic approach, we have investigated the effects of carbon source perturbation on steady-state gene expression in the yeast *Saccharomyces cerevisiae* growing on either galactose or ethanol. For many genes, significant differences between the abundance ratio of the messenger RNA transcript and the corresponding protein product were observed. Insights into the perturbative effects on genes involved in respiration, energy generation, and protein synthesis were obtained that would not have been apparent from measurements made at either the messenger RNA or protein level alone, illustrating the power of integrating different types of data obtained from the same sample for the comprehensive characterization of biological systems and processes. *Molecular & Cellular Proteomics* 1:323–333, 2002.

The concept of discovery science, best illustrated by the human genome project (1, 2), involves the identification of the components of a system without the prior formulation of hypotheses as to how these components function (3). This scientific method has spawned what has become known as the “systems” approach to biology, which involves the comprehensive characterization of the components of a biological system (*i.e.* DNA, RNA, and proteins) as a whole, leading to insights into the responses of these components because of systematic perturbations to the system. The objective of the systems biology approach is to identify markers and mechanisms that are important to the function of the perturbed system, with the ultimate goal of developing computational models that enable the prediction of the response of the system to any given perturbation.

Traditionally, studies measuring the effects of systematic perturbations have been carried out at the level of transcribed mRNA, most commonly using cDNA arrays and chip technologies (4–7), or alternative methods for mRNA analysis such as serial analysis of gene expression (8), differential display (9), and cDNA fingerprinting (10). These technologies have been

used to distinguish diagnostically between cell types (7, 11–15) and to differentiate between states (metabolic, activation, pathological) of a particular cell type (6, 16), as well as for the comprehensive analysis of cellular pathways and processes by targeted perturbations of cells (16–19).

Although the measurement of transcribed mRNA has proven to be very powerful in the discovery of molecular markers and the elucidation of functional mechanisms, alone it is not sufficient for the characterization of biological systems as a whole. This is based on several observations. First, comparison of absolute mRNA transcript abundances measured by serial analysis of gene expression (8) with the corresponding protein abundances expressed in exponentially growing *Saccharomyces cerevisiae* cells has shown that in many cases mRNA abundance is not a reliable indicator of corresponding protein abundance (20), and studies in other systems have reached similar conclusions (21, 22). Furthermore, attenuation of protein abundance because of post-transcriptional control of protein translation (23) and protein modifications (24) cannot be predicted currently from measurement of mRNA abundance. Neither does the mRNA sequence nor abundance predict accurately the subcellular location of expressed proteins, their associations with each other or and/or other biomolecules, or the mechanisms of protein half-life control.

In contrast to the studies discussed above that have investigated the correlation between absolute amounts of transcribed mRNA and protein, the aim of this study was to investigate the correlation of induced changes in mRNA and protein expression. To this end, we have applied the isotope-coded affinity tag (ICAT¹; Applied Biosystems, Foster City, CA) reagent technology (25) to the quantitative analysis of steady-state protein expression in *S. cerevisiae* grown on the two carbon sources, galactose and ethanol; we have also measured the relative abundance of the corresponding mRNA transcripts by cDNA microarray analysis (4). Our results indicate that for many genes the measurement of mRNA response is not predictive of the protein response. Consequently, these different types of data provide complementary information to the elucidation of mechanisms of control that

From the Institute for Systems Biology, Seattle, Washington 98103
Received, January 8, 2002, and in revised form, April 2, 2002
Published, MCP Papers in Press, April 2, 2002, DOI
10.1074/mcp.M200001-MCP200

¹ The abbreviations used are: ICAT, isotope-coded affinity tag; MS/MS, tandem mass spectrometry; r-proteins, ribosomal proteins.

would not be evident from information obtained at only a single level of gene expression. This validates a central idea to the systems approach to biology that it is only through the integration of different levels of information that the system can be described comprehensively.

EXPERIMENTAL PROCEDURES

Preparation of mRNA and Protein—Proteins and mRNA were harvested from the yeast strain YPH499 (*MATa*, *ura3-52*, *lys2-801*, *ade2-101*, *leu2-1*, *his3-200*, *trp1-63*). mRNA and proteins were isolated from separate yeast cultures grown under the same conditions. Starter cultures were first grown in YP-rich medium containing 2% galactose. Cells from this culture were diluted to 0.05 optical density in either the same galactose-containing medium or YP-rich medium containing 2% ethanol and 0.05% galactose. These cultures were then grown to log phase (2×10^7 cells/ml) at 30 °C. From one set of the cultures total RNA was harvested, the poly(A) RNA was purified, and cDNA was synthesized by reverse-transcription as described previously (26). Protein was harvested from the other set of cultures as described (20).

DNA Microarray Analysis—The synthesized cDNA was hybridized to a DNA microarray containing a set of ~6200 known and predicted gene open reading frames fabricated as described (26). The hybridization fluorescence data obtained were processed automatically using a software algorithm (26).

Quantitative Protein Profiling—Quantitative protein analysis was carried out by the ICAT (Applied Biosystems, Foster City, CA) reagent method (25). 2.2 mg of protein isolated from cells grown on the ethanol carbon source were labeled with the isotopically normal (d(0)) form of the ICAT reagent, and likewise 2.2 mg of protein isolated from cells grown on the galactose carbon source were labeled with the isotopically heavy (d(8)) form of the reagent. The samples were labeled, digested, and purified using a multidimensional chromatographic approach as described previously (27). Ten of the purified, ICAT reagent-labeled peptide fractions were lyophilized and redissolved in reverse phase microcapillary liquid chromatography Buffer A.

Mass Spectrometric Analysis—Fused silica microcapillary columns (100 μm inner diameter \times 12-cm) were in-house packed with Magic C18 (5 μm , 200 Å) spherical silica (Michrom BioResources, Auburn, CA). A flame-pulled tip (5 μm diameter) at one end of the capillaries served a dual purpose of retaining beads and as an electrospray ionization source. The voltage (+1.8 kV) was applied behind the column through a gold wire into one arm of a microcross (Upchurch Scientific, Oak Harbor, WA). The other three arms of the cross were used as a receiver for the liquid chromatography flow (75 $\mu\text{l}/\text{min}$), a fused-silica flow restrictor (50 μm inner diameter \times 50-cm) that passed 74.5 $\mu\text{l}/\text{min}$ to waste, and 500 nl/min (split ratio of 0.007) across the packed capillary column that was connected to the ion source. The capillary column was loaded with ~20% of the total of each affinity-purified peptide mixture offline via a pressure cell and then reconnected to the system. After washing for 5 min with 90% Solvent A (0.4% acetic acid and 0.005% heptafluorobutyric acid in water) and 10% Solvent B (0.4% acetic acid, 0.005% heptafluorobutyric acid in 100% acetonitrile), a binary gradient from 10% Solvent B to 35% Solvent B over 1 h was run using an HP1100 solvent-delivery system (Hewlett Packard, Palo Alto, CA). Eluting peptides were analyzed using an LCQ classic ion-trap mass spectrometer (Finnigan MAT, San Jose, CA). The mass spectrometric strategy employed consisted of the ion trap alternating between mass spectral scans detecting peptide ion mass-to-charge ratios and tandem mass spectrometry (MS/MS) scans in which a selected peptide ion species was subjected to collision-induced dissociation. Each scan lasted an average of ~1.3 s. Therefore over the 1-h analysis time ~1300 sequencing

attempts were carried out. The specific mass-to-charge value of each peptide sequenced by tandem mass spectrometry was excluded dynamically from reanalysis for 1 min (28, 29).

