Supplementary Information for:

Structure of the Transcriptional Network Controlling White-Opaque Switching in *Candida albicans*

Aaron D. Hernday, Matthew B. Lohse, Polly M. Fordyce, Clarissa J. Nobile, Joseph L. DeRisi, and Alexander D. Johnson

Supplementary Information

Contents

Supplemental Information

Roles of the specific regulators in white-opaque switching

To better understand the functional logic of the ChIP-chip enrichment patterns, we performed genome-wide transcriptional profiling of opaque strains deleted for *WOR2*, *CZF1, EFG1* or *AHR1*. First, we consider Wor2. In order to determine the logic of Wor2 binding in opaque cells, we compared the expression profiles of two opaque strains in which *WOR1* was constitutively expressed; one carried the wild-type *WOR2* alleles while the other was a homozygous *wor2* deletion strain. Deletion of *WOR2* resulted in the differential expression (≥ 2-fold) of 231 genes (148 upregulated, 83 downregulated), 122 of which are differentially expressed by at least twofold between wild-type white and opaque cells. Of the genes whose expression was increased in the *wor2* deletion strain, many are white-enriched, while a majority of the genes whose expression was decreased are opaque-enriched. Based on Wor2 ChIP-chip enrichment upstream of each target gene, Wor2 is formally a direct activator of thirteen opaque-enriched transcripts and a direct repressor of nine white-enriched transcripts (Fig. S1D).

Czf1 functions as an activator in white cells, but is formally both an activator and a repressor in opaque cells. In opaque cells, 158 genes are differentially expressed more than twofold in response to *CZF1* deletion (107 upregulated, 51 downregulated), 26 of which are directly bound by Czf1 (Fig. S1E).

Efg1 is formally both an activator and repressor in both the white and opaque cell types. Of the 171 genes 171whose expression is affected by *EFG1* deletion, 30 are bound by Efg1. Based on this analysis, we suggest that Efg1 is both a positive and a negative regulator of transcription (Fig. S1F). We note that the two most Efg1-

dependent opaque transcripts are *CZF1* and *WOR3*, suggesting that these three genes may form a feed-forward loop that gives rise to the 100-fold increase in *WOR3* expression observed in opaque cells.

As with Efg1, Ahr1 is formally both an activator and a repressor in white and opaque cell types. In opaque cells, seventeen of the 74 genes that are differentially regulated (19 upregulated, 55 downregulated) in response to *AHR1* deletion are associated with Ahr1 binding (Fig. S1G).

We note that the gene counts for this section as well as the equivalent analysis for white cells in the main text exclude the deleted gene from consideration (i.e. the *AHR1* deletion analysis of opaque cells does not count *AHR1* as one of the genes differentially regulated). Furthermore, we excluded genes whose values were either rounded up to 1.00 (log2) or down to -1.00 (log2) when presented; rounded to 2 decimal places in File S1.

Supplemental Experimental Procedures

Strain construction

A list of strains used in this study can be found in Table S3A. Deletion strains were generated as described previously (Hernday *et al.*, 2010). The *WOR1*, *WOR2*, *WOR3*, *CZF1*, and *EFG1* deletion strains have been previously reported (Zordan *et al.*, 2006; Lohse *et al.*, 2013). C-terminal myc-tagged transcription factor strains were generated using plasmid pADH34 (Nobile *et al.*, 2009) and PCR-directed genomic integration as described previously (Hernday *et al.*, 2010). Switching competent versions of the previously reported *BRG1*, *CRZ2*, and *AAF1* deletions (Homann *et al.*, 2009) were created using the pJD1 Mating Type Locus knockout cassette as previously described (Lin *et al.*, 2013).

Plasmid construction

A list of plasmids used in this study can be found in Table S3B. A list of oligonucleotides used in this study can be found in Table S3C.

