YEAST
Yeast **14**, 1209–1221 (1998)

Yeast Functional Analysis Reports

Transcriptional Profiling on all Open Reading Frames of *Saccharomyces cerevisiae*

NICOLE C. HAUSER[1]*, MARTIN VINGRON[2], MARCEL SCHEIDELER[1], BERNHARD KREMS[3], KLAUS HELLMUTH[4], KARL-DIETER ENTIAN[3] AND JÖRG D. HOHEISEL[1]

[1]*Molecular-Genetic Genome Analysis Group, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany*
[2]*Theoretical Bioinformatics, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany*
[3]*Institut für Mikrobiologie, Johann Wolfgang Goethe Universität, Gebäude 250, Marie Curie Str. 9, D-60439 Frankfurt/M., Germany*
[4]*Biochemie-Zentrum, Universität Heidelberg, Im Neuenheimer Feld 328, D-69120 Heidelberg, Germany*

Open reading frames (6116) of the budding yeast *Saccharomyces cerevisiae* were PCR-amplified from genomic DNA using 12,232 primers specific to the ends of the coding sequences; the success rate of amplification was 97%. PCR-products were made accessible to hybridization by being arrayed at very high density on solid support media using various robotic devices. Probes made from total RNA preparations were hybridized for the analysis of the transcriptional activity of yeast under various growth conditions and of different strains. Experimental factors that proved critical to the performance, such as different RNA isolation procedures and the assessment of hybridization results, for example, were investigated in detail. Various software tools were developed that permit convenient handling and sound analysis of the large data quantities obtained from transcriptional profiling studies. Comprehensive arrays are being distributed within the European Yeast Functional Analysis Network (EUROFAN) and beyond. © 1998 John Wiley & Sons, Ltd.

KEY WORDS — transcription; microarrays; expression profiling

INTRODUCTION

With completion of the sequence analysis of the entire *Saccharomyces cerevisiae* genome (Goffeau *et al*., 1997), new avenues in yeast analysis were opened up. Due to a sometimes surprisingly high degree of similarity even to evolutionarily distant organisms, such as human (for disease genes see Bassette *et al*., 1996, for example), these analyses will deliver information valuable not only to the understanding of yeast itself but also to the comprehension of biology in eukaryotic cells. In consequence, whole genome functional analyses are now being carried out on a wide range of aspects (e.g. Oliver, 1996; Shoemaker *et al*., 1996; Lockhart *et al*., 1996; Velculescu *et al*., 1997), using tools that also only became available with the acquisition of the complete sequence.

One essential issue in studies on the functioning of an entire organism is the determination of its gene activity and the regulation thereof. Transcriptional analysis by hybridizing complex RNA samples to gene arrays, which so far have been mostly made from anonymous cDNA sequences,

*Correspondence to: Nicole C. Hauser, Molecular-Genetic Genome Analysis Group, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany. Tel: (+49) 6221 42 4683; fax: (+49) 6221 42 4682; e-mail: n.hauser@dkfz-heidelberg.de
Contract/grant sponsor: German Ministry of Research and Technology.
Contract/grant sponsor: European Commission (EUROFAN programme).

CCC 0749–503X/98/131209–13 $17.50

Received 24 February 1998
Accepted 10 May 1998

has provided early indications of the usefulness of the information acquired by this means (e.g., Augenlicht *et al.*, 1991; Höög, 1991; Gress *et al.*, 1992; Nguyen *et al.*, 1995; Schena *et al.*, 1995). With the availability of a complete, non-redundant gene set, however, a new quality level has been reached, based on which even complex analyses can be performed (DeRisi *et al.*, 1997; Wodicka *et al.*, 1997). Studies on yeast will also largely benefit from the exploitation of the large amount of genetic, physiological and biochemical knowledge accumulated by the scientific community prior to the genomics age.

Within the European Functional Analysis Network (EUROFAN), we have embarked on the distribution of arrays that consist of the complete set of yeast genes. The common analytical tool will be one important factor in linking the different, specific fields of interest. On the one hand, data will be directly correlated to the respective type of investigation and assist in its analysis; concurrently, however, data cross-referencing will interrelate the individual aspects towards an understanding of the whole picture. In this article, we describe the basic means upon which these transcriptional analyses are based, and assess the protocols and quality criteria that were found to be critical for successful analyses. Also, suitable computational tools are presented which are prerequisites for such studies.

MATERIALS AND METHODS

Open reading frame PCR-amplification

A set of 12,232 PCR primers was commercially obtained (Research Genetics, USA). The DNA was transferred into microtitre dishes and simultaneously diluted to a concentration of 4 μM using a Biomek-2000 robot (Beckman, USA). The oligomer sequences were designed to bind close to either end of the open reading frames (ORFs). They were about 45 nucleotides in length, of which two common tag-sequences accounted for 20 and 22 nucleotides at the 5′-end, respectively [forward primers: d(GGAATTCCAGCTGACCACCATG); reverse primers: d(GATCCCCGGGAATTGCC ATG)]. PCR was done in microtitre dishes using PTC-200 cyclers (MJ Research, USA). For the amplification of the most 3′-located region of genes, a second primer set was designed according to the specifications of the above set but binding about 500 bp upstream of the respective 3′ end. The oligonucleotides were obtained from Interactiva (Germany).

First-round PCR was done on a template of 25 ng genomic DNA in a total volume of 50 μl of 10 mM-Tris–HCl, pH 8·8, 50 mM-KCl, 1·5 mM-$MgCl_2$, 0·2 mM of each nucleotide triphosphate, a mixture of 1 unit of *Taq*-polymerase and 0·25 unit *Pfu*-polymerase (Stratagene, USA) and both primers at a concentration of 0·4 μM each. Initial denaturation was at 95°C for 3 min, followed by 35 cycles of 1 min at 95°C, 45 s at 50°C and 3·5 min at 72°C; post-treatment at 72°C was for 10 min.

