Chapter 2

Identification of Transcription Factor Targets by Phenotypic Activation and Microarray Expression Profiling in Yeast

Gordon Chua

Summary

A major obstacle to identify physiological transcriptional targets is that the conditions that induce the majority of yeast transcription factors (TFs) are unknown. Microarray analyses of deletion mutants indicate that most TFs are inactive under standard growth conditions. To overcome this, we screened an ordered array of yeast open reading frames (ORFs) to identify TFs that confer reduced fitness upon overexpression, suggesting that overexpression results in an activated state (phenotypic activation). Approximately one-third of all yeast TFs exhibited this phenotype. Here, we describe in detail our methodology to characterize these TF overexpression strains including microarray expression profiling, data analysis, and motif searching. Our analyses show that in many cases, the differentially regulated genes correspond to physiological functions and known targets of well-characterized TFs. The expected binding sites of several TFs were also identified in the promoters of these genes. Moreover, novel DNA-binding sequences and putative targets were identified for less-characterized TFs. These results demonstrate that phenotypic activation is an effective approach to rapidly characterize TFs on a large scale, which should also be feasible in other organisms.

Key words: Transcription factor, Overexpression, Phenotypic activation, Yeast, Microarray, FunSpec, RankMotif

1. Introduction

A primary objective to understand the workings of an organism is to obtain a complete mapping of the transcriptional-regulatory network, defined here as interactions between DNA-binding transcription factors (TFs) and the cognate sequences they bind in order to control expression of target genes. One straightforward

I. Stagljar (ed.), *Yeast Functional Genomics and Proteomics*, Methods in Molecular Biology, vol. 548
DOI: 10.1007/978-1-59745-540-4_2,

approach to systematically characterize TF activities would be to microarray-profile all the TF deletion strains in an organism, where differentially downregulated or upregulated genes in the mutants would potentially represent target genes of transcriptional activators and repressors, respectively. However, we discovered that a large majority of yeast TF mutants (~85%) exhibit profiles indistinguishable from microarray noise when grown typically in the laboratory, suggesting that most TFs are not active under these conditions, or that there exists a high degree of functional overlap among TFs *(1)*. To overcome these obstacles, we sought for alternate approaches to systematically characterize TFs and identify their target genes. One such approach stemmed from the construction of an overexpression array of 5,280 yeast strains, each containing a unique gene under control of the strong, inducible *GAL1/10* promoter *(2)*. Phenotypic characterization of the overexpression array revealed that 769 genes (15% of the genome) were detrimental to normal growth when ectopically expressed *(2)*. Interestingly, these toxic overexpressors were most enriched for TFs (32.6% of all TFs), representing a greater than twofold enrichment for this functional class of proteins compared to the rest of the genome *(2)*. On the basis of these observations, we hypothesized that the reduced growth rate is attributed to induction of TF activity by ectopic expression (hence the term "phenotypic activation"; *see* **Note 1**) *(3)*.

Differential gene expression caused by phenotypic activation of TFs is globally detected by DNA microarrays to elucidate the biological function of TFs and identify putative gene targets (**Fig. 1**). Strains containing either a 2-μm vector with a *GAL1/10*-driven TF gene or an empty vector are grown concurrently and induced for 3 h after a raffinose to galactose shift. Total RNA is extracted from these strains using hot phenol, and poly-A mRNA is subsequently isolated using oligo-dT cellulose. To couple Cy3 and Cy5 dyes to the samples, the mRNA is initially labeled with aminoallyl-dUTPs during reverse transcription. Cy3/Cy5-coupled cDNA samples are hybridized in dye reversal onto a spotted microarray containing 60-mer oligonucleotide probes specific for all coding genes of *Saccharomyces cerevisiae (4)*.

The microarray data is subjected to an analysis pipeline designed to filter out microarray noise and elucidate TF function and putative gene targets. We determined that Pearson correlations >0.3 between experimental replicates are indicative that the differentially expressed genes are more likely due to TF overexpression rather than microarray noise *(5)*. The differentially regulated genes are also examined for significant enrichment in Gene Ontology functional categories using FunSpec to identify putative function and targets of TFs *(6)*. Finally, the promoter regions of differentially expressed genes are subjected