Ectopic *WOR1* expression strains were generated by cloning the *WOR1* open reading frame behind the p*MET3* upstream region in pADH33 (Lohse *et al.*, 2013). Following digestion with NcoI, the SAT1-marked ectopic expression construct containing *WOR1* was integrated at the *RP10* locus and confirmed by colony PCR.

CUG to AGC or UCC (SER) codon substitutions were performed on *CZF1*, *WOR2*, and *EFG1* ORFs to allow for expression of the correct protein sequence in *E. coli*. Codon substitutions were introduced by stitching PCR using the oligonucleotides listed in Table S3C. Briefly, mutagenic primers that overlap each CUG codon were used to amplify the intervening DNA sequences between each of the CUG codons. The

resulting codon-optimized fragments were assembled by stitching PCR, digested with SacII and XbaI, and cloned into pDAL649 to generate pADH42 (Efg1) and pADH43 (Czf1). The sequence of pDAL649 is available at GenBank (http://www.ncbi.nlm.nih.gov/genbank), accession # KC202164. Codon-optimized *WOR2* was TOPO-cloned into the pCR-BluntII plasmid (Invitrogen) to generate pRZ67.

The pLIC-H3C variant of the pET28b plasmid was used for protein expression (Hammon *et al.*, 2009). Full length codon optimized Czf1 and Efg1 were PCR-amplified from pADH43 and pADH42, respectively, adding 5' XmaI sites and a 3' XhoI site (Czf1) or XmaI site (Efg1) for cloning into pLIC-H3. Cloning was conducted in the XL-1 Blue background, correct orientation and sequence of the insert was verified through a combination of test digests and sequencing. Plasmids were transformed into the BL21 background for protein expression.

A Wor2 fragment corresponding to the predicted globular region containing the zinc cluster domain (231-343aa) was PCR-amplified from pRZ67 and cloned into a previously reported modified version of pET28b containing a N-terminal 6-His and Maltose Binding Protein (MBP) affinity tags as well as a PreScission protease site (Lohse *et al.*, 2010). The Wor2 fragment was subcloned into this plasmid using the NheI and XhoI restriction sites, giving pRZ97. The plasmid was transformed into the BL21 background for protein expression.

Switching Assays

Plate based quantitative white to opaque switching assays were performed as previously described (Miller and Johnson, 2002; Zordan *et al.*, 2007) using SD+aa+Uri plates at room temperature. Plates were scored for total number of colonies and

switching events (white colonies with one or more opaque sectors and opaque colonies).

Microarrays

Cultures for gene expression microarray analysis were harvested during mid-log phase by centrifugation. Pellets were snap-frozen in liquid nitrogen prior to RNA extraction. Total RNA was extracted by the hot acid phenol method (Hernday *et al.*, 2010) or by using the RiboPure-Yeast RNA kit (Ambion). Total RNA was reverse-transcribed into cDNA and coupled to Cy3 and Cy5 dyes as described previously (Hernday *et al.*, 2010). Equal amounts of white vs. opaque or wild-type vs. mutant cDNA were hybridized to custom-designed Agilent 8x15k microarrays (AMADID #020166), which contain at least two probes per open reading frame, and scanned using a Genepix 4000B scanner (Axon/Molecular Devices). Data was extracted using GenePix Pro version 5.1 and normalized by Global Lowess normalization using the Goulphar script (Lemoine *et al.*, 2006) for R (The R foundation for Statistical Computing). Transformations (i.e. white vs. opaque or wild-type vs. deletion strain) were performed prior to the extraction of median differential expression values for each ORF. Differential expression was determined by using a 2-fold cutoff. Two biological replicates were performed for each condition. Raw gene expression array data are available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo, accession # GSE42134). The analyzed gene expression array data are included in File S1.

With the exception of, profiling of deletion strains was conducted using strains where both copies of the transcriptional regulator of interest had been deleted and a control strain where the *HIS1* and *LEU2* markers were added back at a different locus.

