Funct Integr Genomics (2002) 2:171–180
DOI 10.1007/s10142-002-0065-3

REVIEW

Christine E. Horak · Michael Snyder

Global analysis of gene expression in yeast

Received: 20 May 2002 / Accepted: 28 May 2002 / Published online: 10 July 2002

Abstract In the past decade, there has been an intense effort to comprehensively catalogue the expressed genes in the yeast *Saccharomyces cerevisiae* and to determine the absolute and relative abundance of transcript and protein levels under different cellular conditions. Several methods have been developed to monitor gene expression: DNA microarray analysis, Serial Analysis of Gene Expression (SAGE), kinetic RT-PCR and monitoring expression of β-galactosidase fusion proteins. These techniques have been used to measure transcript and protein abundance in different developmental states and under different environmental stimuli. A wealth of expression data for yeast is now publicly available through several web sites. The expression information that exists has the obvious benefits of providing a better understanding of the gene expression patterns that accompany changes in a yeast cell's environmental and developmental states. This data has also, however, provided clues to unraveling the complicated questions surrounding gene regulation: why and how is gene expression controlled?

Keywords Microarray · SAGE · Kinetic RT-PCR · Beta-galactosidase fusion proteins

Why monitor gene expression?

Understanding the composition of the expressed genome, knowing both what genes are expressed and the extent to which they are represented, provides three different types of information. First, the set of genes expressed at a given time reflects the cellular processes that a yeast cell is undergoing. Therefore, identifying the functional classes of genes that are expressed or repressed under specific conditions allows for a better understanding of the molecular responses to particular stimuli. For example, the genes that are induced during sporulation and their temporal pattern mirror the events of recombination, meiosis, spore formation and spore wall maturation that are occurring within the yeast (Chu et al. 1998; Primig et al. 2000).

Second, expression studies can help elucidate the function of characterized genes and uncharacterized genes. It is now clear that genes which share similar expression patterns often participate in the same cellular processes (Cho et al. 1998; Chu et al. 1998; DeRisi et al. 1997; Gasch et al. 2000, 2001; Primig et al. 2000; Spellman et al. 1998). In the case of sporulation, 33 uncharacterized ORFs that were found to be induced during sporulation in microarray expression studies were found to be required for the formation of wild-type spores by mutation analysis (Rabitsch et al. 2001).

The third type of information that can be deduced from expression studies is information about gene regulation. The patterns and phenomena identified within the expression data can be used to address the complexities and nuances of the regulation of gene expression. Cataloguing the yeast transcriptome and proteome, and their variations, has provided some evidence for understanding why some genes have restricted or limited expression patterns. The data has also provided clues for defining the upstream molecular regulators of the observed expression patterns. This review describes the current methods for monitoring gene expression in yeast, the plethora of available expression data and how it has been used to understand gene regulation.

Methods for measuring yeast gene expression

Several new methods have been developed in recent years to monitor gene expression on a large scale. Using these approaches, transcript or protein abundance can be qualitatively or quantitatively determined. Most of these

C.E. Horak · M. Snyder (✉)
Department of Molecular, Cellular and Developmental Biology,
Yale University, New Haven, CT 06520, USA
e-mail: michael.snyder@yale.edu
Tel.: +1-203-4326139, Fax: +1-203-4326161

M. Snyder
926 Kline Biology Tower, Yale University, 266 Whitney Ave,
New Haven, CT 06520, USA

Table 1 Methods for measuring gene expression

Method	Measures protein or mRNA abundance	Quantitativeness	Comprehensiveness	Facility	Sensitivity	References
Microarray and DNA chip analysis	mRNA	Semi-quantitative	All annotated sequences	Easy; single hybridization	0.5–200 transcripts/cell	(Lashkari et al. 1997; Shalon et al. 1996)
SAGE	mRNA	Quantitative	All expressed sequences	Laborious; sequencing 60,000 tags	0.3–200 transcripts/cell	(Kal et al. 1999; Velculescu et al. 1997)
Kinetic RT-PCR	mRNA	Quantitative	All annotated sequences	Laborious; individual PCR	0.001–200 transcripts/cell	(Holland 2002; Kang et al. 2000)
β-gal fusion library	Protein	Qualitative	All expressed sequences with transposon insertions	Easy; 96-well format filter assay	0.001–200 proteins/cell	(Ross-MacDonald et al. 1999)
2D-gel electrophoresis	Protein	Quantitative	Expressed proteins that can be separated by their molecular weight and isoelectric point	Laborious; LC/MS and scintillation counting for individual proteins	1–1,000 proteins/cell	(Futcher et al. 1999; Gygi et al. 1999)
Antibody arrays	Protein	Semi-quantitative	All annotated sequences	Easy; single hybridization	?	

techniques measure mRNA abundance, such as microarray and DNA chips, SAGE (Serial Analysis of Gene Expression) and kinetic RT-PCR. Relative protein levels have been measured on a large scale using a library of protein β-galactosidase fusions (Ross-MacDonald et al. 1999). Two-dimensional gel electrophoresis has also been utilized to measure protein abundance on a large-scale (Futcher et al. 1999; Gygi et al. 1999). In the future, arrays of antibodies raised against all the yeast proteins may be available to comprehensively monitor protein abundance. Table 1 summarizes each of these methods and denotes each of their advantages and disadvantages.

Microarray and DNA chip analysis

Microarray and DNA chip techniques involve hybridization of fluor-labeled cDNA to an array containing open reading frame (ORF) sequences (Lashkari et al. 1997; Shalon et al. 1996). Upon completion of sequencing the yeast genome, the sequence was annotated to predict all coding sequences. DNA chips and microarrays have been designed to contain sequences of those predefined ORFs. The DNA chips for yeast are commercially available through Affymetrix. They are high-density oligonucleotide arrays, where 25-mers unique to each gene are synthesized directly onto these arrays. DNA microarrays, on the other hand, are arrays of PCR-amplified ORFs. Each annotated sequence is separately amplified and printed in a unique position on the arrays. The advantage of DNA chip and microarray technology is that all of the approximately 6,300 predicted ORFs can be easily monitored simultaneously in a single experiment. A schematic of the microarray analysis approach is shown in Fig. 1A. Figure 1B shows a portion of an ORF microarray that has been hybridized with fluor-labeled cDNA from wild type and *swi4Δ* cells.