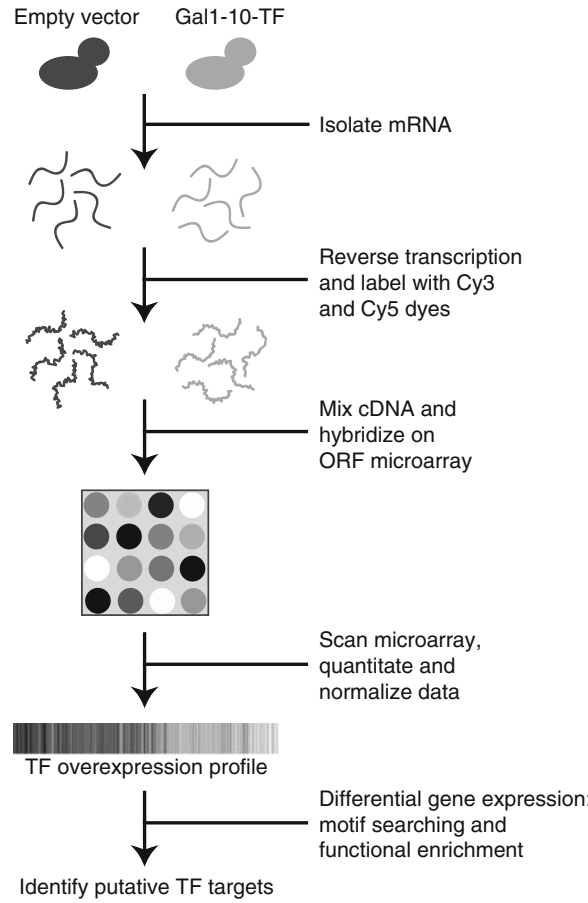

Fig. 1. Flow chart describing the procedure involved in the microarray profiling of transcription factor overexpression strains and data analysis to identify putative targets.

to a motif-finding algorithm called RankMotif to identify putative TF-binding sites *(3)*.

2. Materials

2.1. Cell Culture

1. Synthetic uracil dropout medium with 2% raffinose (SR-Ura).
2. 40% Galactose.
3. 250-ml baffled culture flasks.
4. 50-ml Falcon tubes.
5. Liquid nitrogen.
6. Spectrophotometer (microplate reader).

2.2. Total RNA Isolation

1. AE buffer: 50 mM sodium acetate, pH 5.2, 10 mM ethylenediamine tetraacetic acid (EDTA), 1% sodium dodecyl sulfate (SDS), diethylpyrocarbonate-treated (DEPC) H_2O.
2. Unbuffered liquefied acid phenol (pH = 4.5–5.5) (this is extremely toxic and corrosive and care should be taken not to receive exposure). Store at 4°C.
3. Phenol: chloroform: isoamyl alcohol (25:24:1) (this is extremely toxic and corrosive and care should be taken not to receive exposure). Store at 4°C.
4. Chloroform: isoamyl alcohol (24:1). Store at 4°C.
5. 3 M sodium acetate, pH 5.2, DEPC H_2O.
6. 95–100% Ethanol.
7. 70% Ethanol.
8. Isopropanol.
9. 0.1% DEPC H_2O.
10. Acid-washed 425–600 μm glass beads (Sigma).
11. Plasticware: 10-ml disposal pipettes, 15-ml Falcon tubes.
12. Multi-tube vortexer (VWR).

2.3. Poly-A mRNA Isolation

1. 2× Loading buffer: 40 mM Tris–HCl, pH 7.6; 1 M NaCl; 2 mM EDTA; 0.2% sodium lauryl sarosine (SLS); DEPC water.
2. Middle wash buffer: 20 mM Tris–HCl, pH 7.6; 150 mM NaCl; 1 mM EDTA; 0.1% SLS; DEPC water.
3. Elution buffer: 10 mM Tris–HCl, pH 7.6; 0.1 mM EDTA; DEPC water.
4. Oligo-dT cellulose (SIGMA): This can be recycled by washing the resin with 0.1 N NaOH, which will elute and/or degrade any attached RNA. Store at 4°C.
5. Poly-prep columns (Bio-Rad).
6. 95–100% Ethanol.
7. 0.1 N NaOH.
8. DEPC H_2O.
9. Linear acrylamide (Ambion). Store at –20°C.
10. 3 M sodium acetate, pH 5.2, DEPC water.
11. Plasticware: 10-ml disposal pipettes, 6-ml and 50-ml Falcon tubes.
12. Nutator.

2.4. Reverse Transcription and Aminoallyl Labeling of mRNA

1. PCR purification columns and associated reagents (Qiagen).
2. T18-VN primer (1 μg/μl).

3. Superscript II reverse transcriptase with 5× buffer and 0.1 M dithiothreitol (Invitrogen).
4. 10 mM dNTPs.
5. 1 mM aminoallyl-dUTP (Sigma): Dissolve 1 mg of AA-dUTP in 191 µl DEPC water. Store at –20°C.
6. 1 N NaOH.
7. 0.5 M EDTA.
8. 1 M Tris–HCl, pH 7.6.
9. 80% Ethanol.
10. DEPC H_2O.
11. 42 and 65°C water baths.
12. Speed-Vac.

2.5. Cy3/Cy5 Dye Coupling of cDNA

1. Anhydrous dimethyl sulfoxide (DMSO, Sigma).
2. 2× Bicarb buffer: (one bicarbonate capsule (Sigma), 25 ml water, 125 µl 37% HCl). Note: this buffer should be made fresh and is good only for 1 week.
3. Cy3 and Cy5 monofunctional dye packs (GE Healthcare).
4. 4 M hydroxylamine.
5. PCR-purification columns and associated reagents (Qiagen).