WOR2 is normally required for the formation of opaque cells, presenting a problem for the profiling of the opaque cell type when *WOR2* was deleted. The requirement for *WOR2* can be overcome if *WOR1* is ectopically overexpressed, thus to characterize a *WOR2* deletion opaque we compared the expression profiles of two opaque strains in which *WOR1* was constitutively expressed; one carried the wild-type *WOR2* alleles while the other was a homozygous *wor2* deletion strain.

ChIP-chip

Cultures for chromatin immunoprecipitation were harvested during mid-log phase by centrifugation. Chromatin immunoprecipitation, strand-displacement amplification, dye-coupling and hybridization to a custom 1x244k Agilent tiling microarray (AMADID #016350) was performed as described previously (Hernday *et al.*, 2010). At least two biological replicates were performed for each condition except for the Ahr1-myc ChIP-chip, which was performed once. We used previously reported custom-designed polyclonal antibodies (Bethyl Laboratories) raised against Wor1 (Zordan *et al.*, 2007), Wor2, Czf1, and Efg1 (Lohse and Johnson, 2010). An anti-c-Myc monoclonal antibody (Invitrogen AHO0062) was used for C-terminally Myc-tagged Ahr1 as well as the previously reported Wor3 ChIP-chip dataset (Lohse *et al.*, 2013). We used antibodies raised against the sequences CQANQSASTVAKEEK (residues 541-554) for Efg1 and CKVLRGIVEYRSK (residues 374-385) for Czf1. For Wor2, two antibodies raised against the sequences CYSPNSPYSLPTR (residues 57-68) and CSAVINRVSVADLLK (residues 433-446) were used. Re-analysis of Wor1 ChIP binding data used datasets for both the QVLDKQLEPVSRRPHERER (residues 26-44) and DDAVGNSSGSYYTGT (residues 771-785) peptides that were previously reported (Zordan *et al.*, 2007). Arrays

were scanned using a Genepix 4000B scanner (Axon/Molecular Devices) and processed as described previously (Nobile *et al.*, 2012) with the following exception: Minimum enrichment cut-offs for MochiView peak detection were set to 0.58 for the tagged/wild-type arrays, and 0.27 for the untagged/deletion arrays. The detected peaks were further filtered by subtraction of likely artifactual peaks, based on the fact that these loci showed variable but substantial enrichment in the majority of deletion control ChIP-chip experiments that were performed with antibodies against a deleted target (Table S3D). In cases where two different antibodies were use (e.g. Wor2 and reanalysis of Wor1), we considered sites present in one or both datasets as valid. The list of bound genes for each regulator was generated by assigning called peaks to the 5' intergenic regions of each ORF using MochiView. ChIP enrichment values for each target ORF were selected from the most highly-enriched peak within each bound intergenic region. Raw ChIP-chip data are available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo, accession# GSE42837). The analyzed ChIP-chip data are included in File S1. Plots of the areas of peak enrichment for each regulator are included in Files S2, S5, and S6.

Intergenic region and binding site breakdowns

Unless otherwise noted, all of the analysis described was performed in MochiView. Creation of the intergenic region location set used in this study has been previously described (Lohse *et al.*, 2013). We then created location sets for the intergenic regions bound by each regulator using the "Merge Location Set (Union)" function with "Only keep locations intersected by all contributing location sets" selected for the regulator peaks and intergenic region location sets. We then cleaned up each set

by subtracting out the transcribed ORF location set using the “Merge Location Set (Subtraction)” function.

We then used the “Merge Location Set (Union)” function to create location sets bound by all six regulators, then all five, and so on until we had created sets for all possible combinations. We then used the “Merge Location Set (Subtraction)” function to subtract out all of the higher order events out from lower ones (for example, subtracting out all combinations of five or six regulators bound from the sets with four bound so that the four bound sets contained only instances where four specific regulators were bound).