Most of the yeast expression data that is currently available is based on microarray and DNA chip transcript profiling. This expression analysis has been used to characterize yeast cell response to several different physiological conditions, including glucose limitation (Ferea et al. 1999) and the diauxic shift (DeRisi et al. 1997). Many studies have been done to examine yeast expressional response to different sources of stress. The transcriptional response to such stresses as DNA damage (Gasch et al. 2001), starvation (Chu et al. 1998), oxidizing and reducing agents and heat shock (Gasch et al. 2000) are the subject of a review by Gasch et al. in this issue.

Microarray expression profiling has examined the changes in gene expression throughout several different developmental stages. Multiple studies have been performed to profile expression during the cell cycle and during the stages of sporulation (Cho et al. 1998; Chu et al. 1998; Primig et al. 2000; Spellman et al. 1998). The transcriptional program of mating cells and cells undergoing pseudohyphal growth have also been elucidated (Roberts et al. 2000). In addition, expression analysis has been performed for a myriad of mutant yeast strains under different growth conditions (Hellauer et al. 2001; Holstege et al. 1998; Lopez and Baker 2000; Sudarsanam et al. 2000).

This approach of using arrays containing PCR amplified ORFs is semi-quantitative as the relative fluorescence intensity of the sequence elements is a good indicator of relative differences in gene expression, but is a

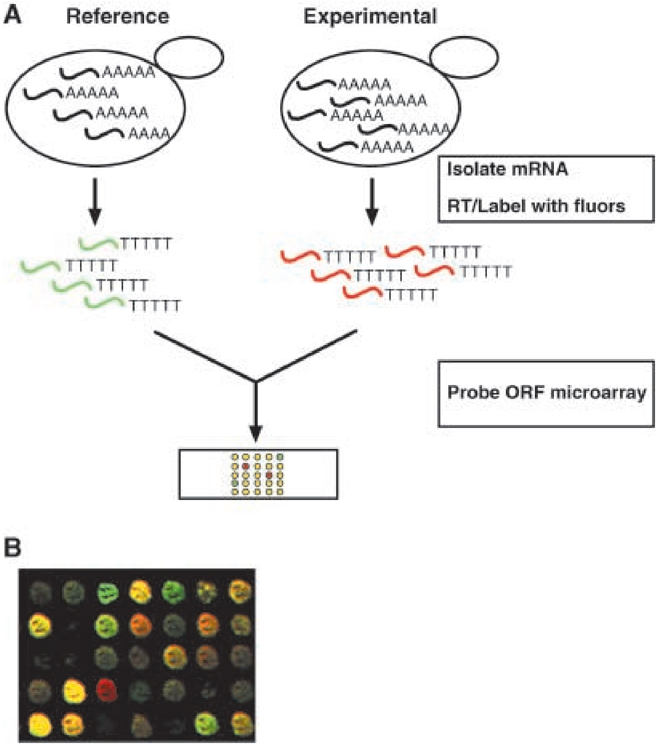

Fig. 1 **A** A schematic of DNA microarray expression analysis in yeast. First, mRNA is isolated from a reference yeast strain (usually wild-type or untreated cells) and an experimental strain (usually deletion or chemically treated cells). Reverse transcription is used to incorporate fluorescent-dye-conjugated nucleotides into cDNA. The fluorescent probes from each yeast strain are then simultaneously probed to a microarray of all yeast open reading frames (ORFs). **B** A portion of a yeast ORF microarray. Individual spots represent unique ORFs. The microarray was probed with a reference cDNA (derived from wild-type cells) labeled with Cy3 dye (*green*) and an experimental cDNA (derived from *swi4Δ* cells) labeled with Cy5 dye (*red*). A *red spot* indicates a relative enrichment of that gene transcript in the experimental mRNA pool, while a *green spot* indicates a relative enrichment for that particular ORF in the reference mRNA population. *Yellow spots* are indicative of equal quantities of that specific ORF in both pools of mRNA. The fluorescence intensity of each spot corresponds with transcript abundance in a population of transcripts

poor indicator of transcript abundance. The fluorescent signal intensity is influenced by the probe or array element length and the melting temperature of the sequences. The use of oligonucleotide arrays can provide more quantitative information on transcript levels. A distinct disadvantage of current microarray methods is that they require a priori information on the sequences that are expressed in the genome and are dependent on thorough and precise annotation of the sequence. Genes that are expressed and have not been annotated are missing from the analysis. This problem may be swiftly remedied by incorporating all genome sequences or genes that are newly identified by other expression techniques, such as SAGE and gene-trapping (see below).

Serial analysis of gene expression

Serial analysis of gene expression (SAGE) also monitors gene transcripts. It involves preparation and concatenation of 10- to 15-bp unique expressed sequence tags from the three prime end of cDNAs and serial sequencing of the concatamers (Velculescu et al. 1997). The unique sequence tags are generated by digestion of double-stranded cDNA with the *NlaIII* restriction enzyme, followed by polyA selection, which will purify the 3′-most *NlaIII* cDNA fragments. A *BsmFI* linker is subsequently ligated to this pool of fragments. Since the *BsmFI* recognition sequence overlaps with the *NlaIII* site and the *BsmFI* enzyme cuts 14 bp away from its recognition element, short, unique sequences are created upon digestion with *BsmFI*. These sequence "tags" are ligated together, PCR-amplified and sequenced. This method is schematized in Fig. 2.

Using this approach, Velculescu et al. studied transcript levels during vegetative growth and in cells arrested at S phase with hydroxyurea and at the G2/M phase with nocadazole (Velculescu et al. 1997). These studies with SAGE have shown that each yeast cell contains approximately 15,000 transcripts that represent at least 76% of the 6,300 predicted yeast genes. A vast majority of these genes (75%) are represented by one or less transcript per cell (Velculescu et al. 1997).

A major advantage of this technique is that it is quantitative and the absolute levels of mRNA can be determined. The number of times the unique tag is sequenced correlates with its abundance in the transcript population. Unfortunately, nearly 60,000 SAGE tags need to be sequenced in order for the analysis to cover much of the transcriptome, which makes this approach laborious (Velculescu et al. 1997) and even this effort is not saturating. However, this method has been valuable for identifying previously un-annotated genes, since it is not dependent on predetermined sequence annotations. Thirty such NORFs (non-annotated ORFs) were identified in logarithmically growing cells (Velculescu et al. 1997). It has also been successfully used in studies monitoring gene expression in wild-type and mutant backgrounds when oleate is used as a carbon source (Kal et al. 1999).