2.6. Microarray Hybridization

1. Quackenbush prehybridization solution: 5× SSC, 1% SDS, and 1% BSA. Store in 50-ml aliquots at –20°C.
2. Hybridization solution: 5× SSC, 25% formamide, 0.1% SDS. Store in 1-ml aliquots at –20°C.
3. Wash 1 solution: 1× SSC, 0.2% SDS.
4. Wash 2 solution: 0.1× SSC, 0.2% SDS.
5. Wash 3 solution: 0.1× SSC.
6. Microscope slide mailers.
7. Lifter cover slips.
8. Curved forceps.
9. UV crosslinker.
10. Speed-Vac.
11. Rack for microscope slides.
12. Hybridization oven (42°C).
13. Hybridization chambers.
14. 250-ml glass dishes.
15. Compressed air (e.g., Dust-off).

2.7. Acquisition, Processing, and Analysis of Microarray Data

1. Laser microarray scanner with integrated software for quantitation and normalization of data (e.g., Axon GenePix 4200A scanner with GenePix Pro/Acuity software – Molecular Devices).

3. Methods

3.1. Cell Culture

1. For each experiment, prepare two overnight cultures of isogenic strains, one containing a 2-μm vector with a *GAL1/10*-driven TF gene and the other containing an empty vector (pEGH, *URA3*$^+$). Inoculate several medium-size colonies (~10) from an agar plate into 100 ml SR-Ura medium in a 250-ml baffled culture flask. In addition, set up a second diluted culture (1/5–1/10×) to ensure that one of the overnight cultures will be in the appropriate cell density range the following day. Shake cultures at 30°C overnight.
2. For each strain, select an overnight culture in mid-log phase (<5 × 10^6 cells/ml) and inoculate a new 95-ml culture in SR-Ura medium at an initial cell density of 7 × 10^5–9 × 10^5 cells/ml.
3. Add 5 ml 40% galactose (2%) into each culture and shake at 30°C for 3 h (*see* **Note 2**).
4. Just prior to 3 h in galactose, determine the optical density at 600 nm (OD_{600}) of each culture. The cell densities should range between 2 × 10^6 and 3 × 10^6 cells/ml. Pair up the empty vector and TF overexpression cultures with the closest cell densities.
5. Harvest cultures in 50-ml Falcon tubes by centrifugation at 2,000 × *g* for 2 min, pour out the supernatant, and freeze cell pellets in liquid nitrogen.
6. Store cell pellets at –80°C.

3.2. Total RNA Isolation

1. Preparation prior to isolation procedure: Preheat AE buffer and acid phenol in a 65°C water bath. For each sample, label three 15-ml Falcon tubes.
2. Remove frozen cell pellets from the freezer (two 50-ml Falcon tubes/100 ml culture).
3. Combine the samples by adding 4 ml AE buffer (65°C) to one of the frozen cell pellets, loosening the frozen pellet by shaking gently or pipetting and pouring the contents into the other Falcon tube. Add 4 ml of acid phenol (65°C) and

~200 µl acid-washed glass beads. **Steps 2** and **3** should be performed quickly to prevent RNA degradation.

4. Vortex the samples in a multitube vortexer for 30 s and place in a 65°C water bath for 4 min. Repeat four times and then incubate 10 min on ice.
5. Centrifuge at 2,000 × *g* for 5 min at 4°C and transfer the aqueous layer into a 15-ml Falcon tube containing 2 ml cold phenol: chloroform: isoamyl alcohol (25:24:1). Vortex three times for 30 s with 1 min intervals in between.
6. Centrifuge at 2,000 × *g* for 5 min at 4°C and transfer the aqueous layer into a 15-ml Falcon tube containing 2 ml chloroform: isoamyl alcohol (24:1). Vortex once for 15 s.
7. Centrifuge at 2,000 × *g* for 5 min at 4°C and transfer the aqueous layer into an empty 15-ml Falcon tube. Add 1/10 volume of 3 M sodium acetate (pH 5.2) (300–400 µl), vortex for 1 min, and then add an equal volume of isopropanol (3–4 ml). Mix well by gently inverting the tubes, and incubate for 10 min at room temperature or –20°C overnight.
8. Centrifuge at 3,200 × *g* for 30 min at 4°C, remove the supernatant and wash the pellet with 1 ml 70% ethanol. Centrifuge at 3,200 × *g* for 5 min at 4°C, remove the supernatant, and air-dry the pellet for 15–20 min.
9. Dissolve the RNA pellet in 1 ml of DEPC water for each sample by pipetting. The samples may be heated at 65°C for 5–10 min to fully dissolve the RNA pellet.
10. Determine the quantity of RNA at OD_{260}. Typical total RNA yield for a 100-ml yeast culture ranges from 500 to 2,000 µg. Store total RNA in –80°C, or continue with poly-A mRNA isolation.