For reactions that failed to yield product under the above conditions, the Expand™ High Fidelity System (Boehringer Mannheim, Germany) was used in a second amplification reaction on 25 ng genomic template DNA under the conditions recommended by the manufacturer. After 2 min denaturation at 94°C, there were 10 cycles of 15 s at 94°C, 30 s at 50°C and 4 min at 68°C, followed by another 26 cycles during which the period at 68°C was extended by 20 s per cycle. Again, there was a final incubation at 72°C to fill-in ends.

Array production

Initially, the PCR-products were transferred to positively charged filters (Pall Biodyne B; DuPont, USA) using 96- or 384-pin gadgets mounted to a BioGrid spotting device (BioRobotics, UK). Flat-bottomed pins of 0·7 mm diameter transferred about 60 nl of liquid per spotting event, while only 15 nl was spotted with pins of 0·4 mm. PCR-products were placed repeatedly onto the same position in order to accumulate DNA. During production, filters were kept humid by being placed onto Whatman 3MM paper soaked in denaturing solution (1·5 M-NaCl, 0·5 M-NaOH).

For the production of higher-density grids, a special pin tool (BioRobotics, UK) was used that produced arrays of densities of up to 750 spots per square centimetre. More recently, DNA was placed onto the surface by a device equipped with an eight-channel array of piezo printing elements (GeSim, Germany). By this technique, samples could be sprayed in multiples of 15 picolitres onto solid support at a grid density of up to 10,000 per centimetre square.

Cell lysis and RNA isolation

Enzymatic lysis Cells from an exponential-phase culture were pelleted, washed twice in water and

once in 1 M-sorbitol, before spheroplasts were produced by incubating 10 OD_{600} of cells in 2 ml 1 M-sorbitol, 0·1 M-EDTA, 14 mM-β-mercaptoethanol and 1 mg/ml Novozym (Sigma, Germany) at 30°C for 10–60 min, shaking carefully. RNA was extracted using the TRIZOL® Reagent (Life Technologies, UK) or RNA-Clean™ solution (AGS, Germany) following the manufacturers' instructions. Since the spheroplasting step causes a stress reaction in the cells prior to lysis, however, other protocols had to be used to circumvent the artificially triggered gene response.

Phenol extraction Yeast cells were grown in a volume of 50 ml to an OD_{600} of 1 and harvested by quick pelleting and direct freezing at −70°C. The pellet was resuspended in 2·5 ml 40 mM-Tris–HCl, pH 7·5, 4 mM-EDTA, 80 mM-LiCl, 4 mM-iodoacetic acid, 8% SDS and 2 ml phenol. After the addition of glass beads (0·5 mm; Braun Melsungen, Germany), the sample was equilibrated at 50°C in a waterbath, stirred on a vortex for 2 min and incubated again at 50°C for 10 min. Cooled down on ice for 20 min, it was spun for 10 min at 4°C. The upper phase was extracted twice with about one volume of chloroform and subsequently ethanol precipitated. The pellet was taken up in 300 ml 0·5 M-NaAc, pH 7·0, and stored frozen after the addition of two volumes of ethanol. Only directly prior to use, the RNA was pelleted and dried, after an additional washing step in 70% (v/v) ethanol.

Mechanical disruption Best results were obtained when the basic features of the above two protocols were merged. Yeast cells were grown to the desired growth phase. They were pelleted for 3 min and resuspended in a very small volume of growth medium. The suspension was sucked into a glass pipette and released as individual drops directly into liquid nitrogen kept in the small Teflon vessel of a micro-dismembrator (Braun Melsungen, Germany). After adding a 7 mm bead made of tungsten carbide, frozen drops equivalent to about 15 OD_{600} units of cells were mechanically broken for 2 min at top speed. The still frozen powder was instantly taken up in TRIZOL® Reagent or RNA-Clean™ solution and treated as instructed.

Probe generation

Probe was generated by a first-strand cDNA synthesis (Nguyen *et al*., 1995). Some 25 to 30 µg of total RNA were mixed with 0·5 µg $(dT)_{15}$ in 11 µl of water. The sample was heated to 70°C for 10 min and subsequently cooled to 43°C. Reverse transcription was performed in a total volume of 30 µl using SuperScript™II (Life Technologies, UK). Incubation was for 1 h at 43°C in the presence of 0·25 mM each of dATP, dGTP and dTTP, 1·66 µM-dCTP and 30 µCi of [α-^{33}P]dCTP (3000 Ci/mmol; Amersham, UK). Subsequently, 1 µl 1% SDS, 1 µl 0·5 M-EDTA and 3 µl 3 M-NaOH were added and the RNA was hydrolysed for 30 min at 65°C and 15 min at room temperature. The solution was then neutralized with 10 µl 1 M-Tris–HCl, pH 8, and 3 µl 2 N-HCl. After the addition of 5 µl 3 M-sodium acetate, pH 5·3, 5 µl tRNA (10 µg/µl) and 60 µl isopropanol, the DNA was precipitated at −20°C for 30 min, pelleted by centrifugation and taken up in 100 µl of water. Alternatively, the unincorporated nucleotides were removed using QIAquick columns (Qiagen, Germany).

Hybridization and image analysis

Probe was denatured at 100°C for 5 min or by adding 1/10 volume of 3 M-NaOH. Arrays were prehybridized for at least 2 h in 5 × SSC, pH 7·5, 5 × Denhardt's solution, 0·5% SDS (Sambrook *et al*., 1989). Probe hybridization was in the same buffer at 65°C for 20 h in a volume of 50 µl per centimetre square, probe concentration being about 10–50 Mcpm. Subsequently, filters were briefly rinsed in 2 × SSC, 0·1% SDS before being washed for 20 min at 65°C in the same buffer and again in 0·2 × SSC, 0·1% SDS at 65°C for 1 h.

Signal detection was performed by a Storm 860 PhosphorImager (Molecular Dynamics, USA) or an adapted CCD camera system (BioImage, USA), which could also be used if the probe had been labelled by incorporation of fluorescence-tagged nucleotides. Images were directly imported into purpose-made software (BioImage HDG Analyser) run on a SUN Ultra workstation, which identified positive spots by matching the actual image with an idealized reference image of the grid. The output files were re-formatted by a self-made software tool and form the basis for the analysis algorithms.