The obtained MS/MS spectra were automatically searched against a data base of predicted proteins derived from the ~6100 open reading frames in the *S. cerevisiae* genome using the SEQUEST algorithm (30). The cleavage specificity for the protease used was not specified for the search, and oxidized methionines and ICAT reagent-labeled cysteines (both the d(0) and d(8) forms) were specified as static modifications in the search parameters. No sequence constraints were included in the data base search to allow for the identification of non-specifically retained, non-cysteine-containing peptides, which were found to constitute less than 10% of the total peptides identified and did not interfere with the analysis. A peptide was considered to be a match if the cross-correlation score for a MS/MS spectrum from a peptide ion was at least 2.0 or the Δ correlation score was at least 0.1 (30). We considered a protein identified if at least two such peptide matches were apparent for the protein. For proteins identified by a single peptide, the veracity of these peptide sequence determinations by SEQUEST was confirmed by manual inspection. Several criteria were used to confirm the sequence matches: (i) Peptide sequences identified as containing either d(0)- or d(8)-labeled cysteine were inspected for characteristic ICAT reagent fragments (peak at m/z of 284 for d(0)-labeled peptides and m/z of 288 for d(8)-labeled peptides); (ii) inspection of spectra for matches between major product ions and the theoretically predicted product ions from the data base-matched peptide; (iii) examination of the chromatographic profiles of the peptides identified as d(0) and d(8) labeled for expected behavior, as the d(8)-labeled peptides consistently elute several seconds before the d(0)-labeled peptides. Only peptides passing all of these criteria were determined to be true sequence matches. Quantification of each identified protein was done by reconstructing the ion-chromatographic trace for the d(0) and d(8) form of each peptide and comparing the peak area for corresponding peptide pairs using XPRESS, a novel quantification software routine that enables visual inspection of reconstructed ion chromatograms for identified peptides (31). The criteria used in determining the accuracy of the quantitative results were as follows: (i) signal-to-noise ratios of at least 10; if one of the labeled peptides (d(0) or d(8)) in a pair was detected at a signal-to-noise ratio less than 10, the abundance was estimated by measuring the noise level over the approximate elution profile of that peptide; (ii) approximately Gaussian-shaped elution profiles; (iii) clear chromatographic separation from coeluting peptides at similar m/z ratios. The final results were summarized using the software tool INTERACT (31). For those proteins from which multiple peptides were identified, the \log_{10} of the abundance ratio for each specific pair of ICAT reagent-labeled peptides were averaged to give an average \log_{10} abundance ratio that was used in the final results.

RESULTS

The correlation between induced changes in mRNA and protein-expression profiles was varied. We investigated the extent to which perturbation-induced changes in gene expression at the protein and mRNA levels were correlated in yeast cells. Cells were grown on the fermentable carbon source, galactose, and the non-fermentable carbon source, ethanol, respectively. Carbon source utilization in yeast has been studied in depth, with being growth on various carbon sources known to cause large changes in gene expression (5, 32). Glucose is the preferred sugar for energy generation in yeast, and a well characterized regulatory network results in

TABLE I
Abundance ratio variations among peptides derived from selected proteins

The number of unique, ICAT reagent-labeled peptides identified by tandem mass spectrometry is given for each protein, along with the average ratio of abundance for the peptides on the galactose carbon source relative to the ethanol carbon source and the standard deviation between these ratios.

Protein	Description	No. of peptides	Gal/Eth
Cdc19	Pyruvate kinase	5	2.7 ± 0.5
Gre3	Aldose reductase	3	1.8 ± 0.2
Cox12	Cytochrome c oxidase	3	0.73 ± 0.09
Ahp1	Alkyl hydroperoxide reductase	3	4.6 ± 0.3
Yef3	Translation elongation factor EF-3A	3	2.1 ± 0.4
Ald6	Cytosolic acetaldehyde dehydrogenase	5	0.87 ± 0.2
Hom2	Aspartate-semialdehyde dehydrogenase	3	0.89 ± 0.1
Psa1	Mannose-1-phosphate guanyltransferase	3	1.4 ± 0.2

