

Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*

David J. Galgoczy*, Ann Cassidy-Stone^{†‡}, Manuel Llinás*, Sean M. O'Rourke*[§], Ira Herskowitz*[¶], Joseph L. DeRisi*, and Alexander D. Johnson*^{¶||}

Departments of *Biochemistry and Biophysics and [†]Microbiology and Immunology, University of California, 600 16th Street, San Francisco, CA 94143-2200

Edited by Stanley Fields, University of Washington, Seattle, WA, and approved November 11, 2004 (received for review October 14, 2004)

The budding yeast *Saccharomyces cerevisiae* has three cell types (a cells, α cells, and a/ α cells), each of which is specified by a unique combination of transcriptional regulators. This transcriptional circuit has served as an important model for understanding basic features of the combinatorial control of transcription and the specification of cell type. Here, using genome-wide chromatin immunoprecipitation, transcriptional profiling, and phylogenetic comparisons, we describe the complete cell-type-specification circuit for *S. cerevisiae*. We believe this work represents a complete description of cell-type specification in a eukaryote.

chromatin immunoprecipitation | mating | transcriptional circuit

A problem of central importance in understanding multicellular organisms is how different cell types are stably maintained. Typically, cell-type specification is based on a transcriptional circuit in which combinations of regulatory proteins determine the final pattern of gene expression that is appropriate to a given cell type. Although unicellular, the yeast *Saccharomyces cerevisiae* has three distinct types of cells, and the cell-specification circuit is combinatorial (refs. 1–3 and Fig. 1). The **a** and α cell types are typically haploid in DNA content and mate with each other in an elaborate ritual that culminates in cellular and nuclear fusion. These events produce the third type of cell, the **a/ α** cell type, which is typically diploid. This cell type cannot mate but, when environmental conditions are appropriate, can undergo meiosis and sporulation, producing two **a** and two α cell types. The patterns of cell-type-specific gene expression are set up by a few sequence-specific DNA-binding proteins acting in various combinations. Three critical proteins ($\alpha 1$, $\alpha 2$, and **a1**) are encoded by the mating-type (*MAT*) locus. A fourth key sequence-specific DNA-binding protein (Mcm1) is encoded elsewhere in the genome. In this article we use the term “cell-type-specification circuit” to refer to the regulatory scheme diagrammed in Fig. 1, because each component and branch of this scheme is necessary and sufficient to establish and maintain three cell types.

In this article we apply three methods [genome-wide chromatin immunoprecipitation (ChIP), genome-wide transcriptional profiling, and phylogenetic comparisons] in an attempt to completely determine the cell-type-specification circuit in *S. cerevisiae* (4–8). The use of three different techniques generated considerably more data than are needed to reconstruct the circuit, and because it is overdetermined, we believe our circuit description to be very accurate, containing at most only a few false negatives or positives.

Materials and Methods

Strains. Isogenic EG123-derived strains (*his4 leu2 trp1 ura3 can1*; see ref. 9) were used for the experiments shown in Figs. 2 A and B and 7. Isogenic $\Sigma 2000$ strains (*his3 Δ ::hisG leu2 Δ trp1 Δ ::hisG ura3 Δ*) and isogenic EG123 strains were used for the experiments shown in Figs. 2 C and 5 B–D. yDG208 (*MATa ura3-52 lys2-801[amb] ade2-101 his3- Δ 200*) and yDG240 (*MATa/MAT α ura3-52/ura3-52 lys2-801 [amb]/lys2-801[amb] ade2-101[och]/ade2-101[och] trp1- Δ 1/trp1- Δ 1*) were used for the experiments

shown in Figs. 3, 5 A and B, and 6. EG123, yDG208, and yDG240 are all derivatives of S288C. For the salt-sensitivity experiment, the **a1- α 2** site in the endogenous *HOG1* gene promoter was replaced by integration of *Kluyveromyces lactis URA3*, which was subsequently replaced by the integration of an oligonucleotide-generated construct to restore the *HOG1* promoter with a modified **a1- α 2**-binding site: GCGTGgCGGATTTTACgCC (lowercase “g” replaced T, A, and T in the wild-type sequence). These nucleotides are highly conserved among **a1- α 2**-binding sites and have been demonstrated to be critical for **a1- α 2** binding and repression (10). The promoter was sequenced to verify correct integration.

DNA Microarrays. Arrays containing both ORFs and intergenic sequences (13,200 elements total) were described previously, as were microarrays containing only ORFs (4, 11).

Expression Microarrays. Transcriptional profiling, yeast growth, RNA isolation, microarray procedures, and data acquisition were performed as described (12) except cultures were not treated with salt. For each strain analyzed, cDNA was made in duplicate, and two microarrays were probed, using a reference sample made from equal parts of 11 strains tested. The average expression ratios (ratio of means) were used for subsequent analysis.

ChIP Experiments. Overnight cultures were grown in yeast extract, peptone, and dextrose (YEPD) for ≈ 16 h at 30°C to an OD₆₀₀ of 0.2. Immunoprecipitation (IP), amplification, and fluorescence labeling were carried out as described (5). Microarray hybridization was carried out as described (13), and data were uploaded to the NOMAD database (<http://ucsf-nomad.sourceforge.net/>) for additional analysis.

Chromosomal maps for ChIP data were generated in PROMOTERPROJECT 2.3 (software created by J.L.D.).

Binding-Site Motif Identification and Phylogenetic Comparisons. For binding-site motif discovery, sequences were submitted to the search tool MEME (14). Motif consensus matrices generated by MEME were submitted to the tool MAST (15) to search for matching sequences within -950 to $+50$ upstream of all genes in the *S. cerevisiae* genome. For $\alpha 2$ -Mcm1 and $\alpha 1$ -Mcm1 analyses, the *E* values assigned by MAST were invariably <0.5 for

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: ChIP, chromatin immunoprecipitation; YEPD, yeast extract, peptone, and dextrose; IP, immunoprecipitation.