The binding site breakdown followed the general approach used for the breakdown of the intergenic regions, with the following changes. The peak calls for each regulator were not combined with intergenic regions; instead the combinations were made and cleaned using only the peak location sets.

White Cell Czf1 Motif Analysis

The 39 Czf1 binding sites (each 500bp) were examined for the presence of the ChIP-chip and MITOMI derived Czf1 motifs as well as an associated Efg1 motif using MochiView (Homann and Johnson, 2010). Six Czf1 binding sites had two distinct instances of clear Czf1 motifs; for these locations both instances were scored, bringing the total of motifs analyzed to 45. Czf1 motifs were characterized as “MITOMI” or “ChIP-chip” based on an analysis of the positions 3, 5, 8, and 9 from the motif presented in Fig. 5A, with the bases T, G, G, and C at the respective positions corresponding to MITOMI and bases A, C, not G, and T corresponding to ChIP-chip. The motif was then assigned to whichever of these groups the majority of these 4 positions corresponded

to. In cases where there were equal numbers of bases falling into each category or in cases where there were not at least two bases in one of the categories, the motif was characterized as “other”. Adjacent Efg1 motifs with the sequence “TGCAT” were characterized as “consensus” and sites which deviated from this were characterized as “other”.

Proximal gene characteristic statistical analyses

Lists of genes downstream of a class of intergenic region (e.g. bound by four regulators) we developed in MochiView using the “Export Gene Proximity Assignments” utility. Maximum number of genes was set to two, maximum search distance was set to 1bp, and the “maximum one gene per strand” option was selected to ensure that only genes whose transcription start site was downstream of the intergenic region were assigned to that intergenic region.

We then determined the length of all intergenic regions in a given set (e.g. bound by four regulators, all intergenics) and calculated the mean length of that set. Control sets were created by taking a subset of intergenic regions, starting from the longest, that had the same mean length as the specific set we were looking at. Intergenic regions in the location set being examined were not excluded from the control set. With defined experimental and control sets, we then determined the total number of locations in each set and the number of locations that passed a specific criteria (e.g. twofold differential regulation upon switching to opaque as determined by RNA-seq). We used both the previously published RNA-seq data (Tuch *et al.*, 2010) as well as recently published microarray data for comparing binding to white-opaque regulation (Lohse *et al.*, 2013). We used the previously published Wor3 deletion microarray data (Lohse *et*

al., 2013) as well as the newly generated deletion microarray data for Efg1, Ahr1, Czf1, and Wor2 for the comparison of binding by a specific regulator and mis-regulation upon deletion of that regulator. Statistical significance was then determined using the Hypergeometric Distribution Test in Excel (Microsoft). This analysis was conducted for datasets where the deleted gene (i.e. *EFG1*) and its promoter were included or excluded from all calculations.

Binding overlap enrichment analyses

We developed three tests to determine if the observed overlap in the set of intergenic regions bound by the regulators in opaque cells was more than would be expected by chance. Each approach makes different assumptions about the relative probabilities of binding to intergenic regions of variable length. For each approach, we limited our analysis to intergenic regions of at least 200bp in size. We note that this cutoff excluded only 17 regions bound by at least one regulator and of these 17, only 2 were bound by multiple regulators. This cutoff left us with 2,954 intergenic regions of average length 1,279bp to consider.

Our first approach treated all intergenic regions as equally probable targets for regulatory binding. For each regulator, r, we calculated the fraction of bound intergenic regions, fr, as the number of intergenic regions bound by the regulator, nr, divided by the total number of intergenic regions, N, in the genome (Eq. 1).

$$f_r = \frac{n_r}{N} \qquad \text{(Eq. 1)}$$

Using equation 1, we calculated the expected number of intergenic regions, E, bound by a set of regulators $[r1, r2, \ldots, rk]$ as the total number of intergenic regions in the genome times the product of the fraction of the bound intergenic regions for each regulator, fi (Eq. 2).