3.3. Poly-A mRNA Isolation

1. Preparation prior to isolation procedure: Remove total RNAs out from freezer and thaw on ice. Save a 10-µl aliquot of each RNA sample for an EPPS/formaldehyde gel to check the integrity of RNA (optional). For each sample, label one poly-prep column, two 6-ml Falcon tubes, and a 1.5-ml microcentrifuge tube. Preheat the elution buffer at 65°C.
2. Preparing columns: Place columns on a wash rack (e.g., Microcentrifuge Tube Rack, ResMer Resin, VWR) with a reservoir underneath (e.g., the lid of a P1000 tip box). Weigh out 0.7 g oligo-dT cellulose (SIGMA) and place in a 50-ml Falcon tube. This amount of oligo-dT cellulose is sufficient for 12–14 columns. Wash the oligo-dT cellulose three times with 50 ml DEPC water and once with 50 ml 0.1 N NaOH. Resuspend the oligo-dT cellulose in 50 ml 0.1 N NaOH.

Mix on Nutator for 10–15 min. Mix the oligo-dT cellulose slurry by inverting the Falcon tube, and quickly add 4 ml slurry into each column using a 10-ml plastic pipette. After the slurry has settled, use the remaining slurry to equalize the volume of oligo-dT cellulose among the columns. Wash the columns by adding 4 ml DEPC water, and then with 2 ml 1× column loading buffer after the water has flowed through the column.

3. Heat the samples to 65°C for 5 min in the water bath. Cool quickly by chilling on ice for 3 min.
4. While samples are cooling, add an equal volume (1 ml) of 2× column loading buffer.
5. Place each column into a 6-ml Falcon tube while inserting a P2 pipette tip between the column and the tube to allow venting and flow through of the sample. Stand the Falcon tubes with columns in a rack.
6. Carefully pour the samples into the columns. The sample should flow through in 2–5 min. If it takes much longer than this, pipette gently to resuspend the resin and let it resettle – the column should then run a bit faster.
7. Reload the columns by pouring the sample from the Falcon tube back into the top of each column, and reinsert the column into the Falcon tube as in **step 5**.
8. Reload the columns again for the third time (repeat **step 7**).
9. After the samples have run through the columns, store them in the –20°C freezer. They can be retrieved later if the mRNA yield is very low.
10. Place the columns into the wash rack. Wash twice with 2 ml 1× column wash buffer.
11. Wash once with 400 µl middle wash buffer.
12. Place the column into a new 6-ml Falcon tube with a P2 pipette tip as described before.
13. Elute three times with 330 µl 65°C prewarmed elution buffer.

A second column run is required to further purify the poly-A mRNA samples.

14. Place columns on the wash rack and wash once with 4 ml DEPC water and then 4 ml 1× column wash buffer.
15. Repeat **steps 3–11**.
16. Place each column into a 1.5-ml microcentrifuge tube with a P2 pipette tip as described before.
17. Elute twice with 250 µl 65°C prewarmed elution buffer.

18. Add 50 µl 3 M sodium acetate (pH 5.2) and 6 µl linear acrylamide into each sample. Vortex the samples for 30 s. Add 1.1 ml of 95% EtOH and vortex the microcentrifuge tube on its side for 30 s. Precipitate mRNA in –20°C overnight.
19. Centrifuge mRNA samples at 16,000 × *g* at 4°C for 30 min. Remove supernatant by pipetting, pulse-centrifuge, and remove the remaining supernatant. Do not wash the samples. Place Kimwipe over the tubes and air-dry samples at room temperature for 30 min. Resuspend in 20 µl DEPC water by pipetting.
20. To remove traces of oligo-dT cellulose that may interfere with subsequent steps, heat the samples at 65°C for 5–10 min, centrifuge at 16,000 × *g* for 1 min, and carefully transfer the supernatant into a new microcentrifuge tube.
21. Determine the quantity of RNA at OD_{260}. Typical total poly-A mRNA yield ranges from 0.5 to >20 µg. Store mRNA in –80°C.
22. Recycling the oligo-dT cellulose and column: Cap the bottom of the columns and add 2 ml of 0.1 N NaOH to each column. Resuspend the oligo-dT cellulose in each column by pipetting, and transfer contents into a 50-ml Falcon tube. Add 1 ml of 0.1 N NaOH to the column, resuspend, and transfer remaining oligo-dT cellulose into the Falcon tube. It is important to remove as much residual oligo-dT cellulose from the column to prevent blockage in future mRNA isolations. Wash oligo-dT cellulose three times with DEPC water. After the final wash, add 50 ml DEPC water and store at 4°C. Rinse the columns twice with DEPC water before storage.