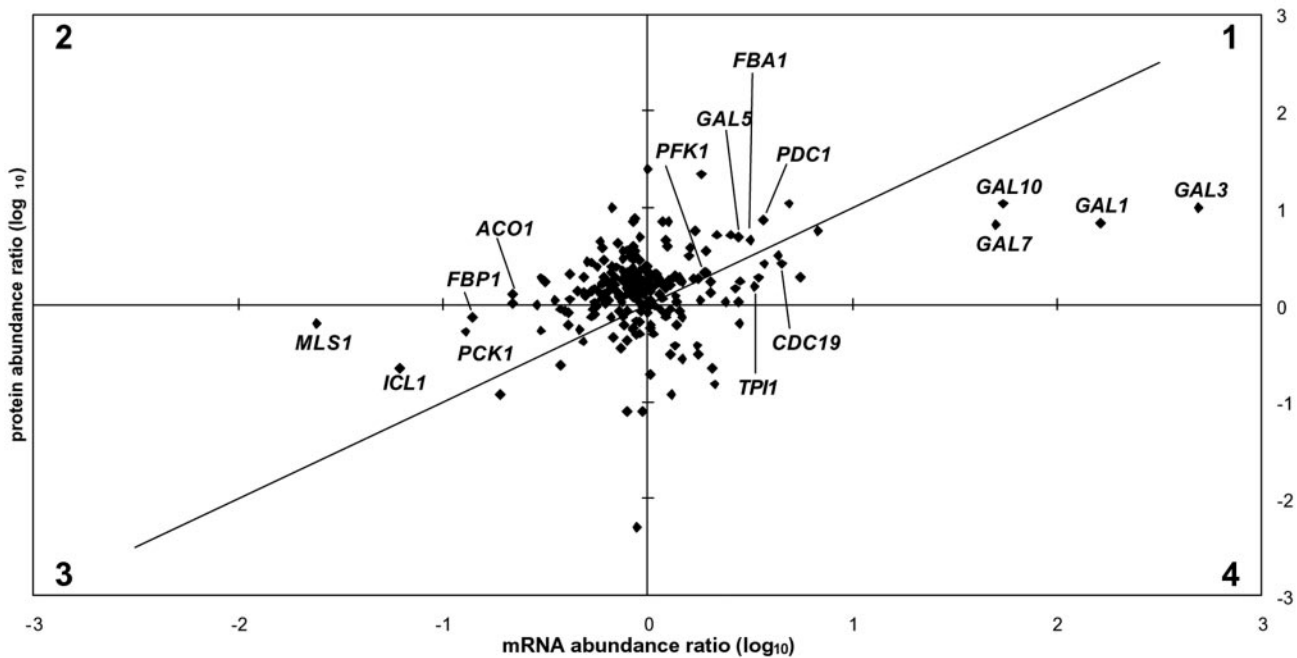
the repression of a plethora of genes responsible for the utilization of other carbon sources when glucose is present (32, 33). Changes in mRNA levels in cells grown on the two carbon sources galactose and ethanol were determined by full genome cDNA arrays (4), and the obtained profiles were subjected to maximum-likelihood analysis (26); changes in the protein profiles were determined by the ICAT reagent method (25). Proteins were isolated from the two cell populations, labeled with the isotopically normal (d(0)) or heavy (d(8)) ICAT reagent, respectively, combined, and proteolyzed. The resulting peptide mixtures were fractionated by strong cation exchange chromatography, and each fraction was subjected to avidin affinity chromatography (27). Ten separate fractions of purified ICAT reagent-labeled peptides were identified and quantified using reverse phase microcapillary liquid chromatography in conjunction with electrospray ionization MS/MS analysis and specifically developed software tools (31). In total, we identified and quantified the relative abundances of 245 unique protein products. We identified and quantified an additional 45 peptides by this method that could not be assigned unambiguously to a specific protein because of sequence similarities between two or more proteins. The quality and accuracy of the quantitative protein data generated initially by the software tool (31) was checked manually, and only those peptides showing satisfactory signal-to-noise ratios in the mass spectrometric analysis were incorporated into the data set presented here. The full set of proteins identified, the abundance ratios for these proteins, and the abundance ratios for each corresponding mRNA transcript are provided in Table I of the Supplemental Material. Fig. 1A shows the abundance ratios of the uniquely identified proteins plotted against the ratios obtained for the products of the same genes obtained at the mRNA level. The plotted *line* indicates the line $y = x$ that would be expected if the mRNA and protein-abundance ratios were correlated perfectly. A non-parametric correlation analysis of the experimental data using the Spearman rank correlation method (34) gives a correlation coefficient of 0.21 ($p < 0.001$). This indicates a positive directional correlation between the mRNA and protein-abundance ratios, although there was a relatively weak

correlation when compared with the perfectly correlated data represented by the plotted *line* in Fig. 1A.

An important aspect of this comparison is an assessment of what constitutes a significant change in abundance at either the mRNA or protein level. For hybridization data obtained from a cDNA microarray, a conservative abundance ratio threshold indicating significant change in abundance in a single experiment is at least a 3-fold change ($\log_{10}(\text{ratio}) \geq 0.48$) (26). As the results presented here were derived from a single microarray experiment, only transcripts showing abundance changes exceeding this threshold were considered to be expressed differentially, to account for possible measurement errors in this data. This threshold has been set based on control experiments comparing equal amounts of differently fluorescently labeled mRNA hybridized to an array and determining the level of variation between the measured ratios and the expected ratio of one. As the level of variation increases for transcripts detected at low fluorescence intensities, mRNA species detected by the analysis software at levels below the threshold where this error model is valid were omitted from the results (26). For protein-abundance ratios measured by the ICAT reagent method we set a 1.5-fold change ($\log_{10}(\text{ratio}) \geq 0.18$) as the threshold indicating significant change. This value was chosen for the following reasons: (i) assessment of variation over numerous experiments. In previous studies using the ICAT reagent technology to measure abundances of control mixtures of proteins we have observed discrepancies of 10–20% between the expected and measured values (25, 27); (ii) assessment of the variation within an experiment. The confidence in the protein-abundance ratio measurements within an experiment can be assessed by the variation in abundance ratios measured from multiple peptides derived from the same protein. In the study described here approximately one-third of the proteins quantified had two or more peptides contained in their sequence identified and used to determine the relative abundance ratios. Table I shows a sampling of proteins identified in which multiple peptides from the same protein were matched to the protein and used to determine the abundance ratio of the protein. On average the variation between peptides for the data shown in

A

mRNA versus protein abundance ratios, Gal/Eth



B

Differentially expressed genes, Gal/Eth

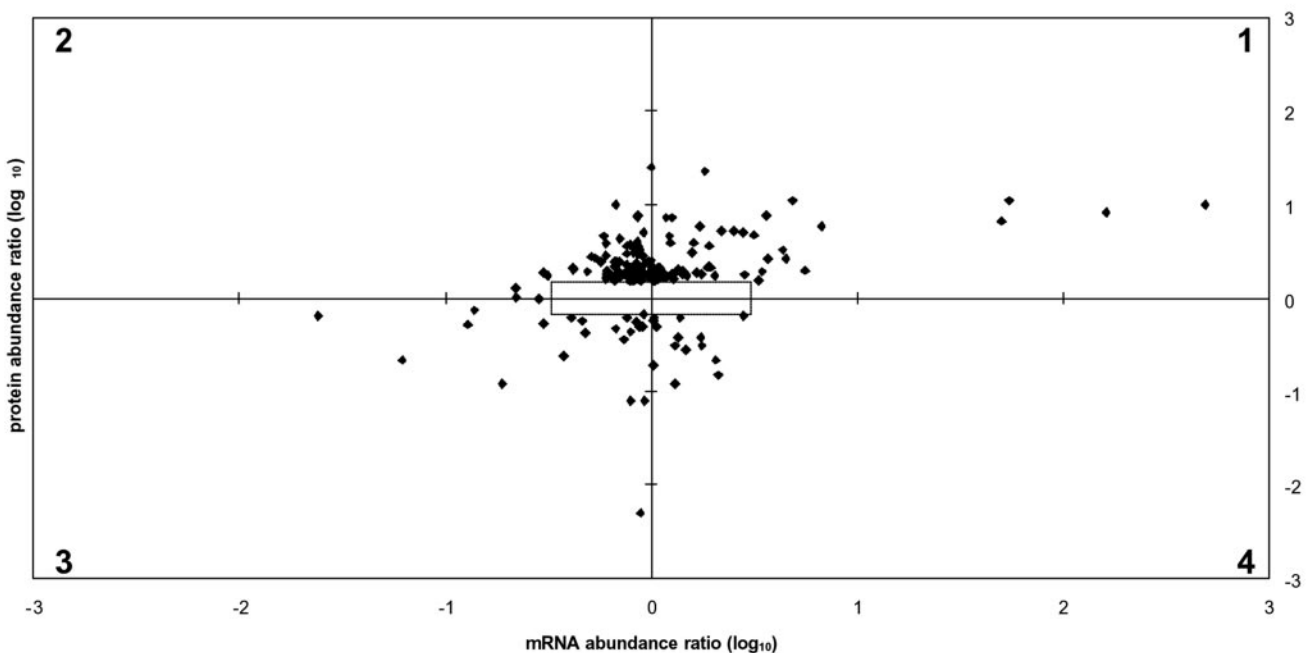


FIG. 1. mRNA abundance ratios versus protein-abundance ratios. The log₁₀ values of the ratio of abundance on the galactose carbon source to the abundance on the ethanol carbon source measured at the mRNA and protein level are plotted against each other for each unique gene product characterized in this study. *A*, all gene products characterized. Some of the key genes involved in carbohydrate metabolism and energy generation are indicated. The plotted *line* indicates data points showing perfect correlation between mRNA and protein abundances. *B*, those data points showing significant differential expression between the two carbon sources.