Kinetic RT-PCR

Given that both SAGE and microarray analysis have a lower limit threshold for detection at 0.3–0.5 transcripts per cell, expression data for at least a portion of 25% of yeast genes is lacking (Holland 2002). Recently, kinetic (or real time) reverse transcription-polymerase chain reaction (RT-PCR) analysis, which is illustrated in Fig. 3A, has revealed that some genes are expressed at levels as low as one-thousandth of a transcript per cell (Holland 2002). With this technique, the concentration of a specific cDNA species is determined by its PCR-product accumulation rate. A fluorescent DNA indicator, such as ethidium bromide, is included in each PCR reaction, so

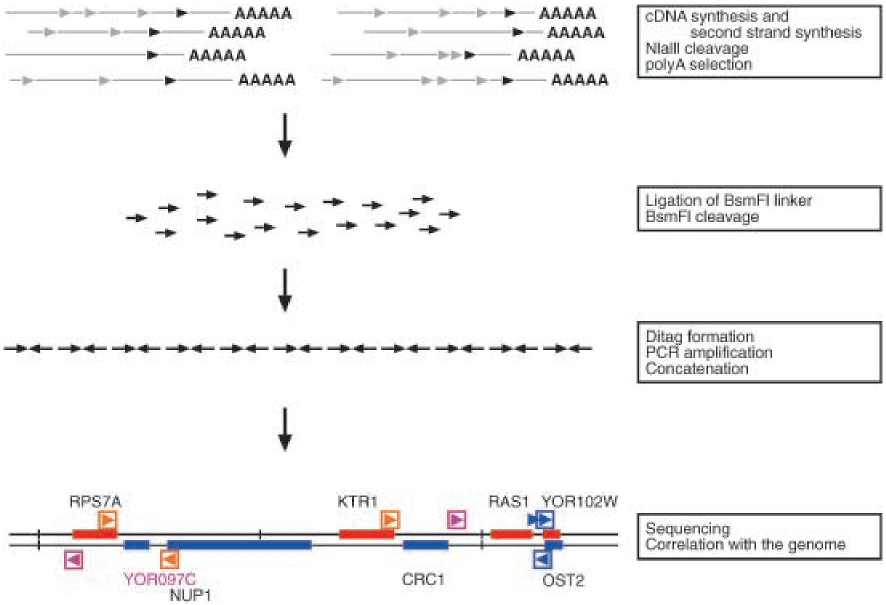

Fig. 2 A schematic of serial analysis of gene expression (SAGE; adapted from Velculescu et al. 1997). Three-prime end cDNA fragments are first created by digestion of double-stranded cDNA with *NlaIII* and polyA selection of the cDNA fragments. The 3′-most *NlaIII* sites are indicated with a *black arrowhead* and all other *NlaIII* sites within a cDNA are represented by *gray arrowheads*. *BsmFI* sites are ligated to selected cDNA fragments and a unique SAGE tag is then generated by digestion with *BsmFI* enzyme, which cuts 14 bp away from its recognition site that overlaps with the original *NlaIII* site. These tags are ligated together to create ditags, which are subsequently amplified by PCR. The amplified ditags are cleaved with *NlaIII* and self-ligated. The concatenated tags are then serially sequenced, quantitated and compared against the genome sequence

that product accumulation can be monitored at each amplification cycle by a kinetic thermal cycler (shown in Fig. 3B). Examples of kinetic RT-PCR curves are illustrated in Fig. 3C. Transcript abundance is related to the inverse log of the number of cycles it takes to reach an arbitrary unit of fluorescence intensity that falls along the linear portion of the product accumulation curve.

The steady state transcript levels of the 65 genes on the left arm of chromosome 3 and 185 transcription factors genes were examined using kinetic RT-PCR. These studies revealed that the most abundant transcripts, which are present at 50–200 copies per cell, have functional roles in protein synthesis and general metabolism, while transcription factors as a class of genes tend to be among the least abundant transcripts (Holland 2002; Kang et al. 2000). Obviously, kinetic RT-PCR is sensitive enough to monitor the expression of these weakly expressed genes, however it has been too arduous to perform on a genome-wide scale.

β-Galactosidase fusion proteins

Much of the published yeast genomic expression data involves monitoring mRNA levels; however, cataloguing the yeast proteome is of key interest. The molecular activities within a cell are more readily characterized by the proteins that are present than by the transcriptome. The prevailing approach for monitoring relative protein levels is a library of expressed ORFs fused to a promoter-less and 5′-truncated β-galactosidase gene. A library of gene chimeras was created by random insertion of a modified transposon containing the β-gal gene and a yeast selectable marker as shown in Fig. 4A (Ross-MacDonald et al. 1999). When the β-gal gene is inserted in-frame with an ORF, it will be transcribed and translated. The level of β-gal activity can be assayed on a filter and may be correlated with the level of expression of the gene to which it is fused. An example of this β-gal fusion colormetric assay is shown in Fig. 4B. This approach has been used to assay for gene expression under vegetative growth, sporulation, in mating-pheromone-treated cells and in mutant backgrounds (Erdman and Snyder 2001; Erdman et al. 1998; Ross-MacDonald et al. 1999). As all the fusions in the library have been sequenced and arrayed in a 96-well format, it is relatively easy to assay for β-gal activity under different conditions by simply plating the library to different growth media. This approach can also be used to identify non-annotated expressed genes and, in fact, 137 of these previously overlooked genes were identified from this fusion library (Kumar et al. 2002c). The major disadvantage of this method is that the library is not complete, in that it contains only about 60% of the 6,300 annotated genes (Kumar et al. 2002b). Another disadvantage is that the colormetric β-gal filter assays used to monitor protein abundance are qualitative, not quantitative. Thus, only relative protein abundance can be derived with this assay.