3.4. Reverse Transcription and Aminoallyl Labeling of cDNA

1. Each experiment will be performed twice with fluor reversal (control sample–Cy3 vs. experimental sample–Cy5, and experimental sample–Cy3 vs. control sample–Cy5). As a result, each mRNA sample will be reverse-transcribed twice. Aliquot equal amounts (1–2 µg) of control and experimental mRNA samples into microcentrifuge tubes. Dry down samples in a Speed-Vac (medium-heat setting to prevent RNA hydrolysis).
2. Dissolve each sample in 10.5 µl DEPC water and 1 µl T18VN primer (1 µg/µl).
3. Denature the samples at 65°C for 5 min.
4. Incubate at 42°C for 5 min to anneal the T18VN primer.
5. While the samples are incubating at 42°C, prepare a master reaction mixture containing 4 µl 5× RT buffer, 2 µl 0.1 M DTT, 1 µl 10 mM dNTPs, 1 µl 1 mM aminoallyl-dUTP, and

0.5 µl superscript II enzyme per sample. Add 8.5 µl of reaction mixture to each sample.

6. Incubate 50 min at 42°C.
7. Add 10 µl of a 1:1 mixture 0.5 M EDTA: 1 N NaOH to each of the samples.
8. Incubate at 65°C for 20 min to hydrolyze the RNA. During this incubation, set up and label a QIAGEN purification column and microcentrifuge tube for each sample, and preheat some water at 65°C.
9. Add 10 µl 1 M Tris–HCl pH 7.6 to each sample.
10. To purify the cDNA samples and to remove Tris to prevent undesired coupling of the monofunctional NHS-ester Cy-dyes to free amine groups in solution, add 60 µl water to each sample to bring the reaction volume to 100 µl.
11. Add 500 µl Qiagen buffer PB to each sample, mix well by pipetting, and apply to the PCR purification columns.
12. Centrifuge the columns at 3,800 × *g* for 1 min at room temperature and discard the flow-through.
13. Add 600 µl 80% ethanol to each column, centrifuge at 3,800 × *g* for 1 min, and discard the flow-through.
14. Repeat the ethanol washes two more times.
15. Centrifuge at 16,000 × *g* for 1 min to dry the column.
16. Elute each sample from the column into a microcentrifuge tube twice with 40 µl 65°C water.
17. Dry the cDNA samples in Speed-Vac (high-heat setting).
18. The cDNA pellet can now be frozen at –80°C for at least several weeks prior to coupling.

3.5. Cy3/Cy5 Dye Coupling of cDNA

1. Resuspend each of the dried cDNA samples in 3.5 µl water.
2. Arrange each pair of cDNA samples of which one will be labeled with Cy3 and the other with Cy5.
3. Preparation of Cy3 and Cy5 dyes: Resuspend each Cy3 and Cy5 dye pack in 15 µl anhydrous DMSO by pipetting up and down 30–50 times, and vortexing for 1 min, prior to pulse centrifugation. Immediately recap and reseal the DMSO. Each dye pack can label 12–14 cDNA samples.
4. Dye coupling: Add 30 µl 2× Bicarb buffer to the Cy3 dye. Quickly aliquot 3.5 µl to each of the Cy3 cDNA samples. Repeat for Cy5 samples.
5. Vortex cDNA samples for 1 min, pulse-centrifuge, and incubate for 30 min at room temperature in the dark.
6. Repeat the previous step.

7. Add 3.5 µl 4 M hydroxylamine to each cDNA sample to quench the reaction. Incubate for 15 min at room temperature in the dark. While the reaction is quenching, set up and label the Qiagen PCR purification columns and microcentrifuge tubes.
8. Purify labeled cDNA away from the dyes: Arrange the tubes in Cy3/Cy5 pairs – they will be combined in this step.
9. For all cDNA sample pairs: Add 70 µl water to the Cy3 cDNA sample. Add 500 µl buffer PB (Qiagen) to the Cy3 cDNA sample, mix by pipetting, combine with the Cy5 cDNA sample, and apply the mix to a Qiagen PCR purification column.
10. Centrifuge at 2,600 × *g* for 1 min and discard the flow-through.
11. Apply 700 µl buffer PE (Qiagen), centrifuge at 2,600 × *g* for 1 min, and discard the flow-through.
12. Repeat the previous step.
13. Centrifuge at 16,000 × *g* for 1 min to dry the column.
14. Elute each cDNA sample from the column into a microcentrifuge tube twice with 30 µl elution buffer (Qiagen). The cDNA samples should be slightly purple in color. The efficiency of Cy3/Cy5 coupling to the cDNA samples can be determined at wavelengths 532 and 635 nm, respectively, with a spectrophotometer.
15. Dry the cDNA samples in a Speed-Vac (high-heat setting). Proceed to prehybridize the microarray slides.

3.6. Microarray Hybridization and Washing

1. Hybridization volumes, reagents, chambers, and conditions vary according to specific expression microarray platforms. The following protocol is used for expression microarrays manufactured by spotting 60-mer oligonucleotides (Open Biosystems) onto polylysine-treated glass microscope slides.
2. Microarray slide prehybridization: Thaw Quackenbush prehybridization buffer at 65°C and then incubate at 42°C.
3. UV-crosslink the microarray slides.
4. Place microarrays into slide mailers, fill with prehybridization buffer, and incubate at 42°C for a minimum of 45 min.
5. At the end of prehybridization, wash the microarrays 4–5 times with water by filling the slide mailer with water, rocking it, and pouring out the water.
6. Dry the microarrays by centrifuging them in slide racks in a benchtop centrifuge containing a swinging-bucket rotor and microplate carriers at 200 × *g* for 5 min. Store the microarrays

in a slide box at 42°C until ready to hybridize the Cy3/Cy5-coupled cDNA samples.