$$E(r_1, r_2, ..., r_k) = N \times \prod_{i=1}^{k} f_i \qquad \text{(Eq. 2)}$$

Our second approach incorporated additional parameters to account for the fact that some regulators bind collections of intergenic regions with mean length that varies from the mean length of all intergenic regions. To account for this skew, for each regulator, *r*, we calculated a length-bias correction, cr, as the fraction of the mean length, $\bar{lr}$, of the intergenic regions bound by the regulator, divided by the mean length, $\bar{L}$, of all intergenic regions in the genome (Eq. 3).

$$c_r = \frac{\bar{l_r}}{\bar{L}} \qquad \text{(Eq. 3)}$$

Using equation 3, we calculated the expected number of intergenic regions bound by a set of regulators as the total number of intergenic regions in the genome, N, times the fraction of the intergenic regions bound by each regulator, fi, weighted by its length-bias correction, ci (Eq. 4).

$$E(r_1, r_2, ..., r_k) = N \times \prod_{i=1}^{k} f_i c_i \qquad \text{(Eq. 4)}$$

Our third approach incorporated parameters to account for the fact that short intergenic regions provide a relatively smaller evolutionary target, and thus are less likely by chance than longer intergenic regions to acquire active binding motifs. For each intergenic region, j, we calculated the region's weighted length, sj, as the product of the region's length, lj, and the inverse of the mean length of all intergenic regions, $\bar{L}$. (eq. 5)

$$s_j = l_j \times \bar{L}^{-1} \qquad \text{(Eq. 5).}$$

For each regulator, r, we calculated a length-corrected fraction of bound intergenic regions, $f'r$, as the sum of weighted lengths for all intergenic regions bound by r, divided by the sum of weighted lengths for all intergenic regions in the genome (Eq. 6).

$$f'_r = \frac{\sum_{j}^{j \in r} s_j}{\sum_{j}^{all} s_j} \qquad \text{(Eq. 6)}$$

Using equation 6, we calculated the number of intergenic regions expected to be jointly bound by a set of regulators, [r_1, r_2, . . ., r_k], as the product of the number of intergenic regions in the genome and each regulator's length-corrected binding fraction (Eq. 7).

$$E(r_1, r_2, ..., r_k) = N \times \prod_{i=1}^{k} f_i' \qquad \text{(Eq. 7)}$$

Using all three methods described above (Eqs., 2, 4, 7), we calculated the expected number of bound intergenic regions for the regulators in the white-opaque switching network, and compared these estimates to the observed number of intergenic binding events.

ChIP-chip enrichment versus number of transcriptional regulators bound analysis

Using the “Map data set(s) to location set” function in MochiView, we mapped enrichment values for all called peaks for the six regulators onto the corresponding intergenic regions. If multiple peaks for a given regulator corresponded to a single intergenic region, the maximum enrichment value was used. If multiple regulators had binding sites corresponding to a single intergenic region, the maximum enrichment value for each of the regulators was assigned to that region. The enrichment values for each intergenic region bound by at least one regulator were then exported from MochiView and enrichment by number of regulators bound visualized using GraphPad Prism (GraphPad Software Inc.).

Motif enrichment at binding sites

Examination of the ability of the various regulator motifs to explain the *in vivo* binding sites was based on previously described procedures (Lohse *et al.*, 2010; Cain *et al.*, 2012; Nobile *et al.*, 2012; Lohse *et al.*, 2013). The analysis was performed using the ChIP-chip developed PSWM for Ahr1 and MITOMI 2.0 PSAMs for the other regulators. The PSWM used for analysis of Ahr1 binding and the PSAMs used for the analysis of binding by the other five regulators are included in File S3.