Table I is ~14% of the average abundance ratio determined for the protein from the peptides, illustrating the high amount of precision and accuracy of the ICAT reagent method for the measurement of relative protein abundances. Similar levels of variation between ICAT reagent-labeled peptides derived from the same protein have been observed in another large-scale study (31); (iii) reproducibility of the experiment. We have reproduced the entire experiment described in this study (*i.e.* cell growth, ICAT reagent labeling of proteins, mass spectrometry) three separate times, giving four independent sets of data including the data set presented here. A sampling of proteins identified in these replicate sets of experiments has shown the ICAT reagent approach to be highly reproducible, with a standard deviation between measurements being on average less than 20% of the average value determined for each of the proteins compared (35). It should be noted that in some cases an identified peptide might exhibit large discrepancies in the measured abundance ratio when compared between separate reproducibility experiments or when compared with data from other peptides derived from the same protein within the same experiment. The vast majority of these discrepancies occur in the case of peptides that are detected at low signal-to-noise ratios and/or overlap in mass with other coeluting peptides; both of these circumstances severely confound the quantitative measurements. However, the XPRESS software tool (31) for quantifying ICAT reagent-labeled peptides facilitates the inspection of the quantitative results for identified peptides, and as such those peptides showing poor data quality can be determined and omitted from the final results. Accordingly, the data quality of all the identified peptides presented in this study, and the studies described above that assessed the variation of the ICAT reagent method, was checked rigorously using the criteria described under “Experimental Procedures,” omitting those peptides showing poor data quality to ensure to the greatest extent possible the accuracy of the data. Data that passes these criteria show an average variation of 20% or less as discussed above, with outliers showing as much as 40% variation and as little as 1% relative to the average for a given protein. This variation is most likely because of matrix effects that cause slight variability in peptide ionization efficiency and instrument detector response. Given the level of variation observed for this data, a 1.5-fold abundance change is a reasonable threshold value for the protein measurements. Fig. 1B shows the data points that exhibited differential expression. The *rectangle* indicated by the *dotted line* shows the region of the plot for data points that did not exhibit significant changes in abundance at either the mRNA or protein level using the criteria set above. In all, 166 of the 245 unique data points showed differential expression when measured at either the mRNA or protein level or both.

In an earlier study we used similar metabolic perturbation in yeast to demonstrate the ability of the ICAT reagent method to identify and quantify proteins in complex mixtures (25). That

study differs from the present one in the growth conditions and the time period between the induction of the perturbation and sample harvest and the sample-processing protocol. In the initial experiment yeast were first cultured in rich medium containing galactose, and a portion of these cells were then transferred to medium containing 2% ethanol. In the present study, the cultured yeast were transferred to medium containing 2% ethanol and 0.05% galactose. The small amount of galactose was added to initiate growth more rapidly on the ethanol, as in the presence of ethanol alone initiation of cell growth may take several days or more, as was the case in the original experiment describing the ICAT reagent method. Therefore the original experiment more closely reflects changes because of separate growth on ethanol or galactose, whereas the present results essentially reflect expression changes because of a carbon source shift from galactose to ethanol. This is reflected in the observation that the ratio of induction of some of the proteins including for Gal1p and Gal10p, which were measured to be 100-fold or greater in the initial experiment (25), are lower in the present experiment. The ability to detect changes in abundance ratio by a factor of 100 or more demonstrated in the earlier study indicates that the decreased magnitude of these ratios relative to the mRNA measurements observed in the present experiment is not because of a lack of dynamic range in the mass spectrometric measurement of these values, but rather it is most likely a biological effect. The complex mixture of labeled peptides generated in the earlier study was separated by one-dimensional (reverse phase) chromatography, whereas the samples generated in this experiment were separated by two-dimensional (cation exchange/reverse phase) chromatography (27). This increased chromatographic space resulted in an increase in the number of uniquely identified and quantified proteins from 34 to 245.

DISCUSSION

It is clear from inspection of the data derived from these 245 genes in Fig. 1B that there is a significant number of genes that show large discrepancies between abundance ratios when measured at the levels of mRNA and protein expression. Particularly evident is the clustering of genes in *quadrant 1* and *2* of Fig. 1. These genes are showing a significant increase in relative protein expression on the galactose carbon source but no significant change in abundance when measured at the mRNA level. To better explain the data presented here, we clustered the genes characterized in this study by their known cellular roles (36) (see Table I of Supplementary Material), and below we discuss the behavior of those genes involved in carbohydrate metabolism, respiration, energy generation, and protein synthesis.

Carbohydrate Metabolism Pathway Genes—Fig. 2 shows the mRNA and protein-expression ratios of genes with known roles in carbon source metabolism in yeast overlaid onto schematic representations of galactose utilization and energy metabolism in yeast cells (5, 32). All five essential proteins