Filter regeneration was done by twice pouring 300 ml of boiling buffer (5 mM-sodium phosphate, pH 7·5, 0·1% SDS) over the filters. The first time, the buffer was immediately cast away, while the second time it was left to cool to room temperature.

Data analysis

Images produced by the PhosphorImager were clear enough to describe the spots merely by the intensity measurement as output by the HDG Analyser. For the further analysis of the resulting greyscale matrix, a software package was written in the high-level mathematical programming language MATLAB (Mathworks Inc., USA) addressing the following issues:

(1) To allow for comparison across filters, spot intensities are normalized by division through the intensities achieved with universal oligonucleotides.
(2) Intensities of spots representing the same ORF are compared in order to judge the reliability of measurement. The variance in the difference of the corresponding intensities is calculated, and spots are not used for the analysis when this difference exceeds four times the standard deviation. For the remaining spots the intensity is defined as the average of the two measurements.
(3) Comparison of the final intensities of different experiments is presented by various means, such as correlation plots or differential screens. Upon clicking on a data point of a plot, the programme provides the information on the corresponding gene.

Western blot analysis

Yeast strains were grown in full synthetic medium (SDC) to an OD_{600} of 0·5–0·7. 10 OD worth of cells were harvested and lysed by vortexing with glass beads in 200 µl SDS buffer and boiling at 95°C. 7·5 µl were separated in a 10% SDS polyacrylamide gel, transferred to nitrocellulose and probed with polyclonal anti-Nsp1p serum EC9–1, which crossreacts with Nup2p (Hurt, 1988).

RESULTS

Array production

For PCR-amplification of the 6116 yeast genes, a primer set from a commercial source (Research Genetics, USA) was used. The reactions were carried out individually in microtitre dishes using as template genomic DNA of yeast strain FY1679 (*MAT***a**/α *ura3–52/ura3–52 trp1D63/TRP1 leu2D1/LEU2 GAL2+/GAL2+*). Each product was checked on agarose gels (Figure 1) for amount and quality. By mixing *Taq*-polymerase with a small amount of *Pfu*-polymerase (Barnes, 1994), a yield of at least 10 µg of DNA was obtained in 90% of the reactions. Fragment length was up to about 6 kb. No apparent correlation between fragment size and yield could be detected. Of the negative samples, another 66% could be recovered by using a commercial kit for long-range PCR, adding up to a total success rate of 97%. These latter samples were up to 11 kb in length. The remaining 3% of unamplified ORFs have an average length of 1·7 kb, which is only slightly more than the 1·4 kb average length of yeast ORFs. No apparent reason for the failure could be detected from the sequence.

In addition to the primers that amplified the complete ORFs, a second set had been designed, priming at a distance of about 500 bp upstream of the 3′ terminus (Figure 1), in order to assess the influence of fragment length on signal intensities. Although, on average, the signal intensities seemed to be slightly stronger on longer fragments, differences were insignificant for the sort of analysis presented here.

Using a robotic device (BioRobotics, UK) equipped with a standard 384-pin tool, the 6116 PCR-products were arrayed on an area equivalent to three microtitre dishes, each fragment being present twice. Using standard spotting protocols, however, ring-structured spots were frequently obtained because of drying effects, which proved difficult to identify by image analysis software. The problem was solved by placing the filters onto a stack of several sheets of 3MM Whatman paper wetted with the denaturing solution that is needed to bind the DNA covalently to the filter. Using the same basic set-up but applying a newly designed micro-pin tool (BioRobotics, UK), all fragments could be placed on an area equivalent to one microtitre dish. Fitting a device based on piezo pipetting technology (GeSim, Germany) for the application of PCR-fragments from 96-well or 384-well plates, respectively, the distances between spots could be reduced significantly and less material had to be transferred due to more compact spots (Figure 2). The high spatial resolution of the device and its reproducible drop size of 15 picolitres enables densities of up to 10,000 spots per centimetre square.

The arrays generated at present for distribution to interested laboratories are in the size-range of microtitre dishes for reasons of convenient handling and because of the fact that such density is compatible with the resolution of standard phosphor- and fluoroimaging devices. Experiments

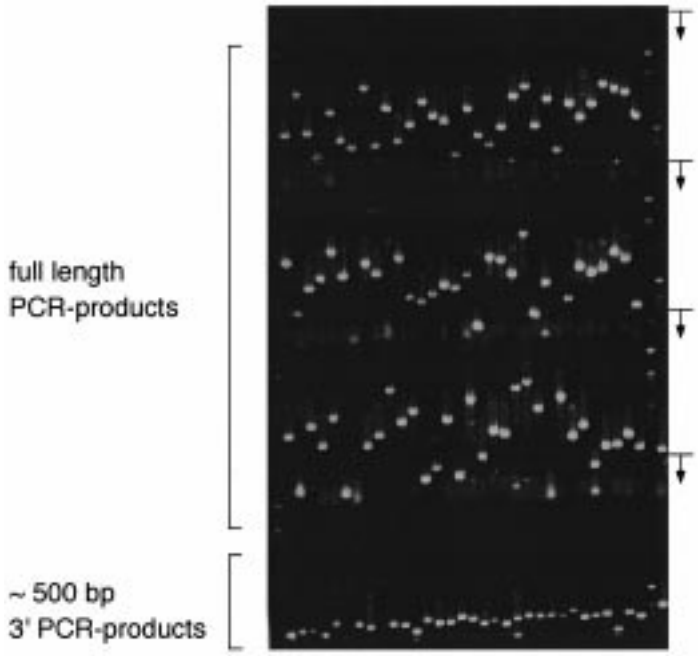

Figure 1. Typical result of a quality check by agarose gel electrophoresis of the PCR-amplification of yeast ORFs. Two primer sets were used for amplification, either priming at the ends of each coding sequence or producing fragments of about 500 bp length from the genes' 3′-termini.