Other technologies

Other methods of protein abundance monitoring are emerging. As described below, two-dimensional gel

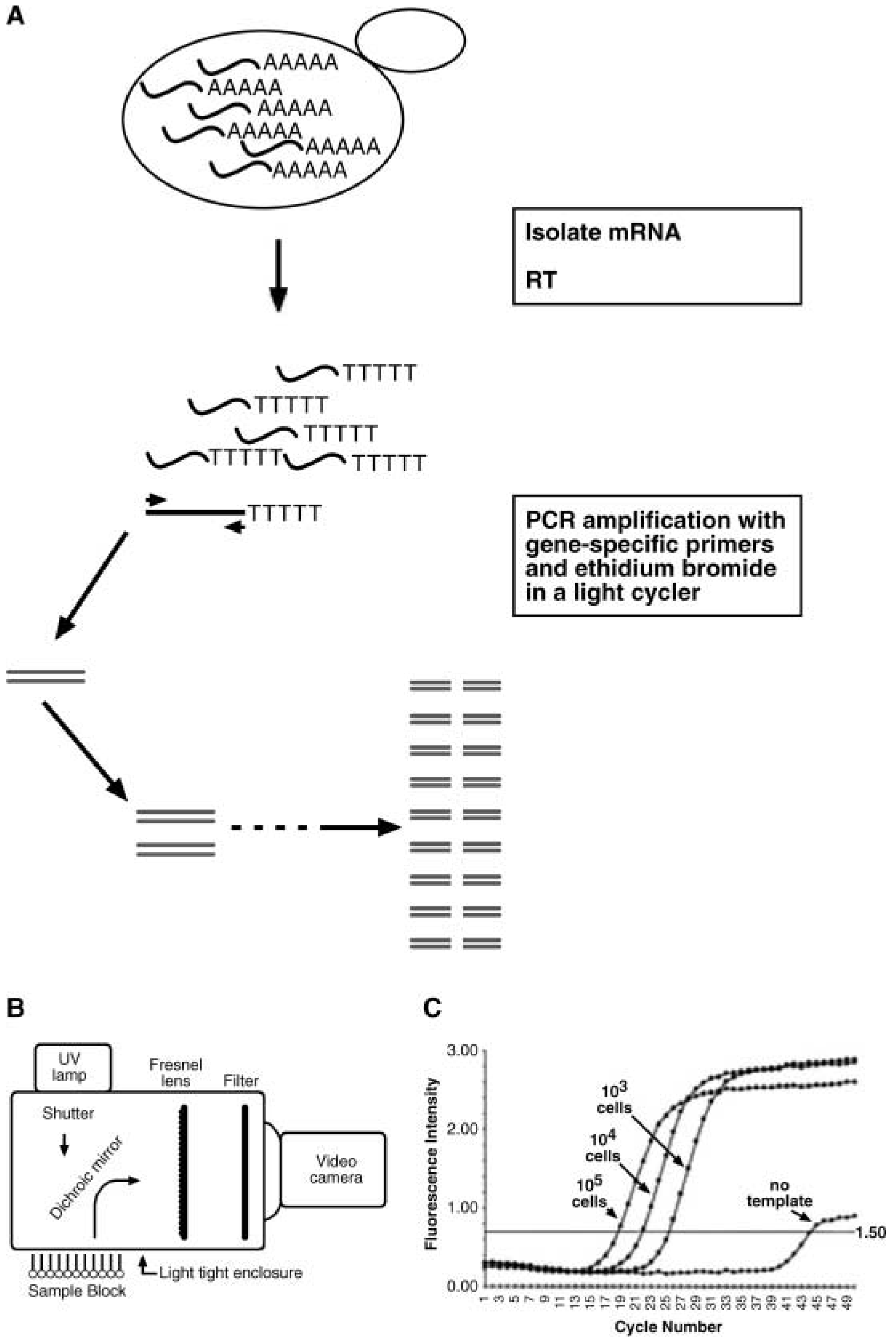

Fig. 3 **A** A schematic of kinetic RT-PCR. First, cDNA is generated and then a specific cDNA species is amplified by PCR with unique primers. Ethidium bromide is incorporated in the PCR as an indicator of DNA concentration (represented as *orange fragments*) which can be monitored by a kinetic thermal cycle at each amplification step. **B** The optical components of the kinetic thermal cycler are illustrated. UV light is shone through the shutter onto the dichroic mirror and onto the PCR sample block. Sample fluorescence reflects off the mirror and through the Fresnel lens and interference filter into a CCD camera (illustration adapted from Kang et al. 2000). **C** Examples of kinetic RT-PCR curves for a single primer pair with 10^3, 10^4 or 10^5 cell equivalents of total RNA and a no template control. A relative fluorescence at 1.5 is designated as the arbitrary fluorescence level (AFL) to compare the kinetic curves. Specific transcript concentration is proportional to the inverse log of the cycle number at which the curve crosses the AFL (results taken from Kang et al. 2000)

electrophoresis has been used to monitor the abundance of over 1,500 proteins. With this method, ^{35}S-labeled proteins are separated by their molecular weight and their isoelectric point by polyacrylamide gel electrophoresis. Protein spots can then be excised and quantitated by scintillation counting and identified by mass spectrometry. While this approach allows for accurate measurements of protein abundance, it is quite laborious. Another drawback is that not all proteins can be fully separated on a two-dimensional gel, so it can never be completely comprehensive.

Perhaps, the most promising approach for easily monitoring protein concentrations is an array of antibodies raised against all yeast proteins. In a manner analogous to DNA microarrays, proteins isolated from yeast can be labeled and hybridized to the antibody array and protein abundance can be surmised from the fluorescent intensity signal. The utility of this method will hinge on the quality and specificity of antibodies that are prepared.