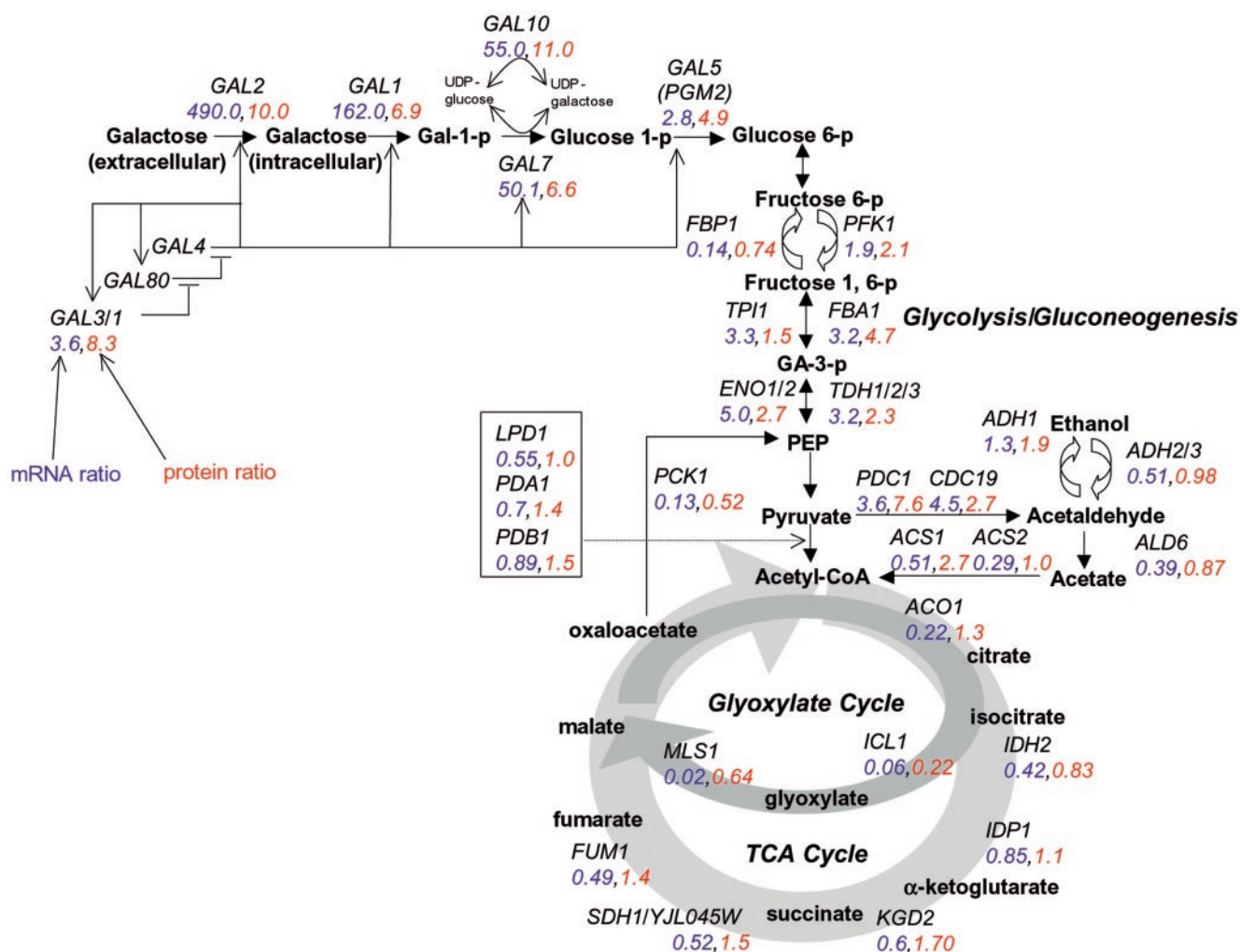


FIG. 2. Schematic of the major carbohydrate metabolism and respiratory genes characterized in this study. For each gene, the measured mRNA abundance ratio is given first, followed by the protein-abundance ratio. For some highly homologous proteins (Gal3/1p, Eno1/2p, Tdh1/2/3p, Adh2/3p, Sdh1p/YJL045W) the peptide sequence identified could not be assigned unambiguously to a single protein. For these genes, the mRNA abundance ratio given is an average value of transcripts corresponding to each of the proteins indicated.

involved in the conversion of galactose to glucose 6-phosphate (Gal1p, Gal2p, Gal5p, Gal7p, and Gal10p) (32) were identified uniquely and quantified; all are repressed in the presence of glucose and induced by galactose (36). An additional protein, Gal3p, could not be characterized unambiguously, as the peptide sequence identified from this protein is also contained in the Gal1p enzyme. In general, the mRNA abundances of the galactose utilization genes showed a larger magnitude of induction on the galactose than the corresponding abundances measured at the protein level. For example, *GAL2* showed an ~500-fold increase in mRNA expression on galactose, whereas it showed an ~10-fold increase when measured at the protein level. Also shown in Fig. 2 are many of the key glycolytic proteins. These genes showed increased expression of similar magnitude on the galactose carbon source at both the mRNA and protein levels, consistent with the metabolic conversion of galactose to glu-

ucose 6-phosphate, which enters into the glycolytic pathway (32). The genes involved in the respiratory conversion of pyruvate to acetyl-CoA, including the E3 component of pyruvate dehydrogenase complex (*LPD1*), the pyruvate dehydrogenase complex E1- α subunit (*PDA1*), and E1- β subunit (*PDB1*), showed no significant change in abundance between the two carbon sources at the mRNA level, while showing small but significant changes at the protein level. In general, although the magnitude of change in abundance may have differed in some cases, the induced changes in expression showed similar patterns when measured at the mRNA and protein level for the major carbohydrate metabolism genes.

Respiratory Genes—The genes involved in the respiratory metabolism of ethanol and their quantitative change at the mRNA and protein level are also shown in Fig. 2. The first step in the respiratory pathway is the conversion of ethanol to acetaldehyde, mediated by the alcohol dehydrogenase 2 en-

TABLE II
mRNA and protein abundance ratios for mitochondrial-located proteins

Gene name	Description ^a	mRNA ^b	Protein ^b
<i>TRX3</i>	Mitochondrial thioredoxin	0.78	1.7
<i>SOD2</i>	Manganese superoxide dismutase	0.74	1.8
<i>GUT2</i>	Glycerol-3-phosphate dehydrogenase	0.53	2.7
<i>HSP60</i>	Mitochondrial chaperonin	0.58	1.4
<i>PET9</i>	ADP/ATP carrier protein	0.42	2.0
<i>ATM1</i>	Membrane transporter	0.85	3.0
<i>MIR1</i>	Phosphate transporter	0.60	1.9
<i>POR1</i>	Outer mitochondrial membrane porin	1.0	1.5
<i>TUF1</i>	Translation elongation factor	0.69	1.9
<i>YLR168C</i>	Putative intramitochondrial sorting protein	0.85	4.0
<i>TOM40</i>	Outer membrane import	0.76	3.5

^a Descriptions were obtained from the Yeast Proteome Database (www.proteome.com).

^b Ratio = abundance on galactose/abundance on ethanol.

zyme (Adh2p) (32). Unfortunately, the peptide sequences identified as being derived from this protein by way of MS/MS analysis are contained in both the Adh2p and the closely related alcohol dehydrogenase 3 isozyme (Adh3p) (37), which catalyzes the reverse reaction, reducing acetaldehyde to ethanol. Therefore the ratio of abundance (which showed no change between the two carbon sources) of the peptides identified could not be assigned unambiguously to either of these enzymes. At the mRNA level, the transcripts from these two genes showed identical ratios (~2-fold increased abundance on the ethanol carbon source), indicating that there is most likely crossover hybridization of these two highly similar transcript sequences (5, 32). Similarly, the peptide sequences determined from the next enzyme in the ethanol utilization pathway, aldehyde dehydrogenase protein (Aldp), could not be differentiated between the cytosolic form (Ald2p) and its putative isoform (Ald3p) (38). The acetaldehyde oxidizing protein (Ald6p) was identified unambiguously. Interestingly it showed no change at the protein level but did exhibit a 3-fold increase in abundance on the ethanol carbon source at the mRNA level. The discrepancies between abundance ratios at the mRNA and protein level of this gene product may indicate a post-transcriptional control mechanism and is consistent with the previously reported finding that this enzyme is also active during fermentative metabolism (39). The acetyl-CoA synthetase protein (Acs1p), showed an ~2.8-fold increase in protein abundance on the galactose carbon source, with no significant change in mRNA abundance. This finding is consistent with an earlier finding that suggests Acs1p is repressed by both glucose and ethanol (40).