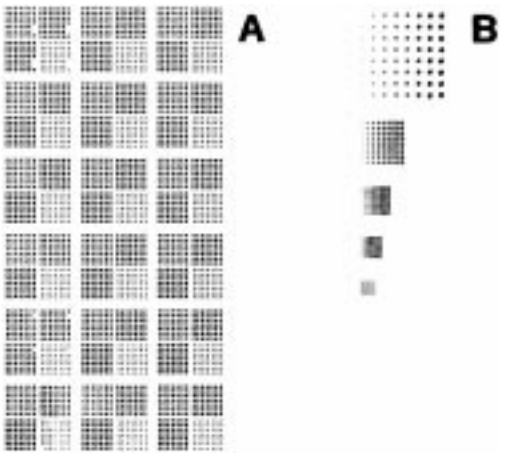

Figure 2. Arrays (original size) produced by a micro-spotting pin tool (A; BioRobotics, UK) and piezo-element robotics (B; GeSim, Germany), respectively, using dye-labelled material. Various spot sizes were chosen. Distances in (B) are (top to bottom) 1500 μm, 750 μm, 500 μm, 350 μm, 250 μm; grids of higher density are possible but cannot be presented in original size. In some instances, spots were placed on top of each other, producing more intense colouring at particular grid positions.

with DNA-chips made of PCR-products that are being bound covalently via their 5′-end to aminated polypropylene are well advanced and will be presented elsewhere (S. Matysiak *et al.*, in preparation).

Filter hybridization

Spotting the PCR-products, no binding of unextended primer molecules occurred. Only amplification products were found attached to the filters, as determined by hybridizing oligonucleotide probes specific to the 5′ tag sequences of the primers. This fact was exploited for quantification of the DNA amounts present at each spot and thus normalization of the signal intensities obtained in subsequent hybridizations with complex probes. Also, the reproducibility of the filter hybridization could be checked by this means. While a sharp reduction of the DNA amount present on the surface could be detected after the first regeneration event, it was found to be constant thereafter (Figure 3). Filters have been re-used in more than 15 hybridizations without significant loss of performance. In the

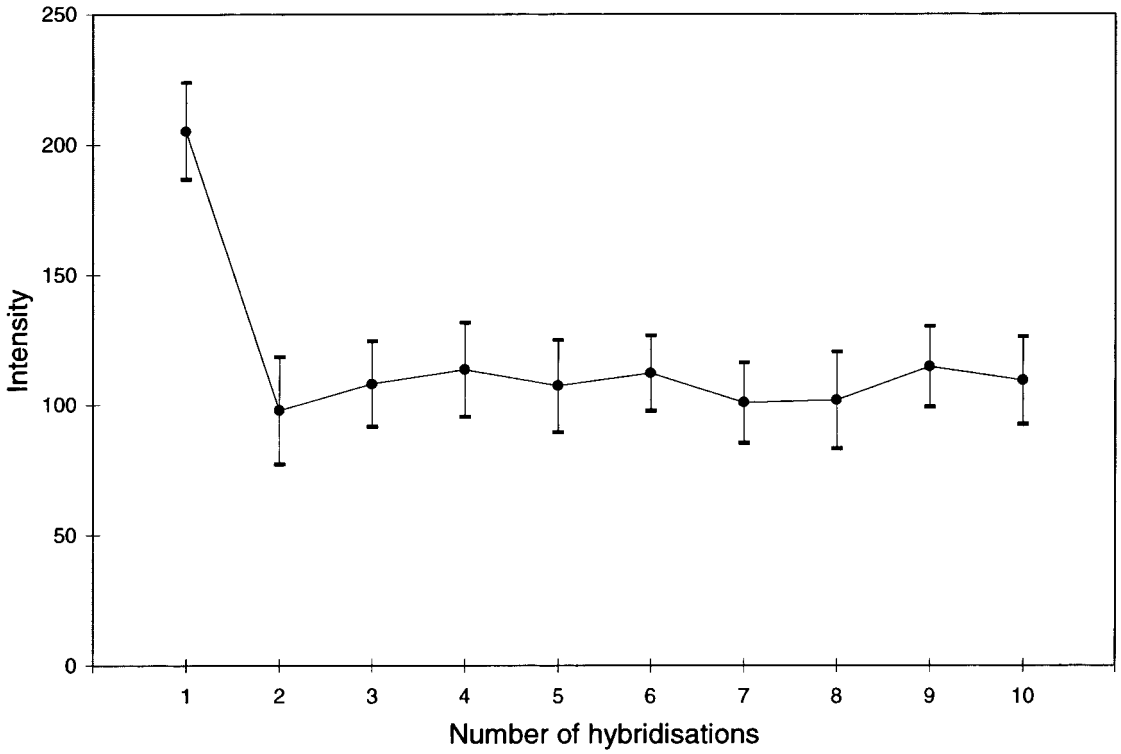

Figure 3. Re-usability of filter arrays. Signal intensities in arbitrary units obtained by hybridizing an oligomer that binds to the common 5′-end of the PCR-products were averaged for 10 different measurements. After the first probe regeneration, signal was reduced by about one half due to a loss of material. Subsequently, however, no significant decrease could be observed.

actual experiments, each filter was subjected to a mock hybridization and regeneration procedure first, before being used in two normalization processes using probe complementary to one primer. Subsequently, the complex RNA–probe hybridizations were taking place, followed by another normalization with the other primer-specific probe for confirmation.

Probe preparation

The method used for RNA isolation was found to be critical for the success of the analysis. Several preparations obtained from different sources yielded insufficient probe, although the quality seemed to be good as judged by OD-measurement and gel analysis. Our own initial protocol relied on the generation of protoplasts prior to the actual RNA extraction with a monophasic solution of phenol and guanidine isothiocyanate. While such treatment provided RNA that worked very well for probe generation by reverse transcription, the technique induced an intracellular stress reaction during the relatively long process of cell wall removal, whereby transcriptional activities were strongly influenced (for an example see Figure 6B).

This effect was sought to be circumvented by freezing the yeast cells immediately after growth and isolating the RNA by phenol and chloroform extractions. This method prevented the artificial transcription but produced RNA of variable quality for reverse transcription. Thus, the basic features of the techniques were merged to form the eventual procedure that proved robust in both aspects, unbiased RNA levels and good probe generation.