Transcript vs protein abundance

Although, gene-to-gene variation exists, transcript and protein levels are not always strongly correlated (Futcher et al. 1999; Gygi et al. 1999). This observed discrepancy

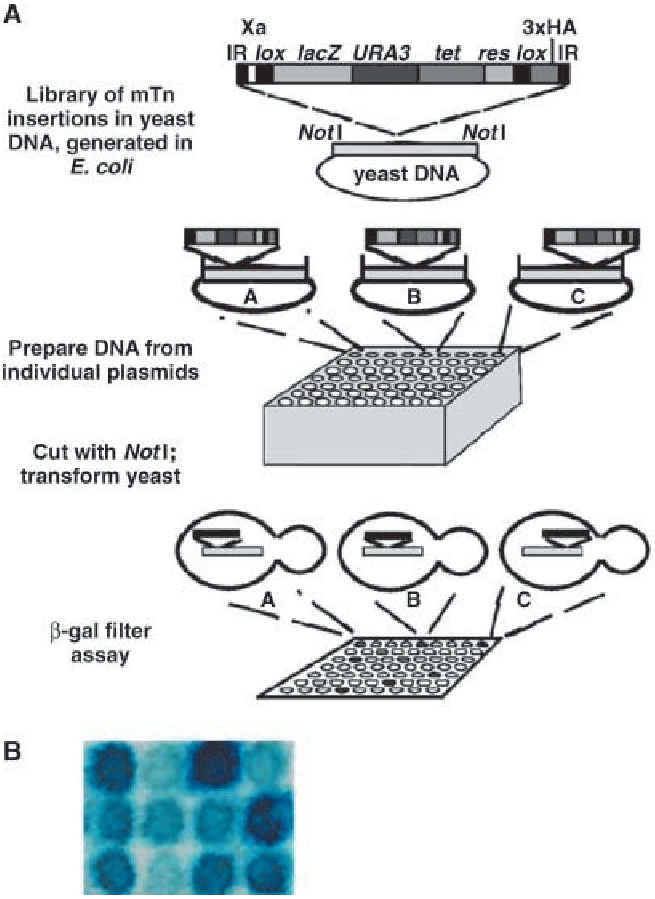
A
Xa
3xHA
IR lox lacZ URA3 tet res lox IR
Library of mTn insertions in yeast DNA, generated in E. coli
NotI
NotI
yeast DNA
A
B
C
Prepare DNA from individual plasmids
Cut with NotI; transform yeast
A
B
C
β-gal filter assay
B

Fig. 4 **A** A schematic of random transposon insertion of the *lacZ* gene into yeast ORFs (adapted from Ross-MacDonald et al. 1999). A multipurpose transposon (*mTn*) containing promoterless *lacZ* gene and a yeast selectable marker is randomly inserted into yeast DNA within *E. coli*. Yeast DNA is isolated from the *E. coli* plasmid sequence by *NotI* digestions. Yeast insertions are transformed into yeast cells by homologous recombination. In-frame *lacZ* insertions are then identified by a β-galactosidase filter assay. **B** The results of a colormetric filter assay for β-gal activity from a small collection of *lacZ* gene fusions. The *intensity of blue color* corresponds to the level of β-gal activity and thus the level of expression of the gene to which the *lacZ* gene is fused

likely reflects differences in relative translation efficiencies or stability of the protein products. A well-studied example of translational control of gene expression comes from the analysis of *GCN4*, a transcription factor involved in the regulation of amino acid synthesis. Gcn4p is translationally induced when histidine is depleted (Natarajan et al. 2001).

Two studies comparing protein abundance data from two-dimensional gel electrophoresis spots that were identified and quantified by tandem capillary liquid chromatography-mass spectrometry with SAGE data found little to moderate correlation between mRNA and protein abundance for the lowest abundance proteins (Gygi et al. 1999). More than 1,500 proteins were analyzed in this manner. Similarly, only two-thirds of the yeast genes identified as induced during sporulation using β-gal fusions were also found to be induced using mRNA microarray analysis (Ross-MacDonald et al. 1999).

Both transcript and protein abundance data will prove invaluable. Both kinds of data provide information about the molecular responses of a cell and, when the data is combined, major modes of regulation can be inferred. Where their abundance is correlative, it can be presumed that abundance is largely regulated at the level of transcription or mRNA processing. When discrepancies are observed in protein and transcript levels, it presumably indicates that abundance is regulated at the translational level or via protein degradation/stabilization.

Yeast expression database resources

Over fifty global expression studies have now been performed using the approaches described above. These studies evaluate yeast expression in about 1,000 physiological conditions and developmental states. The data obtained from these large-scale expression experiments are now publicly available via the web. These websites are summarized in Table 2.

The cell cycle and sporulation expression data can be accessed through two websites that have been devoted to each of these expression data sets (Table 2). These data sets plus most or all other microarray expression data can be accessed through one of three websites: the Expression Connection, the Yeast Microarray Global Viewer (yMGV) and Webminer (Table 2; Heiman and Walter 2000; Marc et al. 2001). The cell cycle analysis project and sporulation project datasets, as well as the Expression Connection and the yMGV can be searched for single gene expression profiles in multiple experiments. yMGV and Webminer provide tools to isolate subsets of genes that share similar expression patterns across multiple sets of data. Identifying these sets of genes can help to ascribe a function to genes within the set or derive a potential mechanism of transcriptional regulation, for example identifying transcription factors involved in regulating expression (Heiman and Walter 2000; Marc et al. 2001). The yTAFNET database focuses on expression profiles of mutated or ectopically expressed yeast transcriptional regulators (Table 2; Devaux et al. 2001b).

Expression data from SAGE and β-gal fusion analysis can be accessed through separate databases. Table 2 also lists these websites. SAGE data for vegetatively growing cells is publicly available and the β-gal fusion expression assay data for vegetatively growing cells and sporulating cells can be accessed via the TRIPLES database (Kumar et al. 2002b).

The websites housing all of the expression information for yeast prove invaluable for identifying the expressional response of a single gene under a myriad of growth conditions. The tools provided by some of these databases easily allow for comparison of expression profiles under different conditions and, as will be discussed below, the expression data can be accessed to begin to understand why and how gene expression is controlled.