Energy-generation Genes—The key regulatory genes involved in both the tricarboxylate cycle and the glyoxylate cycle that were found in this study showed an interesting result. The genes involved in the glyoxylate cycle showed strong increases in transcriptional response on the ethanol carbon source, whereas those genes involved in the tricarboxylate cycle showed no significant amount of change (Fig. 2). At the protein level, we observed a more distinct branching

in the expression pattern between the glyoxylate cycle and the tricarboxylate cycle. Proteins involved in the glyoxylate cycle, specifically isocitrate lyase 1 (Icl1p), malate synthase 1 (Mls1p), and a peptide sequence common to both Mls1p and the nearly identical malate synthase 2 (Dal7p) (36) all showed significantly increased relative abundance on the ethanol carbon source. In contrast, those proteins involved in the other steps of the tricarboxylate cycle (notably Kgd2p and peptide common to Sdh1p and the related protein derived from open reading frame YML045W) showed a slight but significant increase in relative abundance on the galactose carbon source even though the corresponding mRNA levels showed no significant changes. The branch point of these two cycles occurs at the protein Idh1p (41), which showed no significant change in abundance on either the mRNA or protein levels. This is consistent with recently published evidence that Idh1p is regulated allosterically by mitochondrial mRNA species (42), and therefore no change in abundance of the protein on either carbon source may be expected. The allosteric inactivation of Idh1p channels the metabolic flow to its competitor Icl1p and therefore into the glyoxylate cycle in what is known as the glyoxylate bypass (41). The glyoxylate bypass functions as an alternate mechanism of energy generation and synthesis of biological precursors under conditions in which available carbon is scarce (41), thus it should be expected that this pathway would be utilized preferentially in the presence of ethanol but not in the presence of the relatively carbon-rich hexose sugar, galactose.

The steps involved in the tricarboxylate cycle are carried out in the mitochondria, whereas those proteins involved in the glyoxylate cycle are cytosolic (36). Closer examination of the mitochondrial-located proteins characterized here show that these consistently exhibited significantly increased relative protein abundances on galactose, with no significant change at the mRNA level (Table II). This very clearly indicates the preferential channeling of energy generation through the tricarboxylate cycle during growth on galactose initially indicated by the slightly increased expression of the proteins Kgd2p and Sdh1p in Fig. 2. The discrepancies between the

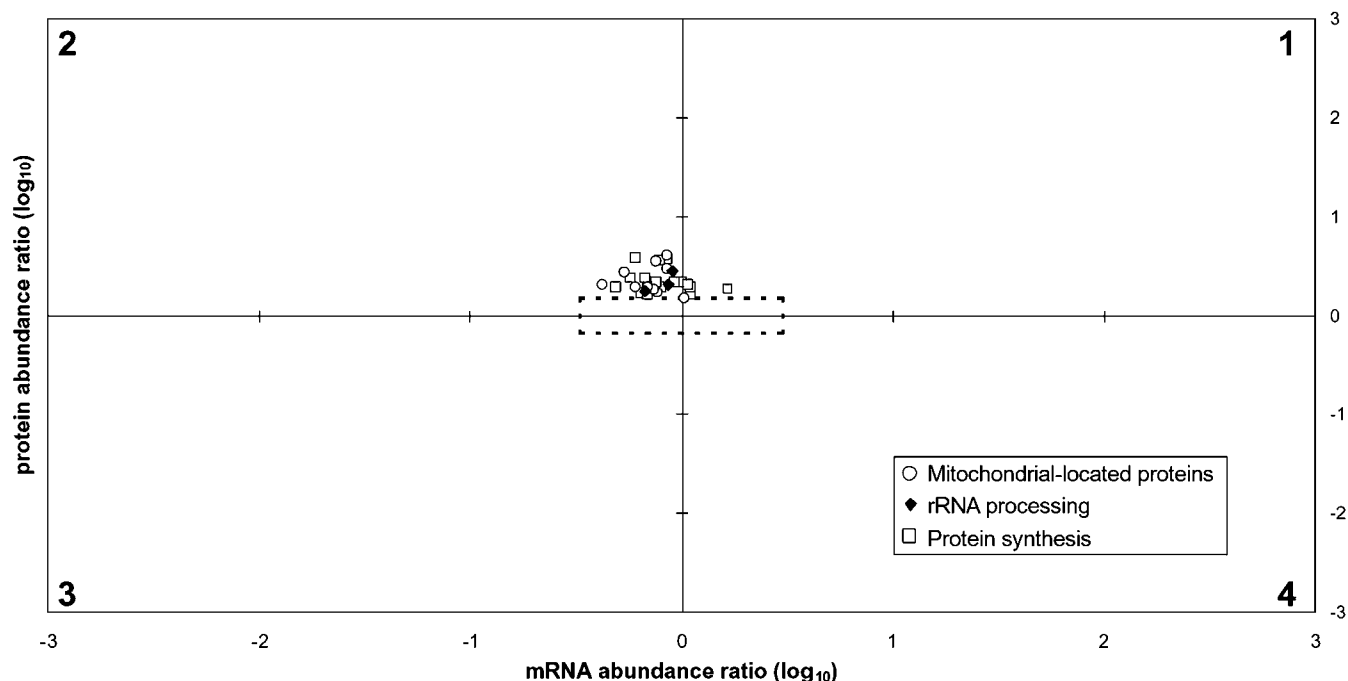


FIG. 3. Data points for proteins that function in the mitochondria, protein-synthesis genes, and rRNA-processing proteins.

mRNA and protein-abundance ratios for these genes indicate a global post-transcriptional control mechanism of these proteins. One possible mechanism of this control may be increased degradation of mitochondrial-located proteins (43) during growth on ethanol, which has been shown to be carbon source-dependent (44). Consistent with this model, the mitochondrial protease Prd1p (36) showed almost a 3-fold increase in abundance on the ethanol carbon source, indicative of increased degradation activity in the mitochondria.

The analysis of the genes involved in respiration and energy generation, taken together with those genes involved in carbohydrate metabolism, indicate three patterns emerging from the comparison of mRNA and protein-abundance ratios: (i) genes that show close agreement between abundance ratios (e.g. glycolysis genes); (ii) genes that show the same general changes in abundance but differ in the magnitude of this change (e.g. galactose metabolism genes); and (iii) genes that are discordant (e.g. mitochondrial-located genes), which may suggest new regulatory mechanisms that are not apparent from either the mRNA or protein data alone.