The cells were instantly shock-frozen by directly releasing drops of the growth culture into liquid nitrogen, kept frozen during mechanical breakage and only thawed when suspended in the organic solvent. By this method, some 250 µg of RNA were obtained within 2 h from 15 OD_{600} units of cells. For labelling, RNA was reverse transcribed

in the presence of a large excess of oligo-dT primer molecules as described (Nguyen *et al.*, 1995). This procedure is optimized for minimizing the portion of poly-A sequence that is reverse transcribed. With no poly-(A:T) sequences present on the array, this fact is inconsequential with respect to the specificity of hybridization but meaningful nevertheless for a reduction of any bias introduced by potential priming differences caused by a transcript's tail length.

Variations in the effectiveness of the labelling procedure was assessed by adding, as a control, a known amount of mRNA of the rabbit β-globin gene (Life Technologies, UK) to RNA isolates prior to the probe preparation. With the rabbit gene present on an array, trans-effects could be checked for directly.

Assessment of sensitivity, reproducibility and quantification

Sensitivity was evaluated by using genes of known activity. Three different strains—wild type (*NUP2*), deletion (*Δnup2*) and overexpression (*Δnup2*+pRS425NUP2; Loeb *et al.*, 1993)—were investigated with respect to the nucleoporin *NUP2* transcript, as an example. Signal intensities of *NUP2* were quantified in comparison to *ACT1*, which is known to show relatively constant transcription levels over a wide range of conditions.

While, unsurprisingly, no signal could be detected for the deletion strain (even after extended exposure), the transcript rate of *NUP2* increased about 25-fold in the overexpressing strain compared to wild type (Figure 4A). Assuming that the *ACT1* RNA represents about 1% of the total mRNA pool in a yeast cell, transcripts as rare as 0·01% (1 part in 10,000) of the total mRNA could be detected on nylon filters, in accordance with results reported earlier for other systems (e.g. Nguyen *et al.*, 1995). The limitation is mostly due to the relatively high background typical of nylon filters. As determined by covalently binding labelled DNA to a filter prior to hybridization, signals of low intensity were simply submerged by noise (data not presented). This effect can be circumvented by using inert materials such as polypropylene.

For *NUP2*, a comparative measurement of the actual protein level was also done on a Western blot (Figure 4B,C). Probing with a polyclonal anti-Nsp1p serum (Hurt, 1988), which cross-reacts with Nup2p because of the high similarity in the FXFG repeats, protein levels of Nup2p could be determined for the three yeast strains relative to the constantly expressed Nsp1p. The protein levels followed the transcript levels, although the increase in the overexpression strain was not quite so big.

The yeast filter arrays were used several times with the same probe to check for the reproducibility of the transcriptional results. Comparing hybridizations performed with the 5′-tag-specific oligonucleotide sequence, done before and after a set of complex hybridizations, the typical correlation coefficient was found to be 0·99 (Figure 5A), indicating that the amount of DNA (and thus signal intensity) at each spot remained constant over this experimental setting. Moreover, these data were used for normalizing the DNA amounts present at each spot. When data sets from actual transcriptional analyses with RNA from the same yeast samples were measured in different experiments (e.g. Figure 5B), a correlation coefficient of 0·97 was typically obtained, demonstrating the high degree of reproducibility of even complex data in duplicate experiments. Nevertheless, the experimental variation was such that the average of at least two identical assays was taken into account in actual analyses, for increased accuracy.

The signal intensities produced by any given RNA sample often varied more than could be covered by the linear range of a phosphorimager device, with some signals being either below the threshold of detection or above the linear range. This was overcome by exposing arrays twice or, in some instances, even three times for different periods. A software routine was designed that merges the two data-sets into one, thus increasing the linear range of detection *in silico*.

Data production and dissemination

With the protocols described above, analyses were carried out under various growth conditions, especially focusing on stress. Typical results of such analyses are shown in Figure 6, again, for better clarity, only for a chromosomal subset of ORFs.

Transcription rates of yeast cells grown in 'standard' culture, YPD medium at 30°C, were compared to those of cells of a second culture subjected to oxidative stress by the addition of 0·5 mM-hydrogen peroxide (Figure 6A; Moradas-Ferreira *et al.*, 1996). The other panel (Figure 6B) shows a comparison between two RNA preparations