Table 2 Useful websites

Expression data analysis	
Expression connection	This database can be accessed through *Saccharomyces* Genome Database and allows one to search for single gene expression profiles from 11 microarray data sets generated at Stanford University and Rosetta Pharmaceuticals. http://genome-www.stanford.edu/cgi-bin/SGD/expression/expressionConnection
Webminer	Boolean searches from multiple microarray datasets to identify genes that are similarly regulated (Heiman and Walter 2000). http://webminer.ucsf.edu/
Yeast microarray global viewer	This database can be searched for single and multiple gene transcript profiles for 50 different microarray studies, but whole sets of genes with similar expression profiles for multiple datasets can be identified. Other interesting data is also available including the least and most variant gene expression (Marc et al. 2001). http://www.transcriptome.ens.fr/ymgv/
Yeast cell cycle analysis project	Cell cycle microarray data from Spellman et al. (1998). http://genome-www.stanford.edu/cellcycle/
The transcriptional program of sporulation	Sporulation microarray data from Chu et al. (1998). http://cmgm.stanford.edu/pbrown/sporulation/
SAGE data	SAGE data from Velculescu et al. (1997). ftp://genome-ftp.stanford.edu/pub/yeast/tables
TRIPLES	Database of expression, localization and phenotypic analysis of the library of transposon insertions (Kumar et al. 2002b). http://ycmi.med.yale.edu/ygac/triples
yTAFNET	Database devoted to microarray expression profiles in specific transcription factor mutants or in response to ectopic transcription factor expression (Devaux et al. 2001b). http://transcriptome.ens.fr/ytafnet/
Promoter analysis	
SCPD	http://cgsigma.cshl.org/jian/
TRANSFAC	http://transfac.gbf.de/TRANSFAC/
The mirage website	http://www.ifti.org/
Systematic determination of genetic network architecture	http://arep.med.harvard.edu/network_discovery
Regulatory sequence analysis tools	http://copan.cifn.unam.mx/~jvanheld/rsa-tools/

Importance of gene regulation

All of these expression studies clearly illustrate that a shift in environmental or developmental states is often accompanied by a shift in the genes that are expressed. Why is gene expression altered depending on cellular status? There are several answers to this question that have been supported by the expression data itself.

For the most part, genes are expressed only under limited conditions because quite simply they are only needed at specific times or under particular conditions and presumably it would be a waste of energy and resources to express them constitutively. For example, genes involved in the process of DNA synthesis are only needed during the synthesis phase of the cell cycle. Most of these genes, in fact, exhibit peak transcript levels in late G1 just before the DNA synthesis phase of the cell cycle (Cho et al. 1998; Spellman et al. 1998).

In accordance with this theory that gene expression is regulated to conserve energy, one would predict then that the largest, most expensive genes in energy terms would have the most restricted expression. A recent study corroborated this prediction. Using the quantitative and semi-quantitative transcript data that is now available from SAGE and microarray experiments, respectively, mRNA concentration per cell was found to be negatively correlated with protein length (Coghlan and Wolfe 2000).

While a majority of genes have restricted expression to conserve energy, some genes' expression may be regulated because their inappropriate expression is toxic. Overabundance of several proteins is known to be lethal, most notably, several regulatory components of karyogamy, including *KAR1* (Rose and Fink 1987). In addition, misexpression of other genes will promote a less advantageous developmental state. For example, the expression of *MUC1* is generally limited to particular starvation conditions and once expressed will promote cell-cell adhesion and hyperinvasive growth (Guo et al. 2000). While this response may be advantageous under nutrient-restricted conditions, the misexpression of *MUC1* under rich growth conditions will slow yeast growth. Microarray and northern analysis have shown that *MUC1* expression is dependent on four different transcription factors via at least three different upstream pathways. The multiple mechanisms employed to regulate *MUC1* expression indicates the importance of restricting the expression of this gene (Pan and Heitman 2000; Roberts et al. 2000).

Gene expression is also linked to cellular and environmental states not only because it may be an energy waste or because it may be harmful, but also because timing of expression may be critical for the order of protein complex formation. Cell cycle microarray analysis revealed that the homologs *BUD8* and *BUD9*, which are involved in bud site selection, have disparate peaks in mRNA abundance during the cell cycle (Spellman et al. 1998). *BUD9* transcript levels peak in the G1 phase, while the homolog *BUD8* peaks in the mitosis phase of the cycle. Promoter swapping studies indicate that the timing of expression of these genes is critical for their localization and function. Bud9p expression from *BUD8* promoter can rescue a *bud8Δ* mutant, but not a *bud9Δ* mutant, and vice versa (Schenkman et al. 2002).

Inherent in the plethora of gene expression data is explanations for why gene expression is controlled. The data would predict that energy and resource conservation is a major reason for restricting gene expression, but gene expression may also be limited to prevent inappropriate physical responses. There is evidence to suggest that temporal expression patterns are required for appropriate protein complex formation for at least some proteins. Pinpointing the molecular mechanisms that control and restrict these important expression patterns is a challenge, but one which has also been aided by genome-wide expression profiling.

Regulation of gene expression

Gene expression is regulated at both the mRNA and protein levels. Thus far, great emphasis has been placed on defining the transcription factors that are important in regulating the gene expression responses. There are approximately 500 known and potential transcription factors and chromatin modifiers in yeast (Kumar et al. 2002a). Each of these general and specific transcriptional regulators can influence the expression of ten to hundreds of genes. The web of transcriptional control is presumably quite complex and features many redundant mechanisms for controlling gene expression. Three methods have currently been used to map this complex circuit of gene regulation.

The first approach utilizes bioinformatics to deduce putative transcriptional regulators by promoter sequence analysis. Clusters of co-regulated genes are identified from individual or multiple expression data sets and the promoter regions of these genes can be searched for similar sequence motifs. Several methods for DNA sequence motif searching are now available through the promoter analysis websites listed in Table 2. Only a handful of transcription factors have well-defined consensus binding motifs, thus this approach is somewhat limited for pinpointing the transcription factors that regulate gene expression clusters. Typically, promoter motif searching has been used to find enrichments for known transcription factor binding sites in the promoters of co-regulated genes. For example, 58% and 52% of the genes in the *CLN2* cluster of cell cycle genes (those genes with peak mRNA abundance in the late G1 phase) contained an MCB or SCB element, respectively, in their promoter regions (Spellman et al. 1998). These elements are the known binding sequences for two transcription factors that are known to regulate the G1/S transition. Sequence motifs can be identified from clusters of genes without a priori information on the binding sequence. The consensus binding sequence for the forkhead transcription factors, Fkh1p and Fkh2p, was identified from the promoter regions of cell-cycle-regulated genes with peak mRNA abundance in the G2 phase of the cell cycle without prior information regarding the consensus binding motifs of these factors (Zhu et al. 2000). Serendipitously, the consensus site of the forkhead proteins was determined around the time that this motif was revealed from the expression data (Zhu et al. 2000). In some cases, it will be possible to experimentally verify the identity of a transcription factor from a promoter sequence motif by a yeast one-hybrid approach. The putative binding sequence is coupled to a reporter and its expression is monitored when in the presence of a library of transcrip-