Protein-synthesis Genes—Another major component of the cluster of data points located in *quadrant 2* of Fig. 1B are genes encoding ribosomal proteins (r-proteins) and other genes involved in protein synthesis (elongation factors, etc.). As a group these fell into the category of discordant genes, as these genes showed an average increase in relative protein abundance on galactose of greater than 2-fold, with no significant changes in abundance indicated at the mRNA level. Amounts of r-protein are known to be proportional to cellular growth rate (45), thus the increased amount of r-proteins in

the presence of the relatively more favorable carbon source galactose is expected. In the case of the r-proteins, the discrepancies between the mRNA and protein-abundance ratios for these genes would seem to contradict previous reports that r-protein gene expression is strictly and entirely regulated at the level of transcription (46, 47). It has, however, also been shown that the abundance of rRNA and r-proteins are under coordinate control (45) and that this control mechanism attempts to keep a balance between the cellular levels of rRNA and r-proteins. Following this model, transcriptional control of rRNA abundances in response to carbon-source perturbations would lead to a subsequent adjustment of r-protein levels. A discrepancy between abundance ratios of r-proteins and their corresponding mRNA transcripts might be expected, at least during the induction phase of the response. Consistent with this model, the rRNA-processing proteins (36) Cbf5p, Sik1p, and Nop1p show increases in relative protein abundances of 2.9-, 2.0-, and 1.7-fold, respectively, on the galactose carbon source (Fig. 3). It has also been suggested that one mechanism for post-transcriptional regulation of r-proteins in response to changes in rRNA levels may be controlled by degradation of excess ribosomal protein (48, 49). Accordingly, the proteins Pre10p, Rpn8p, and Pup3p, members of the 26 S proteasome complex (50), all showed increased abundance (average of greater than 2-fold increase) on the ethanol carbon source, consistent with increased protein degradation on this carbon source.

Together with the mitochondrial-located proteins involved in energy generation and those involved in protein synthesis, these genes make up a significant portion of the data points

that are clustered in *quadrants 1* and *2* of Fig. 1B. These proteins, along with the rRNA-processing proteins, have been plotted separately in Fig. 3. It is evident that a large proportion of protein-abundance ratios measured in this study showed an increase in expression on the galactose carbon source relative to the ethanol carbon source. Accordingly, the median protein-abundance ratio was 1.5 for the entire data set. The reason for this shift in expression ratios is most likely because of biological response and not an introduced experimental effect, as we have reasoned in the case of the energy generation proteins and also the protein-synthesis genes; the reproducibility of these results between replicates of this experiment further supports this conclusion.

We have presented here an extensive study comparing the effects of carbon source perturbations on gene product abundance at the mRNA and protein level in a eukaryotic system. The results shown here illustrate clearly the complementary nature of the information obtained at these different points along the molecular pathway of gene expression. In many cases the response measured at the mRNA level is in accordance with the response at the protein level, which is illustrated here by the genes involved in glycolysis. In other cases there are significant discrepancies between abundance ratios, most notably for the mitochondrial-located energy generation genes and the protein-synthesis genes. These clusters of genes are of most interest, as the differences between the mRNA and protein data may indicate post-transcriptional mechanisms of regulation. By itself this data is of course not sufficient to explain the exact mechanisms of regulation, but it does provide a direction in which to proceed with more specifically targeted experiments. One such direction is to look for conserved gene sequence features among those genes showing prominent differences between mRNA and protein-abundance ratios. This can be done using a publicly available computational tool, AlignACE (51, 52) (atlas.med.harvard.edu/), which identifies potential Cis-regulatory sequence motifs common to a set of gene sequences. Applying AlignACE to the gene sequences for the protein-synthesis genes and the rRNA-processing genes discussed above, we found a highly conserved sequence between these genes that has been described previously as a ribosomal processing element, as it was found to be conserved specifically between rRNA-processing genes (52). The fact that this sequence element is also conserved in the majority of the protein-synthesis genes characterized in this study indicates a possible target in the gene sequence that may be responsible for the coordinate control of the expression of these gene products. Further experiments are necessary to determine whether this sequence element is responsible for the post-transcriptional regulation of these genes that has been indicated in this study.

Despite significant improvements in methodologies for assaying protein expression, such as the ICAT reagent technology used here, the number of unique protein products (245) characterized in this study is still well below the number of

proteins expected to be expressed in *S. cerevisiae* (53, 54). The ability to amplify very low abundance transcripts (7) provides the mRNA-based approaches an advantage in the ability to globally characterize gene products, even those of very low copy number. However, improvements in proteomic approaches using mass spectrometry are now beginning to enable truly proteome-wide characterization of protein expression, as a study described recently that has identified almost 1500 expressed proteins from a total cell yeast lysate (55). Given these recent improvements, along with the development of more sensitive and accurate mass spectrometric instrumentation (56), the sensitivity of protein-expression analysis by mass spectrometry is approaching that of mRNA-based techniques. Additionally, relative quantification of protein expression is more accurate than measurement of expression ratios at the mRNA level, as problems such as cross-hybridization and sequence-specific effects that can confound hybridization-based mRNA measurements do not limit the accuracy of protein measurements. This study, along with others using the ICAT reagent, has shown the ability of this approach to measure accurately small changes (less than 2-fold) in protein expression (19, 25, 31). In this study, the capability to resolve very small changes in protein abundances allowed for the observation of the slight expression changes of the proteins involved in the tricarboxylate cycle, which led to the larger finding of the increased abundances of the mitochondrial-located energy generation proteins.

Finally, in the quest to assign function to gene sequences, the measurement and comparison of mRNA and protein abundances is only a start to characterizing biological systems. Other mechanisms of protein regulation, such as protein modifications and intermolecular protein interactions, are also crucial to protein function. Therefore, there is a need for the development of more comprehensive methodologies that characterize these modifications and interactions, and recent reports document significant progress toward these goals (57, 58). It is clear that continued advancements in the comprehensive analysis of protein products, in conjunction with the already mature methods in global measurement of mRNA expression, will enable the measurement and characterization of cellular circuitry on a system-wide scale.

Acknowledgment—We thank Tim Galitsky for helpful comments on this work.

* This work was supported in part by a grant from the Merck Genome Research Institute and Grant IR33CA84698 from NCI, National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains Supplemental Material.

‡ Supported in part by a National Institutes of Health post-doctoral genome training fellowship.

§ Supported in part by National Institutes of Health Grant HG0041.

Present address: Dept. of Cell Biology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115.

¶ Supported by a National Institutes of Health pre-doctoral genome training fellowship and an ARC fellowship. Present address: Whitehead Inst. for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142.

|| Present address: BioVisioN GmbH and Co. KG, Hannover D-30625, Germany.