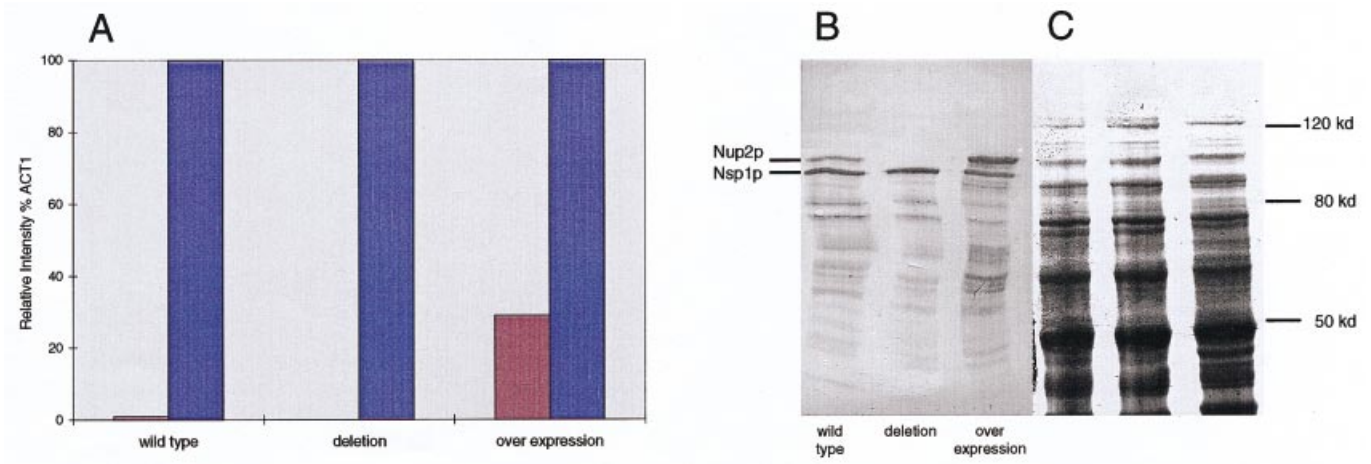

Figure 4. Sensitivity and quantification. Probe was made from yeast grown in synthetic medium (SDC for wild type and deletion strain, SDC-leu for the overexpressing strain) and hybridized to filter arrays. (A) Signal intensities obtained on the actin gene *ACT1* (blue) and the nucleoporin gene *NUP2* (red) are shown. Left, the results with wild-type yeast are presented, with *NUP2* being expressed at a level of about 1% that of actin. With mutant strains in which the nucleoporin gene was deleted (middle) or overexpressed due to the high copy-number plasmid pRS425NUP2 (right; Loeb *et al.*, 1993), the intensities accurately reflected the transcription ratios as determined on Northern blots (data not presented). Concomitantly to RNA levels, the actual amount of protein expressed in the respective yeast strains was determined by Western blotting. Signals were obtained from detection with an antibody (B) and Coomassie blue staining (C), the latter demonstrating that equivalent amounts of protein were loaded onto each lane.

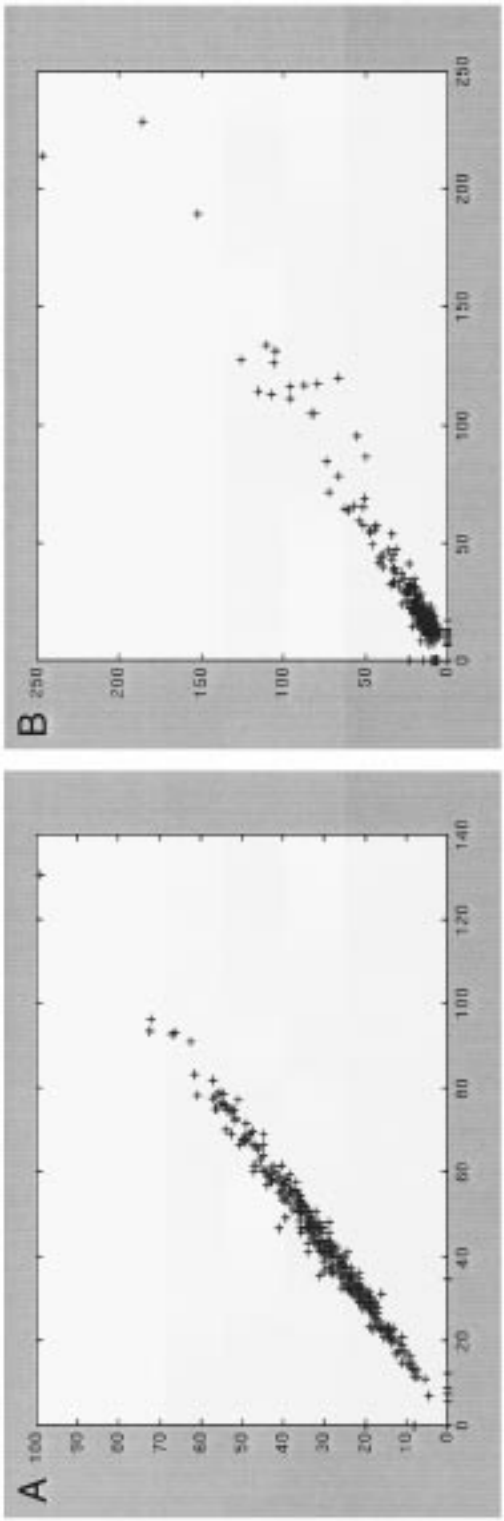

Figure 5. Reproducibility of analysis. Here, the results of the 272 ORFs of chromosome VIII are shown; signal intensities are given in arbitrary units. Each cross represents an ORF; the signal intensity in either experiment is indicated by its location within the two-dimensional matrix. (A) Two different arrays were hybridized with an oligonucleotide probe specific to the common 5′-end of the PCR-products, being the 6th and 12th experiments, respectively, that were carried out with these filters. Correlation between both experiments was high (correlation coefficient 0·996). (B) Two complex probes made from total RNA that had been harvested from identical culture conditions were hybridized. Although slightly less reproducible than the experiments with oligomer probes, correlation between the resulting signals was still very high, with a correlation factor of 0·977.

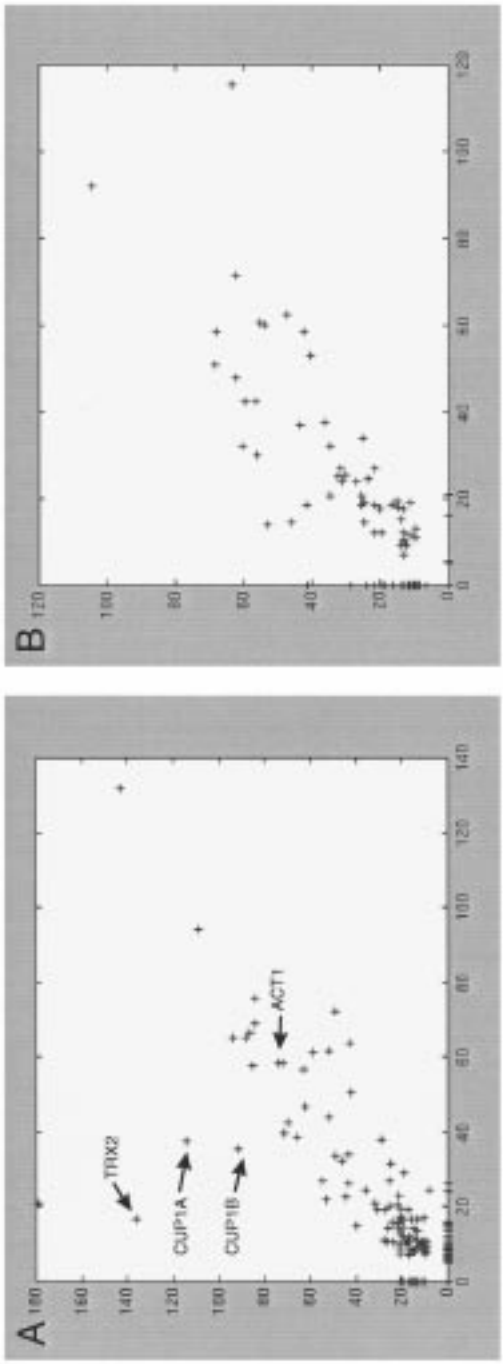

Figure 6. Comparison of transcription rates. RNA was isolated from standard yeast culture (signal intensities along the *x*-axis in both panels) and cells grown in the presence of hydrogen peroxide (*y*-axis in A) or isolated by 'enzymatic lysis' of the yeast cells (*y*-axis in B); signal intensities are given in arbitrary units. The relative transcription rates obtained for a representative gene subset are shown, demonstrating the variation in transcription caused by oxidative or preparative stress, respectively. The arrows in (A) point to some specific transcripts as indicated.