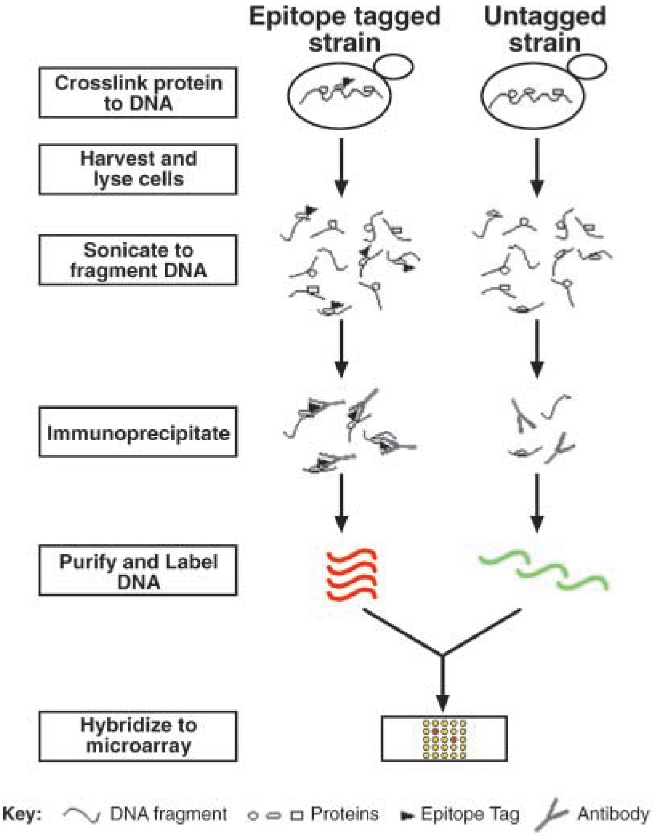

Fig. 5 A schematic drawing of the ChIp-chip technique for epitope-tagged DNA-binding proteins (adapted from Horak and Snyder 2002). The approach first involves chemical crosslinking of in vivo protein-DNA complexes. Yeast cells are subsequently lysed and their chromatin is sheared to 500- to 1,000-bp fragments by sonication. An epitope-tagged transcription factor and its bound DNA fragments can then be immunopurified with an antibody raised against the epitope tag. DNA fragments are isolated and labeled for hybridization to a yeast microarray of promoter fragments (see cover)

tion factors fused to the activation domain of a well-studied transcription factor, Gal4p.

Many attempts have been made to identify the full complement of transcription factor targets by expression profiling experiments when the activity of a specific transcription factor is altered. The yTAFNET database is devoted to these expression experiments (Table 2; Devaux et al. 2001b). Target identification from expression studies is based on the preconception that a target gene's transcript levels will be altered by more than twofold in the transcription factor mutant. Given that a majority of yeast genes are represented by one transcript per cell or less, more subtle variations in mRNA abundance may be biologically relevant. In general, simply deleting the transcription factor of interest will not give striking differences in transcript levels as observed via microarray analysis. This is not a surprise given the complexity of gene expression, the significant proportion of redundancy that exists and the fact that many transcription factors control the temporal pattern of gene expression rather than absolute transcript levels. Microarray analysis of yeast ectopically expressing specific transcription factors or expressing dominant transcription factor mutants has proven more useful in determining their target genes. Gene targets of Pdr1p, a transcription factor involved in pleiotropic drug resistance, were identified by expression analysis of a hyperactive allele of Pdr1p and using a fusion the Pdr1p DNA-interacting domain to the Gal4p transcriptional activation domain, which is constitutively active (DeRisi et al. 2000; Devaux et al. 2001a).

The third approach to transcription factor identification is chIp-chip, which also involves microarray analysis, but examines DNA-binding targets of transcription factors rather than transcript profiling. ChIp-chip analysis, which stands for *ch*romatin *i*mmuno*p*recipitation and DNA *chip* analysis, identifies sequences that are directly bound by a specific transcription factor of interest in vivo, by probing a microarray of promoter sequences with labeled DNA that has been immunoprecipitated with the protein. A schematic of this method is shown in Fig. 5. Using this technique, transcription factor targets can be found regardless of the factor's contribution to gene transcriptional regulation; thus genes that are only marginally influenced by a specific transcription factor can be isolated. Another advantage of this approach is that it allows for the direct detection of transcription factor targets, whereas expression analysis will also identify genes that are indirectly regulated by modulating transcripion factor activity. ChIp-chip has been performed with only a handful of yeast transcription factors so far, including many of the key factors in regulating the cell cycle (Iyer et al. 2001; Lieb et al. 2001; Ren et al. 2000; Simon et al. 2001). The binding data obtained from this type of analysis in combination with the information gathered from the other two approaches will be required to construct a complete map of transcriptional regulation in yeast.

Information regarding the yeast transcriptome and the proteome under numerous cellular conditions has been rapidly generated in recent years as a result of the development of several genomic approaches for monitoring transcript and protein abundance. Undoubtedly, the volume of expression data will grow, and more sensitive and direct approaches for expression analysis will emerge in the coming years, which will enhance the understanding of regulation of gene expression.

Acknowledgements Many thanks to Anuj Kumar for his critical review and comments on this manuscript. C. Horak is supported by a Howard Hughes Predoctoral Fellowship. Research in M. Snyder's laboratory is supported by the National Institutes of Health.