^{*}Present address: Section of Molecular and Cellular Biology, 1 Shields Avenue, University of California, Davis, CA 95616.

[§]Present address: Institute of Molecular Biology, 1229 University of Oregon, Eugene, OR 97403-1229.

[¶]Deceased April 28, 2003.

^{||}To whom correspondence should be addressed. E-mail: ajohnson@cgl.ucsf.edu.

© 2004 by The National Academy of Sciences of the USA

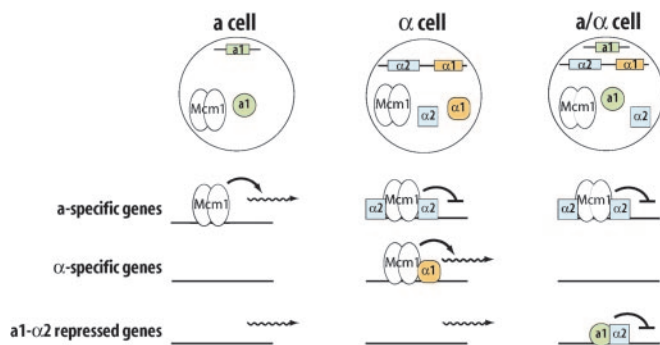


Fig. 1. Cell-type regulation in *S. cerevisiae*.

genes identified by ChIP, expression, or both. For the $\alpha 1$ - $\alpha 2$ analysis, some genes confirmed by ChIP, expression analysis, and phylogenetic comparison had E values as high as 50, reflecting the greater degeneracy of the $\alpha 1$ - $\alpha 2$ site. All DNA sequences with MAST E values < 5 for $\alpha 2$ -Mcm1 and $\alpha 1$ -Mcm1 analyses and < 100 for the $\alpha 1$ - $\alpha 2$ analysis were aligned with homologous sequences from four other closely related *Saccharomyces sensu stricto* species (7, 8) by using CLUSTALW at the *Saccharomyces* Genome Database (<http://yeastgenome.org>).

Salt-Sensitivity Assays. To test salt sensitivity, strains were grown overnight in YEPD liquid medium at 30°C to an OD_{600} of 0.5, sonicated, diluted, and plated on YEPD plates with and without 1 M NaCl. Plates were incubated at 30°C for 2 (YEPD) or 3 (YEPD plus NaCl) days and then photographed.

Results

a-Specific Genes (Genes Repressed by $\alpha 2$ -Mcm1). To identify the complete set of a-specific genes, we first carried out ChIPs in α cells by using an antibody directed against the $\alpha 2$ protein. For this analysis, we directly compared DNA precipitated from α cells, in which $\alpha 2$ is present, to otherwise isogenic a cells, in which it is absent (see Fig. 1). Immunoprecipitated DNA was amplified randomly by PCR amplification, labeled with Cy3 or Cy5 fluorescent dyes, and competitively hybridized to DNA microarrays containing both ORFs and intergenic elements from the *S. cerevisiae* genome (16).

Four iterations of this experiment were carried out, and all led to the same conclusions. The results of one experiment are shown in Fig. 2 *A* and *B*. For this experiment, the DNA was sheared minimally; because the shear length was relatively large (median, ≈ 1 kb), each genomic position bound by $\alpha 2$ gave rise to a cluster of enriched microarray elements when mapped on the genome (Fig. 2). As shown in Fig. 2, 45 of the 46 highest elements with the highest α/a ratios fell into eight clusters in the genome. We systematically lowered the fluorescence ratio threshold and observed that the additional elements now included in the data set did not assort into any new clusters in the genome. Thus, the point at which the clustering broke down was chosen as the significance threshold. We note that this clustering is particularly useful when a protein is bound to relatively few sites in a genome; although high resolution is compromised, the distinction between signal and noise is very clear. From these results, we can conclude with high confidence that $\alpha 2$ occupies eight positions in the genome of this S288C strain of *S. cerevisiae* grown at 30°C in YEPD medium. Based on previous results, we believe that each genomic site corresponds to a dimer of $\alpha 2$ and a dimer of Mcm1 bound together to one or more twofold-symmetric DNA sequences (1–3). Based on prior *in vivo* dimethyl sulfate protection studies (9), we believe the occupancy of these sites to be close to 100%. Because the microarrays