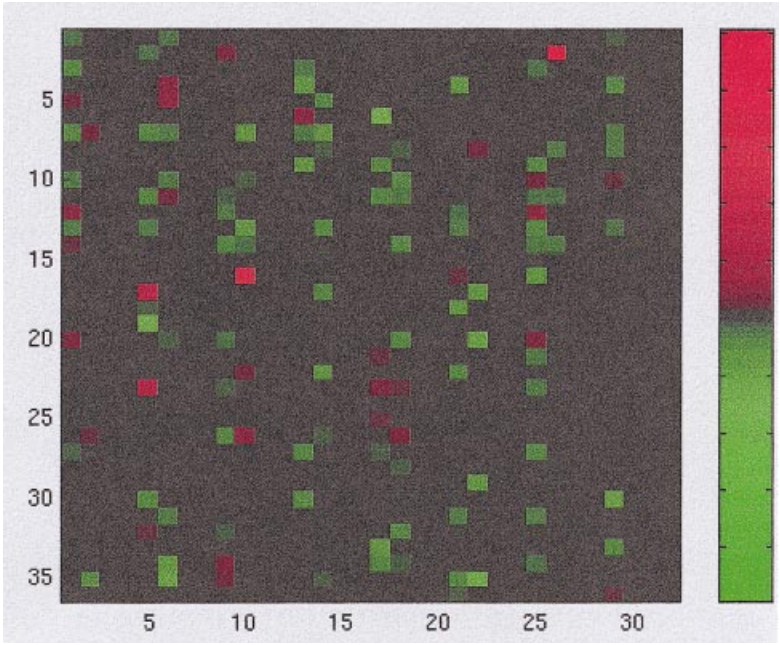

Figure 7. Differential screening plot of the results shown in Figure 6A; the degree of overexpression under oxidative stress (red) or at standard conditions (green) is indicated by the intensity of the colouring.

from standard growth condition, one done by the 'enzymatic lysis' protocol, which includes the generation of spheroplasts, the other the result of the 'mechanical disruption' procedure (see Materials and Methods).

The results can be presented in two formats: in two-dimensional plots (Figure 6) the entire data set from the relevant exposure is shown. On screen, a coordinate can be clicked upon, providing further information on the relevant ORF. Alternatively, a differential screening plot indicates only the significant transcriptional differences by both colour and intensity (Figure 7).

While the former format is a superior presentation in terms of overall quality assessment, the latter is a more user-friendly tool. From Figure 6, it can be seen that, by causing oxidative stress, significant changes in the transcription of various genes occurred. Prominent for the presented ORFs was an increase of the transcription rate of *TRX2* (Thioredoxin II) by a factor of 7, for example. This effect, and many other results generated during various studies, were confirmed by inverse format experiments on Northern blots. In all cases, the results obtained on arrays corresponded well to the signals from the Northern blot assays (results not shown).

The large amount of data produced by even a single experiment on the complete set of yeast genes prevents a broad and simultaneously detailed discussion within this or any other manuscript, even more so because it is rather likely that additional data from other experimental approaches are needed for a real understanding of many effects. While particular findings will be presented in future publications, the majority of our data, such as the above, are made available to the public via our web-site (http://www.dkfz-heidelberg.de/abth0701/index.htm) for further interpretation.

DISCUSSION

A system is described that allows a detailed analysis of transcriptional activity in yeast cells. While the technical principle is not new, the various aspects that influence such analysis had to be investigated thoroughly to warrant sensible data acquisition. Surprisingly to us, the method of isolating the RNA was one of the major issues that had to be addressed, especially with respect to reproducibility. Data dissemination was, and to some extent still is, another problem. Software tools are essential to deal with the amount of data

even at the initial, relatively low level of mere transcriptional profiling. It will become even more crucial with increasing complexity.

Only complex functional investigations at several levels and their interrelation will provide the information that is necessary for a more comprehensive grasp of cellular activity and regulation. A comparison of transcriptional activity and actual protein expression (e.g. Figure 4) or promoter activity (see Niedenthal *et al.*, 1996) is an obvious and important step toward this end. Beyond this, the transcriptional analysis has to be knitted tightly into a network of more biochemically and physiologically oriented studies that provide the added value for better interpretation. This is only feasible, however, by linking information from a network of laboratories with different areas of expertise via commonly used experimental tools, thus generating data of a matching type. For this reason, we currently distribute the arrays in a format that can be handled in any laboratory. Widespread dissemination of appropriate software packages will be a more difficult process, also because the specific requirements, apart from basics as outlined here, might vary considerably between different laboratories.

For the preparation of microarrays, we have concentrated on the use of aminated polypropylene as a support medium. Amination was performed in our laboratory by a new technique yielding a high density of active amino groups present on the sheets (S. Matysiak *et al.*, in preparation). Covalent bonding of DNA is being performed via a functional group at the 5′ termini. Thus, the existing PCR-products had to be re-amplified using a pair of appropriately modified oligonucleotide primers binding to the common 5′ tag-sequences. This re-amplification of the entire set of 6116 ORFs is currently under way at the conditions that yielded the initial amplicons, with the success rate of the secondary amplification being identical to that of the primary PCR.