** To whom correspondence should be addressed: Inst. for Systems Biology, 1441 N. 34th St., Seattle, WA 98103. Tel.: 206-732-1204; Fax: 206-732-1255; E-mail: raebersold@systemsbiology.org.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921
2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., et al. (2001) The sequence of the human genome. *Science* **291**, 1304–1351
3. Aebersold, R., Hood, L. E., and Watts, J. D. (2000) Equipping scientists for the new biology. *Nat. Biotechnol.* **18**, 359
4. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470
5. DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686
6. Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., and Davis, R. W. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13057–13062
7. Lockhart, D. J., and Winzler, E. A. (2000) Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836
8. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science* **270**, 484–487
9. Liang, P., and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967–971
10. Ivanova, N. B., and Belyavsky, A. V. (1995) Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting. *Nucleic Acids Res.* **23**, 2954–2958
11. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1997) Gene expression profiles in normal and cancer cells. *Science* **276**, 1268–1272
12. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537
13. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511
14. Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Botstein, D., Brown, P. O., and Weinstein, J. N. (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* **24**, 236–244
15. Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., and Brown, P. O. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24**, 227–235
16. Marton, M. J., DeRisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett, D. E., Jr., Hartwell, L. H., Brown, P. O., and Friend, S. H. (1998) Drug target validation and identification of secondary drug target effects using

- DNA microarrays. *Nat. Med.* **4**, 1293–1301
17. Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728
18. Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., Tyers, M., Boone, C., and Friend, S. H. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880
19. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934
20. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730
21. Anderson, L., and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
22. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I. (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**, 7357–7368
23. McCarthy, J. E. (1998) Posttranscriptional control of gene expression in yeast. *Microbiol. Mol. Biol. Rev.* **62**, 1492–1553
24. Parekh, R. B., and Rohlf, C. (1997) Post-translational modification of proteins and the discovery of new medicine. *Curr. Opin. Biotechnol.* **8**, 718–723
25. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999
26. Ideker, T., Thorsson, V., Siegel, A. F., and Hood, L. E. (2000) Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J. Comput. Biol.* **7**, 805–817
27. Griffin, T. J., Loboda, A., Gygi, S. P., Rist, B., Jilkine, A., Ens, W., Standing, K. G., and Aebersold, R. (2001) Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **73**, 978–986
28. Haynes, P. A., Fripp, N., and Aebersold, R. (1998) Identification of gel-separated proteins by liquid chromatography-electrospray tandem mass spectrometry: comparison of methods and their limitations. *Electrophoresis* **19**, 939–945
29. Shabanowitz, J., Settlage, R. E., Marto, J. A., Christian, R. E., White, F. M., Russo, P. S., Martin, S. E., and Hunt, D. F. (2000) in *Mass Spectrometry in Biology and Medicine* (Burlingame, A. L., Carr, S. A., and Baldwin, M. A., eds), pp. 163–177, Humana Press, Totowa, NJ
30. Eng, J., McCormack, A. L., and Yates, J. R., III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989
31. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951
32. Johnston, M., and Carlson, M. (1991) in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression* (Broach, J. R., Pringle, J. R., and Jones, E. W., eds) Vol. 2, pp. 193–281, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
33. Johnston, M. (1999) Feasting, fasting and fermenting. Glucose sensing in yeast and other cells. *Trends Genet.* **15**, 29–33
34. Lehmann, E. L. (1975) *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco
35. Gygi, S. P., Rist, B., Griffin, T. J., Eng, J., and Aebersold, R. (2002) Proteome analysis of low abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J. Proteome Res.* **1**, 47–54
36. Hodges, P. E., Payne, W. E., and Garrels, J. I. (1998) The Yeast Protein Database (YPD): a curated proteome database for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **26**, 68–72
37. Young, E. T., and Pilgrim, D. (1985) Isolation and DNA sequence of ADH3, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**, 3024–3034
38. Navarro-Avino, J. P., Prasad, R., Miralles, V. J., Benito, R. M., and Serrano, R. (1999) A proposal for nomenclature of aldehyde dehydrogenases in *Saccharomyces cerevisiae* and characterization of the stress-inducible ALD2 and ALD3 genes. *Yeast* **15**, 829–842

39. Wang, X., Mann, C. J., Bai, Y., Ni, L., and Weiner, H. (1998) Molecular cloning, characterization, and potential roles of cytosolic and mitochondrial aldehyde dehydrogenases in ethanol metabolism in *Saccharomyces cerevisiae*. *J. Bacteriol.* **180**, 822–830
40. van den Berg, M. A., de Jong-Gubbels, P., Kortland, C. J., van Dijken, J. P., Pronk, J. T., and Steensma, H. Y. (1996) The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. *J. Biol. Chem.* **271**, 28953–28959
41. Cozzone, A. J. (1998) Regulation of acetate metabolism by protein phosphorylation in enteric bacteria. *Annu. Rev. Microbiol.* **52**, 127–164
42. Anderson, S. L., Minard, K. I., and McAlister-Henn, L. (2000) Allosteric inhibition of NAD⁺-specific isocitrate dehydrogenase by a mitochondrial mRNA. *Biochemistry* **39**, 5623–5629
43. Rep, M., and Grivell, L. A. (1996) The role of protein degradation in mitochondrial function and biogenesis. *Curr. Genet.* **30**, 367–380
44. Luzikov, V. N., Novikova, L. A., Tikhonov, A. N., and Zubatov, A. S. (1983) Correlation between the rate of proteolysis of mitochondrial translation products and fluidity of the mitochondrial inner membrane in *Saccharomyces cerevisiae* yeast. Alteration of the rate of proteolysis under glucose repression. *Biochem. J.* **214**, 785–794
45. Kief, D. R., and Warner, J. R. (1981) Coordinate control of syntheses of ribosomal ribonucleic acid and ribosomal proteins during nutritional shift-up in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **1**, 1007–1015
46. Mager, W. H., and Planta, R. J. (1991) Coordinate expression of ribosomal protein genes in yeast as a function of cellular growth rate. *Mol. Cell. Biochem.* **104**, 181–187
47. Planta, R. J. (1997) Regulation of ribosome synthesis in yeast. *Yeast* **13**, 1505–1518
48. el Baradi, T. T., van der Sande, C. A., Mager, W. H., Raue, H. A., and Planta, R. J. (1986) The cellular level of yeast ribosomal protein L25 is controlled principally by rapid degradation of excess protein. *Curr. Genet.* **10**, 733–739
49. Tsay, Y. F., Thompson, J. R., Rotenberg, M. O., Larkin, J. C., and Woolford, J. L., Jr. (1988) Ribosomal protein synthesis is not regulated at the translational level in *Saccharomyces cerevisiae*: balanced accumulation of ribosomal proteins L16 and rp59 is mediated by turnover of excess protein. *Genes Dev.* **2**, 664–676
50. Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, J., and Deshaies, R. J. (2000) Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439
51. Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16**, 939–945
52. Hughes, J. D., Estep, P. W., Tavazoie, S., and Church, G. M. (2000) Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **296**, 1205–1214
53. Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367
54. Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Jr., Hieter, P., Vogelstein, B., and Kinzler, K. W. (1997) Characterization of the yeast transcriptome. *Cell* **88**, 243–251
55. Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
56. Chalmers, M. J., and Gaskell, S. J. (2000) Advances in mass spectrometry for proteome analysis. *Curr. Opin. Biotechnol.* **11**, 384–390
57. Zhou, H., Watts, J. D., and Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375–378
58. Oda, Y., Nagasu, T., and Chait, B. T. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382