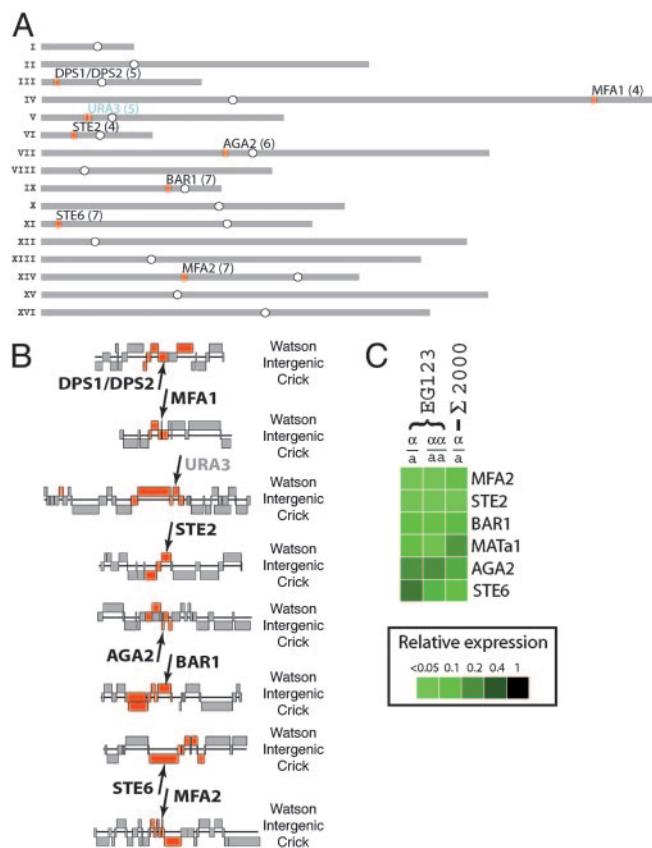


Fig. 2. Repression of the a-specific genes by $\alpha 2$ -Mcm1. (*A*) Positions on the yeast genome occupied by $\alpha 2$ in α cells, as determined by a genome-wide ChIP experiment. The 46 elements with the highest α/a fluorescence ratios are mapped as red bars on the chromosomes (see Results), with the number of enriched elements at each position indicated in parentheses. (*B*) A detailed view of the $\alpha 2$ -enriched elements described for *A*. The strains used in this experiment contained a reporter construct that included $\alpha 2$ -Mcm1-binding sites integrated at *URA3* (indicated in gray lettering); *URA3* is not otherwise a target of $\alpha 2$ -Mcm1. (*C*) a-Specific genes as determined by transcriptional profiling on genomic ORF microarrays. Shown are all genes with expression levels < 0.5 -fold in α cells compared with a cells and in a/α diploid cells compared with a/a diploid cells. Because of a PCR-amplification failure, the *MFA1* spot was unusable in these arrays. It is known from other studies (reviewed in refs. 1–3) that *MFA1* mRNA levels are dramatically lower in α cells compared to a cells.

represent the entire genome, we can conclude also that $\alpha 2$ does not occupy any other specific sites in the genome at these stoichiometries.

Six of these positions represent the control regions of previously known a-specific genes: *STE2*, *STE6*, *MFA1*, *MFA2*, *BAR1*, and *AGA2* (10). The seventh position occupied by $\alpha 2$ in the α strain corresponds to a reporter gene, containing $\alpha 2$ -Mcm1-binding sites, that was integrated at the *URA3* locus as a positive control. The eighth position corresponds to two adjacent $\alpha 2$ -Mcm1-binding sites (termed *DPS1/DPS2*) that regulate a re-combinational enhancer involved in the gene-conversion events that underlie mating-type interconversion (17, 18).

We independently investigated the a-specific genes by examining genome-wide transcriptional differences between isogenic a and α cells by using a series of isogenic strains differing only at the mating-type locus (Fig. 2*C*). As indicated, the transcriptional profiling is in excellent agreement with the genome-wide ChIP.

As a third approach for verifying our assignment of a-specific genes, we analyzed the genomic positions identified thus far in

this analysis by using the computer program MEME (14) to identify shared sequence motifs and generate an **a**-specific consensus sequence, which matched the previously described $\alpha 2$ -Mcm1-binding site (1–3). Using a second tool, MAST (15), we searched the *S. cerevisiae* genome for additional matches to this consensus sequence. Matches then were validated by aligning them with the syntenic regions from four closely related *Saccharomyces sensu stricto* genomes (7, 8). Among genomic sites identified by MAST, only the $\alpha 2$ -Mcm1 sites at the positions identified in our ChIP and one additional site, located adjacent to the *ASG7* gene, were evolutionarily conserved. *ASG7* was identified previously as an $\alpha 2$ -Mcm1-regulated gene, and although the molecular function of its product is not known, it is expressed only after pheromone induction (19); thus, it would not have been identified as an **a**-specific gene in the transcriptional profiling experiment shown in Fig. 2C. It was also not identified in the ChIP experiments as being significantly above background. These results indicate that in the absence of pheromone, $\alpha 2$ -Mcm1 does not significantly occupy the promoter region of *ASG7* in α cells.

From all three of these analyses we conclude that, in α cells of a standard laboratory strain, $\alpha 2$ -Mcm1 significantly occupies seven positions in the genome. Six sites control transcription of **a**-specific genes, and the seventh site controls the activity of a recombinational enhancer. The regulation of the **a**-specific genes is therefore all direct: $\alpha 2$ -Mcm1 does not control **a**-specific gene transcription indirectly through intermediary gene-regulatory proteins.

Genes Repressed by $\mathbf{a1-\alpha 2}$. To identify genes directly regulated by $\mathbf{a1-\alpha 2}$, we carried out ChIP in **a**/ α cells by using an antibody directed against $\alpha 2$. As before, we used IPs from **a** cells (in which $\alpha 2$ is not expressed) as the reference sample for these experiments. To maximize resolution, the DNA was sheared more extensively than that in the previous experiment. The overall IP enrichment in these experiments was lower than those described above, probably because a monomer of $\alpha 2$ is present at each $\mathbf{a1-\alpha 2}$ site, whereas a dimer is present at each $\alpha 2$ -Mcm1 site (3). Because of this lower enrichment, we carried out three separate ChIPs from each strain and hybridized each IP to two microarrays to give a total of six sets of data. To combine these data sets, we converted fluorescence ratios into percentile ranks for each microarray and then calculated a median percentile rank across all experiments for each microarray element (5, 20). The distribution of median percentile ranks (Fig. 3) forms a trough at approximately the 97 percentile point, above which lie a set of DNA fragments that are highly enriched in the replicate IP experiments (5, 20). To be conservative, we chose a cutoff slightly to the left of the inflection point (see Fig. 3 *Inset*); this choice is expected to include some noise (which, as described below, was eliminated later by other criteria) but was unlikely to exclude any fully occupied $\mathbf{a1-\alpha 2}$ sites. This cutoff gave 71 elements representing 36 different positions in the genome. The six **a**-specific genes and the recombinational enhancer were all represented in this data set, because $\alpha 2$ -Mcm1 is bound to these genes in **a**/ α cells just as it is in α cells. (The inclusion of this set of genes demonstrates that the circuit description is independent of differences in experimental conditions, such as extent of DNA shearing and the way the data were processed.) When these positions were placed aside, 29 genomic positions, each potentially representing a site occupied by $\mathbf{a1-\alpha 2}$, remained. Because some of the genomic positions were represented by only a single array element (as opposed to a cluster of elements), we used an additional criterion to determine the genomic positions that represent bona fide $\mathbf{a1-\alpha 2}$ sites: we required that the immunoprecipitated element contain at least one DNA sequence that resembles a known $\mathbf{a1-\alpha 2}$ recognition sequence and that these sites be conserved in the related *Saccharomyces sensu stricto*