While a re-use of arrays seems to be a non-issue for medical applications, it represents an important advantage for the sort of analysis described here. Beside reduced cost, the re-use of identical arrays significantly improves the accuracy of quantitative studies on small transcriptional differences. The polypropylene surface will also make the use of fluorescent dye labelling possible. Preliminary analyses on microarrays indicate that a sensitivity of around a single mRNA molecule per cell seems within range. The length of the PCR-products attached to the support could also be of significant influence to the signal intensities (S. Matysiak *et al.*, in preparation). As an alternative way of probe labelling, work is still ongoing on the attachment of fluorescence dyes directly to the mRNA molecules specifically *via* their CAP-function, thereby avoiding potential bias caused by the reverse transcription process.

Technically improved analysis will be possible by using peptide nucleic acids (PNAs) as substrate on the arrays (Weiler *et al.*, 1997, 1998) rather than PCR-products. For yeast, we are currently working on a comprehensive set of PNA oligomers. Several features of PNA foster superior results: duplex stability of PNA:DNA or PNA:RNA hybrids is high, with dissociation temperatures of 16-mer sequences being in a range close to those of the PCR-products; nevertheless, PNA oligomers exhibit a higher selectivity than DNA oligonucleotides, let alone PCR-products; probe accessibility is better, since intramolecular folding of the probe is diminished due to very low ion concentration in the hybridization buffer, which is made possible because of the uncharged nature of PNA; and last, PNA invades double-stranded nucleic acids by replacing one strand while binding to its complementary sequence.

Any oligomer approach, however, will only be possible if the sequence of an organism's genome or at least its genes is known. Analysis based on PCR-products could be practicable and beneficial even in the long run. For micro-organisms, for example, a minimal clone tiling path could be selected from the results of genomic template mapping, a technique originally developed for ordered, low-redundancy DNA sequencing (Johnston *et al.*, 1997; Scholler *et al.*, 1998). Because of the short intergenic stretches, a microbial genomic clone map also represents a complete, dense and perfectly normalized gene inventory. RNA profiling experiments on such a substrate—placed onto arrays as PCR-products of the mapped but otherwise anonymous genomic clones—could render the sequencing of an entire microbial genome unnecessary. Only genes and related genomic regions that exhibit an interesting response during specific experiments would be chosen for analysis. By this selective approach, functional studies could well precede sequencing with only a few tens of thousands of base pairs of high potential having to be sequenced rather than millions of bases of unqualified importance.

ACKNOWLEDGEMENTS

We thank Kerstin Hartlieb for her skilful technical assistance during primer dilution and PCR-amplification. This work was funded by the German Ministry of Research and Technology (BMBF) and the European Commission as part of the German Yeast Functional Analysis Network and the EUROFAN programme, respectively. Financial support for travelling provided by INTAS is also gratefully acknowledged.

REFERENCES

Augenlicht, L. H., Taylor, J., Anderson, L. and Lipkin, M. (1991). Patterns of gene expression that characterise the colonic mucosa in patients at genetic risk for colonic cancer. *Proc. Natl. Acad. Sci. USA* **88,** 3286–3289.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91,** 2216–2220.

Bassette Jr., D. E., Boguski, M. S. and Hieter, P. (1996). Yeast genes and human disease. *Nature* **379,** 589–599.

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.

Goffeau, A., *et al*. (1997). The yeast genome directory. *Nature* **387** (Suppl.), 1–105.

Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G. and Lehrach, H. (1992). Hybridization fingerprinting of high density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm. Genomes* **3,** 609–619.

Höög, C. (1991). Isolation of large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucl. Acids Res.* **19,** 6123–6127.

Hurt, E. C. (1988). A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae*. *EMBO J.* **7,** 4323–4334.

Johnston, M., *et al*. (1997). The complete sequence of *Saccharomyces cerevisiae* chromosome XII. *Nature.* **387** (Suppl.), 87–90.

Lockhart, D. L., Dong, H., Byrne, M. C., *et al*. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14,** 1675–1680.

Loeb, J. D. J., Davis, L. I. and Fink, G. R. (1993). NUP2, a novel yeast nucleoporin, has functional overlap with other proteins of the nuclear pore complex. *Mol. Biol. Cell* **4,** 209–222.

Moradas-Ferreira, P., Costa, V., Piper, P. and Mager, W. (1996). The molecular defences against reactive oxygen species in yeast. *Mol. Microbiol.* **19,** 651–658.

Nguyen, C., Rocha, D., Granjeaud, S., *et al*. (1995). Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* **29,** 207–216.

Niedenthal, R. K., Riles, L., Johnston, M. and Hegemann, J. H. (1996). Green fluorescent protein as a marker for gene expression and subcellular localisation in budding yeast. *Yeast* **12,** 773–786.

Oliver, S. G. (1996). From DNA sequence to biological function. *Nature* **379,** 597–600.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

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.

Scholler, P., Heber, S. and Hoheisel, J. D. (1998). Optimisation and automation of fluorescence-based DNA hybridization for high-throughput clone mapping. *Electrophoresis*, in press.

Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. and Davies, R. W. (1996). Quantitative phenotypic analysis of yeast deletion mutants using highly parallel molecular bar-coding strategy. *Nature Genet*. **14,** 450–456.

Velculescu, V. E., Zhang, L., Zhou, W., *et al*. (1997). Characterisation of the yeast transcriptome. *Cell* **88,** 243–251.

Weiler, J., Gausepohl, H., Hauser, N., Jensen, O. N. and Hoheisel, J. D. (1997). Hybridization based DNA screening on peptide nucleic acid (PNA) oligonucleotide arrays. *Nucl. Acids Res.* **25,** 2792–2799.

Weiler, J., Matysiak, S., Gausepohl, H. and Hoheisel, J. D. (1998). New developments in hybridization based DNA screening on oligomer arrays. In Epton, R. (Ed.), *Solid Phase Synthesis—Peptides, Proteins and Nucleic Acids*. Mayflower Worldwide, in press.

Wodicka, L., Dong, H., Mittmann, M., Ho, M.-H. and Lockhart, D. J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae. Nature Biotechnol.* **15,** 1359–1367.