References

Cho R, Campbell M, Winzeler E, Steinmetz L, Conway A, Wodicka L, Wolfsberg T, Gabrielian A, Davis R (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2:65–73

Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown P, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. Science 282:699–705

Coghlan A, Wolfe K (2000) Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast 16:1131–1145

DeRisi J, Iyer V, Brown P (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

DeRisi J, van den Hazel B, Marc P, Balzi E, Brown P, Jacq C, Goffeau A (2000) Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. FEBS Lett 470:156–160

Devaux F, Marc P, Bouchoux C, Delaveau T, Hikkel I, Potier M, Jacq C (2001a) An artificial transcription activator mimics the genome-wide properties of the yeast Pdr1 transcription factor. EMBO Rep 2:493–8

Devaux F, Marc P, Jacq C (2001b) Transcriptomes, transcription activators and microarrays. FEBS Lett 498:140–144

Erdman S, Snyder M (2001) A filamentous growth response mediated by the yeast mating pathway. Genetics 159:919–928

Erdman S, Lin L, Malczynski M, Snyder M (1998) Pheromone-regulated genes required for yeast mating differentiation. J Cell Biol 140:461–483

Ferea T, Botstein D, Brown P, Rosenzweig R (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. Proc Natl Acad Sci USA 96:9721–9726

Futcher B, Latter G, Monardo P, McLaughlin C, Garrels J (1999) A sampling of the yeast proteome. Mol Cell Biol 19:7357–7368

Gasch A, Spellman P, Kao C, Carmel-Harel O, Eisen M, Storz G, Botstein D, Brown P (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11:4241–4257

Gasch A, Huang M, Metzner S, Botstein D, Elledge S, Brown P (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of yeast ATR homolog Mec1p. Mol Biol Cell 12:2987–3003

Guo B, Styles C, Feng Q, Fink G (2000) A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. Proc Natl Acad Sci USA 97:12158–12163

Gygi S, Rochon Y, Franza R, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720–1730

Heiman M, Walter P (2000) Prm1p, a pheromone-regulated multispanning membrane protein, facilitates plasma membrane fusion during yeast mating. J Cell Biol 151:719–730

Hellauer K, Sirard E, Turcotte B (2001) Decreased expression of specific genes in yeast cells lacking histone H1. J Biol Chem 276:13587–13592

Holland M (2002) Transcript abundance in yeast varies over six orders of magnitude. J Biol Chem

Holstege F, Jennings E, Wyrick J, Lee T, Hengartner C, Green M, Golub T, Lander E, Young R (1998) Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95:717–728

Horak CE, Snyder M (2002) ChIp-chip: a genomic approach for identifying transcription factor binding sites. (Guide to yeast genetics and molecular and cell biology) Methods Enzymol 350B:469–483

Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409:533–538

Kal A, van Zonneveld A, Benes V, van den Berg M, Koerkamp M, Albermann K, Strack N, Ruijter J, Richter A, Dujon B, Ansorge W, Tabak H (1999) Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. Mol Biol Cell 10:1859–1872

Kang J, Watson R, Fisher M, Higuchi R, Gelfand D, Holland M (2000) Transcript quantitation in total yeast cellular RNA using kinetic RT-PCR. Nucleic Acids Res 28:e2

Kumar A, Agarwal S, Heyman J, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, Cheung K-H, Miller P, Gerstein M, Roeder G, Snyder M (2002a) Subcellular localization of the yeast proteome. Genes Dev 16:707–719

Kumar A, Cheung K-H, Tosches N, Masiar P, Liu Y, Miller P, Snyder M (2002b) The TRIPLES database: a community resources for yeast molecular biology. Nucleic Acids Res 30:73–75

Kumar A, Harrison P, Cheung K-H, Lan N, Echols N, Bertone P, Miller P, Gerstein M, Snyder M (2002c) An integrated approach for finding overlooked genes in yeast. Nat Biotechnol 20:58–63

Lashkari D, DeRisi J, McCusker J, Namath A, Gentile C, Hwang S, Brown P, Davis R (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci USA 94:13057–13062

Lieb J, Liu X, Botstein D, Brown P (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28:327–34

Lopez M, Baker H (2000) Understanding the growth phenotype of the yeast gcr1 mutant in terms of global genomic expression patterns. J Bacteriol 182:4970–4978

Marc P, Devaux F, Jacq C (2001) yMGV: a database for visualization and data mining of published genome-wide yeast expression data. Nucleic Acids Res 29:e63–3

Natarajan K, Meyer M, Jackson B, Slade D, Roberts C, Hinnebusch A, Marton M (2001) Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Mol Cell Biol 21:4347–4368

Pan X, Heitman J (2000) Sok2 regulates yeast pseudohyphal differentiation via a transcription factor cascade that regulates cell-cell adhesion. Mol Cell Biol 22:8364–8372

Primig M, Williams R, Winzeler E, Tevzadze G, Conway A, Hwang S, Davis R, Esposito R (2000) The core meiotic transcriptome in budding yeasts. Nat Genet 26:415–423

Rabitsch K, Toth A, Galova M, Schleiffer A, Schaffner G, Aigner E, Rupp C, Penkner A, Moreno-Borchart A, Primig M, Esposito R, Klein F, Knop M, Nasmyth K (2001) A screen for genes required for meiosis and spore formation based on whole-genome expression. Curr Biol 11:1001–1009

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Itamar S, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. Science 290:2306–2309

Roberts C, Nelson B, Marton M, Stoughton R, Meyer M, Bennett H, He Y, Dai H, Walker W, Hughes T, Tyers M, Boone C, Friend S (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 287:873–878

Rose M, Fink G (1987) *KAR1*, a gene required for function of both intranuclear and extranuclear microtubules in yeast. Cell 48:1047–1060

Ross-MacDonald P, Coehlo P, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung K-H, Sheehan A, Symoniatis D, Umansky L, Heidtman M, Nelson F, Iwasaki H, Hager K, Gerstein M, Miller P, Roeder G, Snyder M (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature 402:413–418

Schenkman L, Caruso C, Page N, Pringle J (2002) The role of cell cycle-regulated expression in the localization of spatial landmark proteins in yeast. J Cell Biol 156:829–841

Shalon D, Smith S, Brown P (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6:639-645

Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. Cell 106:697–708

Spellman P, Sherlock G, Zhang M, Iyer V, Anders K, Eisen M, Brown P, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 9:3273–3297

Sudarsanam P, Iyer V, Brown P, Winston F (2000) Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 97:3364–3369

Velculescu V, Zhang L, Zhou W, Vogelstein J, Basrai M, Bassett D, Hieter P, Vogelstein B, Kinzler K (1997) Characterization of the yeast transcriptome. Cell 88:243–251

Zhu G, Spellman P, Volpe T, Brown P, Botstein D, Davis T, Futcher B (2000) Two yeast *forkhead* genes regulate the cell cycle and pseudohyphal growth. Nature 406:90–94