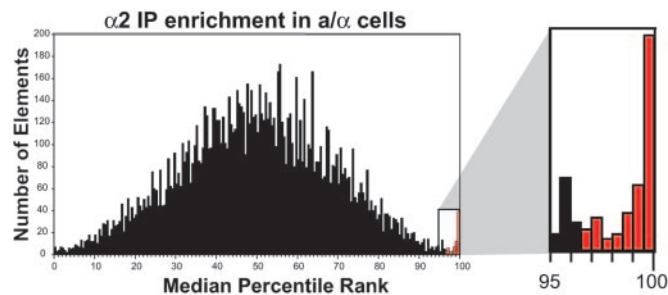


Fig. 3. Reproducibility of ChIPs performed with anti- $\alpha 2$ antibodies in **a**/ α cells. ChIPs from **a**/ α cells and α cells were carried out three independent times, and each pair was hybridized to two arrays for a total of six arrays. The median percentile rank values for each element across the six microarrays are displayed in the histogram, which gives the number of array elements for each median percentile rank value (bin size = 0.5 percentile). The increase from the 96th to 100th percentile ranking reflects a reproducibly high enrichment of a set of elements across the six separate microarrays. Based on the inflection point position we used a significance cutoff at the 96.5 percentile rank (elements with percentile ranks above this cutoff are represented by red histogram bars). These highly enriched elements represent $\approx 0.5\%$ of the total number of elements on the microarray.

species (7, 8). Given the strong conservation of $\mathbf{a1-\alpha 2}$ sites against background intergenic sequences (for example, see Fig. 4), this analysis provided an effective secondary screen to eliminate noise in the ChIP experiment.

The requirement for a conserved binding site eliminated 6 of the 29 positions, leaving 23 confirmed positions at which $\mathbf{a1-\alpha 2}$ is bound in **a**/ α cells. The positions that were eliminated were each represented by only a single element, and their median percentile ranks were generally just above our cutoff. Thus, our conservative evaluation of the inflection point of the ChIP data (Fig. 3) did result in the inclusion of a few false positives that were subsequently eliminated by additional criteria. To check for false negatives (i.e., occupied sites that might have been missed by the ChIP experiment), we also searched all intergenic regions of the *S. cerevisiae* genome for additional sites that matched, at least loosely, the $\mathbf{a1-\alpha 2}$ consensus sequence that were conserved across the other *Saccharomyces sensu stricto* species. Seven potential sites were found with scores significantly above background. None of these seven sites met our additional criteria and are therefore not included in our final analysis. (We do note that three of these sites, adjacent to the *RPL15A*, *REX2*, and *SPT23* genes, fall just under our ChIP significance threshold and could represent sites of relatively low $\mathbf{a1-\alpha 2}$ occupancy.) Thus, the ChIP experiments and the phylogenetic comparisons converged on 23 genomic positions occupied by $\mathbf{a1-\alpha 2}$ in **a**/ α cells.

We independently examined $\mathbf{a1-\alpha 2}$ -regulated genes by transcriptional profiling as described above (Fig. 5). A comparison of the 23 genomic positions at which $\mathbf{a1-\alpha 2}$ was bound (Fig. 5A) with the transcriptional profiling of genetically matched diploid strains (Fig. 5B and C) revealed four points. First, at three locations in the genome, a single $\mathbf{a1-\alpha 2}$ site regulates a pair of divergently transcribed genes (Fig. 6). Second, a comparison of the ChIP and transcriptional profiling data revealed genes (e.g., *TEC1* and *SST2*) that are indirectly regulated by $\mathbf{a1-\alpha 2}$, because $\mathbf{a1-\alpha 2}$ is not bound nearby, nor are there sequences in the



Fig. 4. Evolutionary conservation of the *STE18* $\mathbf{a1-\alpha 2}$ -binding site. Vertical green lines indicate nucleotides identical in the DNA sequences of *S. cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, *Saccharomyces bayanus*, and *Saccharomyces kudriavzevii* (7, 8).