7. Hybridizations: When the coupled cDNA samples are dry and the microarray slides are ready, resuspend each sample in 40 µl hybridization buffer. Add oligo spike-ins specific for control probes on the microarray if required.
8. Just prior to hybridization, heat the samples at 65°C for 3 min to denature, and then incubate at 42°C until ready for hybridization.
9. Place a microarray slide (array side up) at the bottom part of the chamber.
10. Pipette a Cy3/Cy5-coupled cDNA sample onto the microarray, taking care not to touch it with the pipette tip. Avoid the formation of bubbles since they will prevent uniform hybridization on the microarray.
11. Place one edge of the lifter cover slip on one side of the microarray slide with the lifter side down, and carefully lower the other edge onto the sample using a pair of curved forceps. Gently center the lifter cover slip over the microarray with forceps.
12. Fill the humidifying well at the end of the hybridization chamber with 4× SSC.
13. Seal the upper and lower portions of the hybridization chamber with screws/clips and incubate at 42°C overnight (or at least 6 h).
14. Record the label on the microarray and description of the samples in a hybridization sheet.
15. Repeat **steps 9–14** for the next sample.
16. Microarray slide washes: After hybridization is complete, fill three 250-ml glass dishes with Wash 1, Wash 2, and Wash 3.
17. Open a hybridization chamber, and pick up a microarray slide with forceps. Submerge the microarray slide in Wash 1. The lifter cover slip should fall off almost immediately when the microarray slide is submerged. Be careful to avoid scratching the array.
18. Dip the microarray slide up and down 20 times in Wash 1.
19. Dip the microarray slide up and down 20 times in Wash 2.
20. Dip the microarray slide up and down 20 times in Wash 3.
21. Quickly dry the microarray slide (array side first) by immediately blowing off the wash solution with a dust gun or some other compressed gas, going from one side to the other in about 5 s. Store the microarray slide in a slide box.
22. Repeat **steps 17–21** for the next microarray slide.

3.7. Acquisition, Processing, and Analysis of Microarray Data

1. Scan microarrays, quantitate images, and normalize data by Lowess smoothing with the laser scanner and integrated software according to the manufacturer's instructions *(7)*. Each pair of fluor-reversal experiments is combined by averaging the ratios of normalized median intensity of TF overexpression vs. the empty vector control for each spot on the microarray.
2. Examine the microarray data for specific upregulation of the TF mRNA in the corresponding TF overexpression experiment to confirm the induction of the TF gene in the presence of galactose and the correct orientation of the microarray samples (**Fig. 2**).
3. Calculate the Pearson correlation between each pair of fluor-reversal experiments. A Pearson correlation of >0.3 between replicates is indicative that the differential gene expression in an experiment is more likely caused by TF overexpression rather than microarray noise *(5)*. This correlation threshold may be different for other microarray platforms (Agilent, Affymetrix, Nimblegen, etc.).
4. Input subsets of upregulated and downregulated genes separately at various thresholds of differential expression into FunSpec (http://funspec.med.utoronto.ca/) to identify significant functional enrichment of Gene Ontology categories (*see* **Note 3**). The segregation of upregulated and downregulated genes is required to characterize transcriptional activators and repressors, respectively. The significant functional

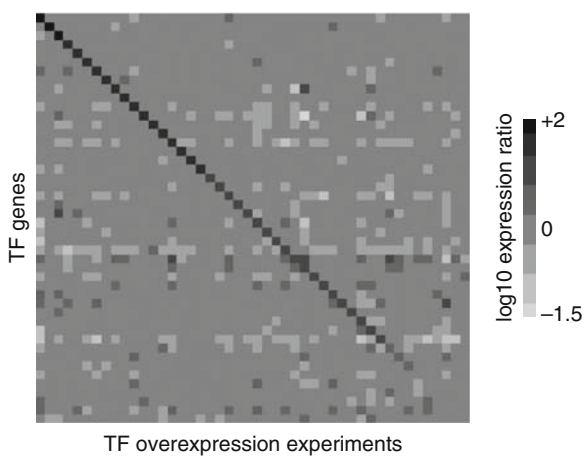

Fig. 2. Induction of transcription factor genes in 46 overexpression strains exhibiting phenotypic activation. The majority of transcription factor genes are highly induced in the corresponding transcription factor overexpression strains from microarray profiling (*diagonal*). The induction of transcription factor genes along the diagonal are highest and lowest in the upper left and lower right, respectively.

categories appearing in the list of differentially regulated genes should provide information on the TF's putative function and targets since both are expected to be involved in a common biological process (**Fig. 3**).

5. Subject the microarray data of each TF overexpression experiment to motif searching using RankMotif (*see* **Note 4**). This probabilistic-inference algorithm is applied to the promoter regions of differentially expressed genes to identify prevalent 8-mer sequences that represent putative TF-binding sites (**Fig. 3**). Motif predictions can be validated using gel mobility shift assays involving the DNA-binding domain of the TF and two tandem copies of the predicted motif *(3)*.

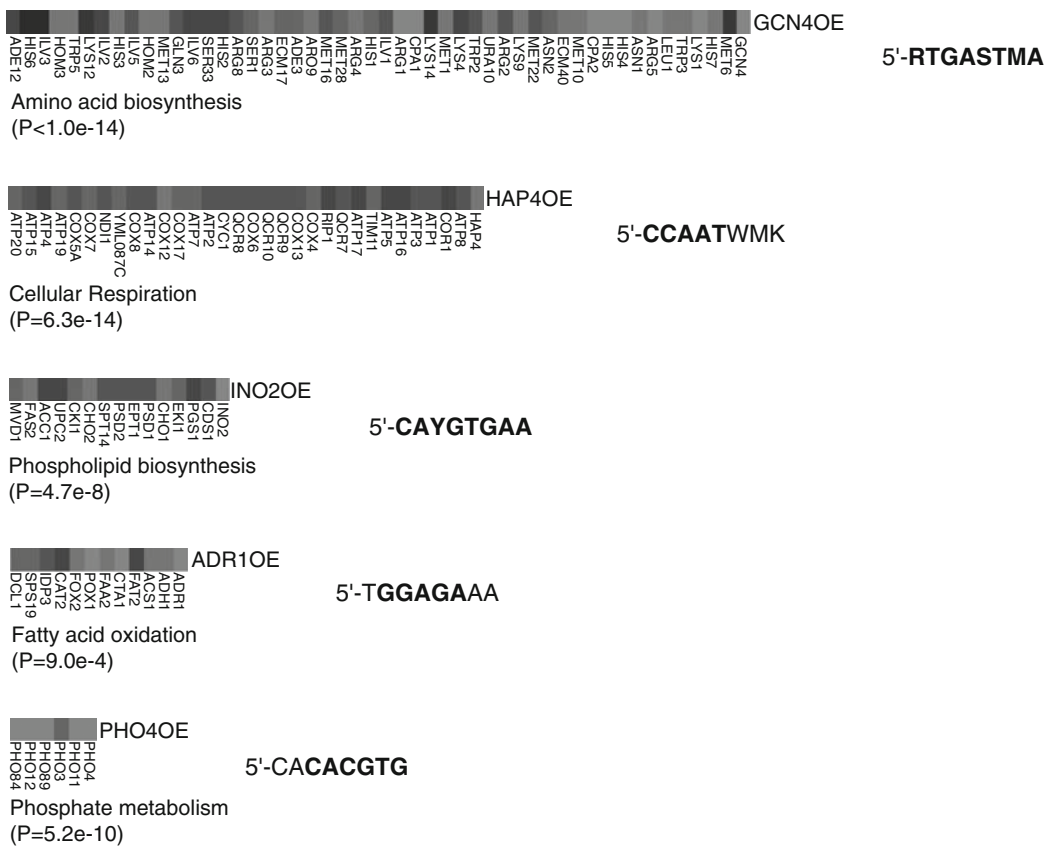

Fig. 3. Putative targets identified for several transcription factors by phenotypic activation and microarray expression profiling. Genes upregulated in response to transcription factor overexpression were searched for (1) significant functional enrichment of gene ontology categories (*P* values) using FunSpec and; (2) overrepresented 8-mer sequences in promoter regions using RankMotif (S = GC; W = AT; R = AG; Y = CT; K = GT; M = AC). Nucleotides of predicted motifs that match known motifs are shown in *bold*.

4. Notes

1. Microarray expression profiles recovered from 23 well-characterized TFs that failed to exhibit an overexpression-related growth defect were all muted, indicating that toxic overexpression is a strong indicator of TF activation *(3)*. It is possible that some of the nontoxic TF overexpressors simply contained epitope-tagged proteins that were nonfunctional. There is currently another overexpression array available which uses a different fusion protein tagging format from that used in this study, and this array is also designed to allow rapid swapping of various epitope tags *(8)*. These assorted versions of the overexpression array can then be screened for novel occurrences of phenotypic activation. In addition, many TFs in the helix–loop–helix, basic–leucine–zipper, and Gal4 classes are obligate dimers and some proportion of these would be obligate heterodimers, which may require overexpression of both subunits in order to obtain hyperactivation of the TF *(9)*. Therefore, one possible strategy to tease out the target genes of these TFs is to identify associations of distinct TFs by protein–protein interaction techniques followed by co-overexpression to phenotypically activate them.
2. Despite obtaining meaningful microarray expression profiles from the phenotypic activation of TFs, identifying and validating their target genes and binding motifs remain challenging. A large part of the difficulty lies in eliminating or filtering out the differential gene expression caused by the induction of secondary pathways resulting from the toxic effects of TF overexpression. One possible approach to overcome this obstacle is to microarray expression-profile earlier and later time points of TF induction. This would conceivably better discriminate primary targets from secondary ones and lead to better predictions by motif-finding algorithms. We had selected a single 3 h induction time point for each TF gene under control of the *GAL1/10* promoter because time course studies initially on several TFs had indicated that this induction period was optimal with not much advantage gained from earlier or later times *(3, 10)*. This is likely not the case for all TFs.
3. One complication in identifying significant enrichment of Gene Ontology categories and motif sequences from microarray expression data is that the optimum arbitrary threshold to define genes that are differentially regulated is difficult to determine. The optimum cut-off selection of differentially regulated genes will certainly vary among expression profiles of different TFs. As a result, several analyses of distinct subsets of differentially expressed genes based on various arbitrary