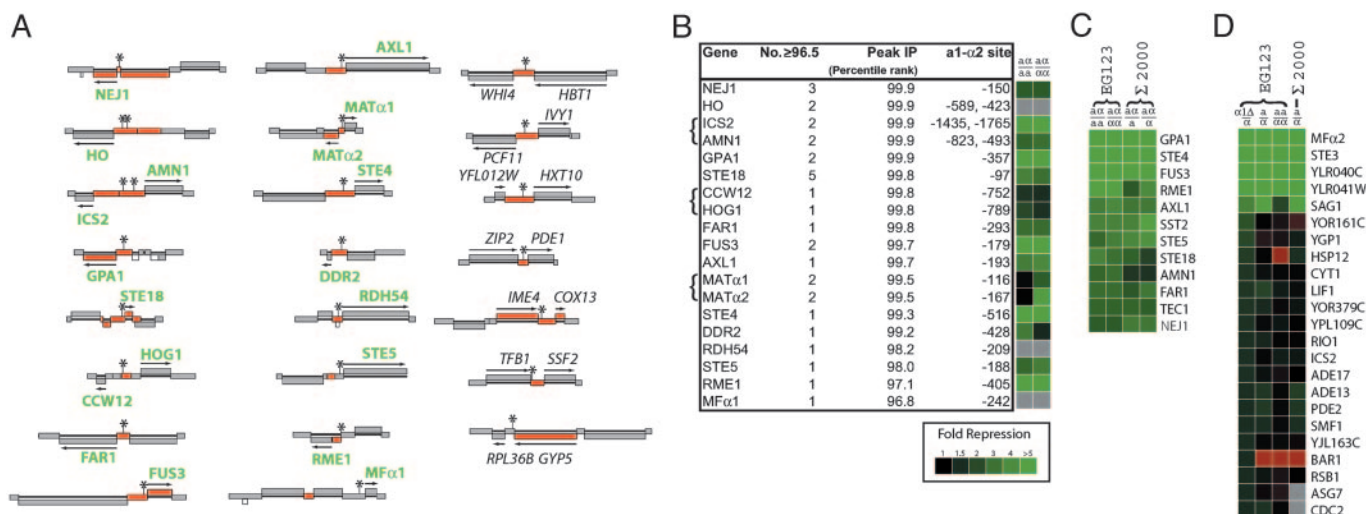


Fig. 5. Analysis of $a1-\alpha2$ and $\alpha1$ -Mcm1 regulation. (A) A detailed view of the 23 sites in the a/α cell genome that were enriched in anti- $\alpha2$ IPs for which conserved $a1-\alpha2$ sites (indicated by asterisks) were identified. Genes that were repressed at least 1.5-fold in a/α cells (versus a/a or α/α cells; see B and C) or that were previously published as being repressed in a/α cells are indicated in bold, green lettering. Enriched elements that contain (or are adjacent to) phylogenetically conserved $a1-\alpha2$ -binding sites but do not seem to regulate an adjacent gene are indicated with the surrounding genes in black lettering. (B) $a1-\alpha2$ -regulated genes. Expression patterns for each gene denoted in green in A are displayed. For each gene, the number of microarray elements yielding IP percentile ranks ≥ 96.5 and the highest percentile rank among those elements are indicated. Brackets indicate genes that are divergently transcribed and are controlled by the same binding site (or in some cases a small cluster of sites). The positions of phylogenetically conserved $a1-\alpha2$ -binding sites relative to the start of the ORF are indicated. The degree of transcriptional repression in a/α cells (versus a/a or α/α cells) is indicated by the intensity of green color in boxes to the right of the table (gray indicates no data). (C) Haploid-specific genes identified by transcriptional profiling. We compared transcriptional profiles of a/α cells with a/a and α/α cells irrespective of ChIP results. This analysis revealed genes directly repressed by $a1-\alpha2$ and genes indirectly regulated. Genes that were transcriptionally down-regulated at least 2-fold in EG123 a/α cells and 1.5-fold in $a/\alpha \geq 2000$ cells are displayed. (D) α -Specific genes identified by transcriptional profiling. Expression levels from a $mat\alpha1\Delta$ and a wild-type α strain are compared (first column), and all genes down at least 1.5-fold in expression are displayed in order of ascending ratios. Also displayed are three sets of expression ratios comparing a and α (or a/a and α/α diploid) strains. Only the first five genes reproducibly score as α -specific genes across all four experiments. **YLR041W* is annotated as a "dubious" ORF in the *Saccharomyces* Genome Database (<http://yeastgenome.org/>). It overlaps *YLR040C* almost completely, and the two ORFs cannot be distinguished because of cross-hybridization. We believe that *YLR040C* is an α -specific gene (see Results) and that it cross-hybridizes to the *YLR041W* spot.

upstream promoter regions of these genes related to the consensus $a1-\alpha2$ site. Third, under the media conditions used in these experiments, no genes were expressed more highly in a/α

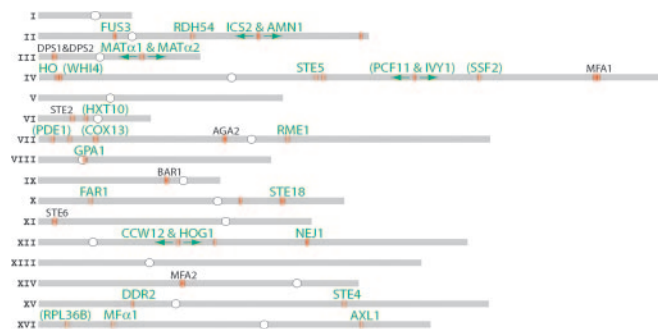


Fig. 6. Comprehensive mapping of sites in the yeast genome occupied by $a1-\alpha2$ and $\alpha2$ -Mcm1. All microarray elements with median percentile ranks > 96.5 are mapped as red bars on the chromosomes (see Fig. 3 and the text for the rationale of this cutoff point). The α -specific genes were also identified in this experiment, and they are indicated by black lettering. Genes lettered in green denote elements enriched in the IP that contain a phylogenetically conserved binding site that regulates the indicated gene. In some cases (e.g., *CCW12* and *HOG1*), a single $a1-\alpha2$ site regulates two divergently transcribed genes. Green letters in parentheses denote positions enriched in the IP that contain a phylogenetically conserved $a1-\alpha2$ sequence, but in these cases the neighboring gene does not seem to be transcriptionally regulated by $a1-\alpha2$. Red vertical bars with no corresponding lettering indicate elements enriched in the IP (≥ 96.5 percentile rank) but that fail to pass subsequent criteria (see Results). These elements are presumed to represent noise in the ChIP experiment.

cells than in a and α cells; i.e., there were no " a/α -specific" genes. Fourth, seven sites occupied by $a1-\alpha2$ in a/α cells and conserved in the *sensu stricto* strains did not seem to regulate known genes. There are two plausible explanations for this last observation. (i) The gene may simply not be expressed under the conditions (30°C, rich medium) used for the experiment. (ii) These $a1-\alpha2$ sites control the production of transcripts that have not been annotated. As described in Discussion, we believe that this latter explanation holds for at least one of these $a1-\alpha2$ -binding sites.

Among the 19 annotated genes identified as being directly regulated by $a1-\alpha2$ are all previously reported $a1-\alpha2$ -regulated genes (*HO*, *STE5*, *MFα1*, *MATα1*, *MATα2*, *FUS3*, *GPA1*, *RME1*, *AXL1*, *NEJ1*, *STE4*, *STE18*, *FAR1*, *AMN1*, and *RDH54* (1–3, 21–24). This result indicates that our false-negative rate is very low, possibly zero. Four additional genes (*HOG1*, *ICS2*, *DDR2*, and *CCW12*) were not recognized previously to be under cell-type control and were identified in this study. The regulation of *HOG1* is addressed below.

α -Specific Genes (Genes Activated by $\alpha1$ -Mcm1). We next turned to the final class of cell-type-specific genes, the α -specific genes. For this analysis, we omitted the ChIP experiment and examined whether it would be possible from transcriptional profiling and phylogenetic analyses to determine, with high confidence, all the genes activated by $\alpha1$ -Mcm1. We first analyzed the cell-type transcriptional patterns and identified four genes with transcription that is up-regulated at least 3-fold by $\alpha1$ (Fig. 5D). Using MEME, we developed a weighted consensus sequence and searched all promoter regions in the *S. cerevisiae* genome for matching sequences by using MAST.

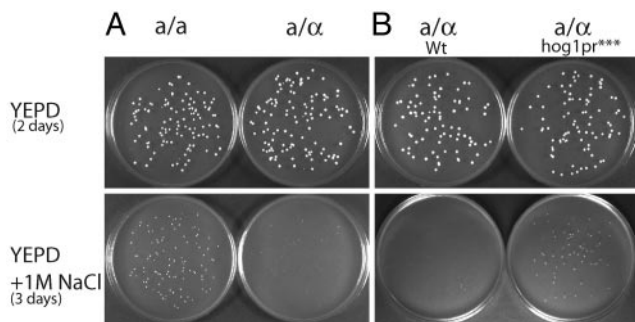


Fig. 7. $\alpha 1$ - $\alpha 2$ repression of *HOG1* confers sensitivity to NaCl in a/α cells. (A) Isogenic a/a and a/α diploid strains. (B) An a/α wild-type strain and an isogenic a/α strain containing three point mutations that disrupt the $\alpha 1$ - $\alpha 2$ -binding site in the *HOG1* promoter (*hog1pr****).

MAST easily picked out the four individual sequences that went into the consensus as well as an additional sequence lying upstream of the *MF α 1* gene, a gene encoding the α pheromone, α -factor. *MF α 1* is known to be regulated by $\alpha 1$ -Mcm1 (1–3), but a failure during microarray production rendered the *MF α 1* spot unusable. When this sequence was incorporated into a new consensus sequence and the genome was searched again, the five sites that went into the weighted consensus were identified by MAST as the strongest matches, and all had low *E* values (≤ 0.0011). The next closest match in an intergenic region had an *E* value of 3.5. This relatively poor match, however, was not phylogenetically conserved across the *Saccharomyces sensu stricto* species; in contrast, the five bona fide sites showed strong evolutionary conservation. We examined more closely the upstream regions of all the genes with regulation that seemed to weakly depend on $\alpha 1$ (i.e., showing a 1.5- to 3-fold effect on the transcriptional microarray) and found that none contained even weak matches to the weighted $\alpha 1$ -Mcm1 consensus sequence (*E* value ≤ 100). Based on this analysis, we conclude that $\alpha 1$ -Mcm1 positively regulates *MF α 1*, *MF α 2*, *STE3*, *SAG1*, and *YLR040C*. The four former genes had been determined previously to be targets of $\alpha 1$ -Mcm1 regulation (1–3). We analyzed mating in an α strain in which *YLR040C* was deleted but found no obvious defect (data not shown). Given its expression pattern, it is likely that this gene plays a role in mating, perhaps under a special set of conditions. Based on this analysis, we conclude that these five genes comprise the complete set of $\alpha 1$ -Mcm1-regulated genes.

Osmotic Sensitivity Is Regulated by Cell Type. We further investigated the direct repression of the *HOG1* gene by $\alpha 1$ - $\alpha 2$. Hog1 is a mitogen-activated protein kinase that, after activation by high osmolarity, enters the nucleus and phosphorylates several transcriptional regulators that, in turn, control genes that help the cell cope with osmotic stress (for reviews, see refs. 25–27).

Our transcriptional profiling experiments indicated that *HOG1* is repressed 1.6- to 2-fold in a/α cells. To test whether the down-regulation of *HOG1* by $\alpha 1$ - $\alpha 2$ has a biological consequence, we compared the sensitivity of different cell types to 1 M NaCl. As shown in Fig. 7, a/α cells are significantly more salt-sensitive than are a/a cells, consistent with the higher levels of *HOG1* transcript in the latter cell type. To demonstrate that this effect is caused by repression of *HOG1* by $\alpha 1$ - $\alpha 2$, we introduced mutations into the $\alpha 1$ - $\alpha 2$ -binding site upstream of *HOG1* and observed that this mutation increased the salt resistance in the a/α strain (Fig. 7B). Thus, the regulation of *HOG1* by $\alpha 1$ - $\alpha 2$ has a significant biological consequence and shows that osmotic sensitivity is under cell-type control in *S. cerevisiae*.