thresholds are required. One approach that circumvents this problem is the use of threshold-independent statistical methods such as the Wilcoxon–Mann–Whitney (WMW) test to identify significant enrichment of various categories. In this method, the expression ratios of all genes from a particular microarray experiment are initially ranked from the highest to the lowest. The WMW test determines whether the differences in the median expression ratio ranks between genes belonging to a given category and those that are not are statistically significant. This approach has been successfully applied to microarray data to determine functional enrichment, motif finding, and comparison of various genome-wide datasets *(3, 11, 12)*.

4. The probabilistic-inference algorithm RankMotif has proven more robust for microarray expression data with considerable secondary effects than other motif-finding algorithms such as BioProspector *(3, 13)*. For example, motif searches by BioProspector frequently outputs the stress–response element 5′-CCCCT-3′ found in the promoters of numerous upregulated genes as a secondary response to TF overexpression. To overcome this problem, RankMotif is designed to search for a motif that is specific for a particular experiment as well as an additional motif that is overrepresented in multiple experiments. RankMotif is available upon request to Quaid Morris (University of Toronto).

Acknowledgments

This author would like to thank Professor Timothy R. Hughes for his support, advice, and critical reading of this manuscript. This work was supported by the Charles H. Best Postdoctoral Fellowship to G.C. and the Natural Sciences and Engineering Research Council to T.R.H.

References

1. Chua, G., Robinson, M. D., Morris, Q., and Hughes, T. R. (2004). Transcriptional networks: reverse-engineering gene regulation on a global scale. *Curr. Opin. Microbiol.* **7**, 638–646.
2. Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S. G., Cyert, M., Hughes, T. R., Boone, C., and Andrews, B. (2006). Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330.
3. Chua, G., Morris, Q. D., Sopko, R., Robinson, M. D., Ryan, O., Chan, E. T., Frey, B. J., Andrews, B. J., Boone, C., and Hughes T. R. (2006). Identifying transcription factor functions and targets by phenotypic activation. *Proc. Natl. Acad. Sci. USA* **103**, 12045–12050.

4. Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M., and Friend, S. H. (2000). Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126.
5. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D., and Hughes, T. R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol. Cell. Biol.* **24**, 5534–5547.
6. Robinson, M. D., Griqull, J., Mohammad, N., and Hughes, T. R. (2002). FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics* **3**, 35.
7. Mnaimneh , S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W. -T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., Trochesset, M., Morse, D., Krogan, N. J., Hiley, S. L., Li, Z., Morris, Q., Grigull, J., Mitsakakis, N., Roberts, C. J., Greenblatt, J. F., Boone, C., Kaiser, C. A., Andrews, B. J., and Hughes, T. R. (2004.) Exploration of essential gene functions via titratable promoter alleles. *Cell* **118**, 31–44.
8. Gelperin, D. M., White, M. A., Wilkinson, M. L., Kon, Y., Kung, L. A., Wise, K. J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., Gerstein, M., Dumont, M. E., Phizicky, E. M., Synder, M., and Grayhach, E. J. (2005). Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* **19**, 2816–2826.
9. Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000). An overview of the structures of protein–DNA complexes. *Genome Biol. 1*, REVIEWS001.
10. 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.
11. Berger, M. F., Philippakis, A. A., Qureshi, A. M., He, F. S., Estep, P. W. III, and Bulyk, M. L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435.
12. Chen, X., Hughes, T. R., and Morris, Q. (2007). RankMotif ++: a motif-search algorithm that accounts for relative ranks of K-mers in binding transcription factors. *Bioinformatics* **23**, i72–i79.
13. Liu, X., Brutlag, D. L., and Liu, J. S. (2001). BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.* **6**, 127–138.

http://www.springer.com/978-1-934115-71-8

Yeast Functional Genomics and Proteomics
Methods and Protocols
Stagljar, I. (Ed.)
2009, XIII, 304 p., Hardcover
ISBN: 978-1-934115-71-8
A product of Humana Press