$\alpha 2$ -Mcm1 repressed genes (α -specific genes)		
MFA1	α -factor	mating
MFA2	α -factor	mating
AGA2	α -cell agglutinin	mating
STE2	α -factor receptor	mating
STE6	α -factor transporter	mating
BAR1	α -factor protease	mating
DPS1 & DPS2	MAT recombinational enhancer	mating type switching
* ASG7	(inhibitor of zygote self-stimulation)	mating
* $\alpha 2$ -Mcm1 regulated only in the presence of pheromone		
$\alpha 1$ -Mcm1 induced genes (α -specific genes)		
MF α 1	α -factor	mating
MF α 2	α -factor	mating
SAG1	α -cell agglutinin	mating
STE3	α -factor receptor	mating
YLR040C	?	?

$\alpha 1$ - $\alpha 2$ regulated genes (a subset of the haploid-specific genes)		
$\alpha 1$ - $\alpha 2$ repressed		
HO	Mating cassette recombination endonuclease	mating type switching
MAT α 1	Mating type transcriptional activator	mating type regulation
MAT α 2	Mating type transcriptional repressor	mating type regulation
GPA1	Pheromone signaling G-protein α	mating
STE4	Pheromone signaling G-protein β	mating
STE18	Pheromone signaling G-protein γ	mating
STE5	Pheromone signaling MAPK scaffold	mating
FUS3	Pheromone signaling MAPK	mating
FAR1	Pheromone-induced cell cycle arrest	mating
MF α 1	α -factor mating pheromone	mating
CCW12	Cell wall mannoprotein involved in agglutination	mating
AXL1	Pheromone-induced morphogenesis and fusion	mating
NEJ1	Non-homologous end-joining DNA repair	DNA repair
RDH54	Recombinational repair	DNA repair
RME1	Transcriptional repressor of meiosis	meiosis
HOG1	Osmotic response MAPK	Osmotic sensing
AMN1	Mitotic exit regulation; haploid-specific clumping	?
ICS2	?	?
DDR2	?	?
$\alpha 1$ - $\alpha 2$ bound, but role unclear		
(TFB1, SSF2)		
(PCF11, IVY1)		
(ZIP2, PDE1)		
(HBT1, WHI4)		
(IME4, COX13)		
(GYPS, RPL36B)		
(YFL012W, HXT10)		

Fig. 8. Summary of the circuit controlled by the *MAT* locus: the complete set of genes regulated by the mating-type transcriptional regulators. In this list are also included sites in the genome significantly occupied by $\alpha 1$ - $\alpha 2$ but in which no genes were observed to be transcriptionally regulated. For each of these sites, two genes flanking the $\alpha 1$ - $\alpha 2$ -bound intergenic sequences are given in parentheses (see *Results* and *Discussion*).

Discussion

In this article we attempt to identify all the target genes directly regulated by the *S. cerevisiae* mating-type locus. The basic regulatory scheme is summarized in Fig. 1, and the direct target genes of each of these regulators are listed in Fig. 8. As summarized in Fig. 8, we identified six α -specific genes (plus a recombinational enhancer) and five α -specific genes. Each of these genes is a direct target of the mating-type-encoded regulators $\alpha 2$ and $\alpha 1$, respectively, and are tightly shut off in the inappropriate cell types. With the possible exception of one α -specific gene (*YLR040C*), all these genes are involved directly in some aspect of mating. We identified 19 genes directly regulated by $\alpha 1$ - $\alpha 2$, 15 of which were known to be directly regulated by $\alpha 1$ - $\alpha 2$. Taken as a whole, the $\alpha 1$ - $\alpha 2$ -regulated genes are involved in a variety of biological processes (Fig. 8). Unlike the α - and α -specific genes (which are tightly regulated), the $\alpha 1$ - $\alpha 2$ -regulated genes show a range of repression values, with some genes being merely turned down in the a/α cell type (see Figs. 2 and 5). Presumably, the strength and consequent occupancy of the binding site at least partly determines this gene-to-gene variation. Another factor could be the spatial relationship between the $\alpha 1$ - $\alpha 2$ -binding site and the major enhancer (upstream activating sequence) of the gene, because this spacing is also known to affect the level of repression (28).

We verified that one process, osmosensing, is indeed under cell-type control (Fig. 7). We show that $\alpha 1$ - $\alpha 2$ regulation of *HOG1* is responsible for rendering a/α cells more sensitive to high NaCl than a/a cells. We currently do not understand why

this regulation exists, but there are several possibilities. For example, osmolarity response is regulated carefully during mating, presumably to reduce the chance of cell lysis during the time the two cells join and fuse (29). Once an **a** and an α cell have mated successfully, **a1- α 2** is formed, and according to this idea, osmotic sensing can be down-regulated. Another possible reason for the *HOG1* down-regulation in an **a**/ α cell is that higher *HOG1* expression in **a** and α cells could be needed to dampen the crosstalk between pheromone and osmotic signaling, which use some of the same signaling components. It is known that **a** and α cells deleted in *HOG1* will activate the pheromone-response pathway inappropriately in response to osmotic stress (30). Elevated expression of *HOG1* might be necessary to prevent this crosstalk in **a** and α cells but not in **a**/ α cells, in which the pheromone pathway is shut off. Finally, it is possible that the cell-type regulation of the osmotic response pathway plays some particular but yet unknown role in the interaction of *S. cerevisiae* and its natural surroundings.

An unexpected finding from this study concerns seven **a1- α 2** sites in the genome that are significantly occupied by **a1- α 2** in **a**/ α cells and show excellent phylogenetic conservation but do not seem to control transcription of an adjacent gene. As discussed in *Results*, there are several possibilities for these observations; perhaps the most intriguing is the existence of unannotated transcripts controlled by **a1- α 2**. There is some

evidence for this possibility for the case of *IME4*. An occupied **a1- α 2** site identified in this study is in a position to regulate a haploid-specific *IME4* antisense transcript identified by Shah and Clancy (31); preliminary evidence suggests that regulation of this transcription by **a1- α 2**, in principle, could control *IME4* production through an antisense mechanism (C. Hongay and G. Fink, personal communication). This idea raises the possibility that the other six bona fide **a1- α 2**-binding sites may also regulate unannotated transcripts, some of which may also produce antisense transcripts.

Finally, we note that although genome-wide ChIP, transcriptional profiling, and phylogenetic footprinting are all powerful techniques for analyzing transcriptional circuits, each has certain limitations. As we show here, the combination of all three approaches can produce an unambiguous and accurate description of a complex transcriptional circuit.

We thank Sarah Green and other members of the Johnson laboratory for helpful suggestions and assistance; Adam Carroll for invaluable help with microarrays; and Mary Clancy and Gerald Fink for communicating unpublished results. This work was supported primarily by National Institutes of Health Grant RO1 GM37049 (to A.D.J.). D.J.G. is a National Science Foundation Predoctoral Fellow, and S.M.O. was supported by a National Institutes of Health training grant, the Markey Program in Biological Sciences, the Herbert W. Boyer Fund, and a University of California San Francisco Chancellor's Fellowship.

- Herskowitz, I., Rine, J. & Strathern, J. (1992) in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, eds. Jones, E. W., Pringle, J. R. & Broach, J. R. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 2, pp. 583–657.
- Sprague, G. F., Jr., & Thorner, J. (1992) in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, eds. Jones, E. W., Pringle, J. R. & Broach, J. R. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 2, pp. 657–744.
- Johnson, A. D. (1995) *Curr. Opin. Genet. Dev.* **5**, 552–558.
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) *Science* **278**, 680–686.
- Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. & Brown, P. O. (2001) *Nature* **409**, 533–538.
- Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000) *Science* **290**, 2306–2309.
- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A. & Johnston, M. (2003) *Science* **301**, 71–76.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. (2003) *Nature* **423**, 241–254.
- Keleher, C. A., Redd, M. J., Schultz, J., Carlson, M. & Johnson, A. D. (1992) *Cell* **68**, 709–719.
- Jin, Y., Zhong, H. & Vershon, A. K. (1999) *Mol. Cell. Biol.* **19**, 585–593.
- Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O. & Davis, R. W. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 13057–13062.
- O'Rourke, S. M. & Herskowitz, I. (2004) *Mol. Biol. Cell* **15**, 532–542.
- Bennett, R. J., Uhl, M. A., Miller, M. G. & Johnson, A. D. (2003) *Mol. Cell. Biol.* **23**, 8189–8201.
- Bailey, T. L. & Elkan, C. (1994) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36.
- Bailey, T. L. & Gribskov, M. (1998) *Bioinformatics* **14**, 48–54.
- Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. (2001) *Nat. Genet.* **28**, 327–334.
- Wu, X. & Haber, J. E. (1996) *Cell* **87**, 277–285.
- Szeto, L., Fafalios, M. K., Zhong, H., Vershon, A. K. & Broach, J. R. (1997) *Genes Dev.* **11**, 1899–1911.
- Zhong, H., McCord, R. & Vershon, A. K. (1999) *Genome Res.* **9**, 1040–1047.
- Buck, M. J. & Lieb, J. D. (2004) *Genomics* **83**, 349–360.
- Fujita, A., Oka, C., Arikawa, Y., Katagai, T., Tonouchi, A., Kuhara, S. & Misumi, Y. (1994) *Nature* **372**, 567–570.
- Valencia, M., Bentele, M., Vaze, M. B., Herrmann, G., Kraus, E., Lee, S. E., Schar, P. & Haber, J. E. (2001) *Nature* **414**, 666–669.
- Whiteway, M., Hougan, L., Dignard, D., Thomas, D. Y., Bell, L., Saari, G. C., Grant, F. J., O'Hara, P. & MacKay, V. L. (1989) *Cell* **56**, 467–477.
- Nagaraj, V. H., O'Flanagan, R. A., Bruning, A. R., Mathias, J. R., Vershon, A. K. & Sengupta, A. M. (2004) *BMC Genomics* **5**, 59.
- Gustin, M. C., Albertyn, J., Alexander, M. & Davenport, K. (1998) *Microbiol. Mol. Biol. Rev.* **62**, 1264–1300.
- Sprague, G. F., Jr. (1998) *Genes Dev.* **12**, 2817–2820.
- O'Rourke, S. M., Herskowitz, I. & O'Shea, E. K. (2002) *Trends Genet.* **18**, 405–412.
- Johnson, A. D. & Herskowitz, I. (1985) *Cell* **42**, 237–247.
- Nelson, B., Parsons, A. B., Evangelista, M., Schaefer, K., Kennedy, K., Ritchie, S., Petryshen, T. L. & Boone, C. (2004) *Genetics* **166**, 67–77.
- O'Rourke, S. M. & Herskowitz, I. (1998) *Genes Dev.* **12**, 2874–2886.
- Shah, J. C. & Clancy, M. J. (1992) *Mol. Cell. Biol.* **12**, 1078–1086.