The motif enrichment at specific binding site analysis was limited to sites where only a single regulator was bound as well as all combinations of binding sites with at least five observed instances. Random control group location sizes were based on the mean location set size for all instances where a given number of regulators (e.g. four) were bound. Creation of control groups followed previously described methods (Lohse *et al.*, 2010; Cain *et al.*, 2012; Nobile *et al.*, 2012; Lohse *et al.*, 2013).

In brief, we used MochiView to assign the maximum score for each motif for each location in each location set. We then used Excel (Microsoft) to calculate the fraction of each experimental or control location set with a motif score greater than a specific threshold (specific score for one or more motifs).

Transcriptional regulator list

We based our transcriptional regulator list on the set of 283 transcriptional regulators listed in Dataset S1 originating from Homann et al. (Homann *et al.*, 2009). We manually added Wor1 (Orf19.4884) and Wor3 (Orf19.467) to this list. We removed MTLα2 (Orf19.10708) from this list because the experiments were performed in MTLa/a cells. These changes bring the number of genes in our Transcriptional Regulator list to 284.

Supplemental References

Cain, C. W., Lohse, M. B., Homann, O. R., Sil, A., and Johnson, A. D. (2012) A conserved transcriptional regulator governs fungal morphology in widely diverged species. *Genetics* **190**: 511-521.

Hammon, J., Palanivelu, D. V., Chen, J., Patel, C., and Minor, D. L., Jr. (2009) A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. *Protein Sci* **18**: 121-133.

Hernday, A. D., Noble, S. M., Mitrovich, Q. M., and Johnson, A. D. (2010) Genetics and molecular biology in Candida albicans. *Methods Enzymol* **470**: 737-758.

Homann, O. R., Dea, J., Noble, S. M., and Johnson, A. D. (2009) A phenotypic profile of the Candida albicans regulatory network. *PLoS Genet* **5**: e1000783.

Homann, O. R., and Johnson, A. D. (2010) MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol* **8**: 49.

Lemoine, S., Combes, F., Servant, N., and Le Crom, S. (2006) Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* **7**: 467.

Lin, C. H., Kabrawala, S., Fox, E. P., Nobile, C. J., Johnson, A. D., and Bennett, R. J. (2013) Genetic control of conventional and pheromone-stimulated biofilm formation in Candida albicans. *PLoS Pathog* **9**: e1003305.

Lohse, M. B., Hernday, A. D., Fordyce, P. M., Noiman, L., Sorrells, T. R., Hanson-Smith, V.*, et al.* (2013) Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains. *Proc Natl Acad Sci USA* **110**: 7660-7665.

Lohse, M. B., and Johnson, A. D. (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* **78**: 331-343.

Lohse, M. B., Zordan, R. E., Cain, C. W., and Johnson, A. D. (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proc Natl Acad Sci USA* **107**: 14105-14110.

Miller, M. G., and Johnson, A. D. (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* **110**: 293-302.

Nobile, C. J., Fox, E. P., Nett, J. E., Sorrells, T. R., Mitrovich, Q. M., Hernday, A. D.*, et al.* (2012) A recently evolved transcriptional network controls biofilm development in Candida albicans. *Cell* **148**: 126-138.

Nobile, C. J., Nett, J. E., Hernday, A. D., Homann, O. R., Deneault, J. S., Nantel, A.*, et al.* (2009) Biofilm matrix regulation by *Candida albicans* Zap1. *PLoS Biol* **7**: e1000133.

Tuch, B. B., Mitrovich, Q. M., Homann, O. R., Hernday, A. D., Monighetti, C. K., De La Vega, F. M., and Johnson, A. D. (2010) The Transcriptomes of Two Heritable Cell Types Illuminate the Circuit Governing Their Differentiation. *PLoS Genet* **6**: e1001070.

Zordan, R. E., Galgoczy, D. J., and Johnson, A. D. (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci USA* **103**: 12807-12812.

Zordan, R. E., Miller, M. G., Galgoczy, D. J., Tuch, B. B., and Johnson, A. D. (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* **5**: e256.

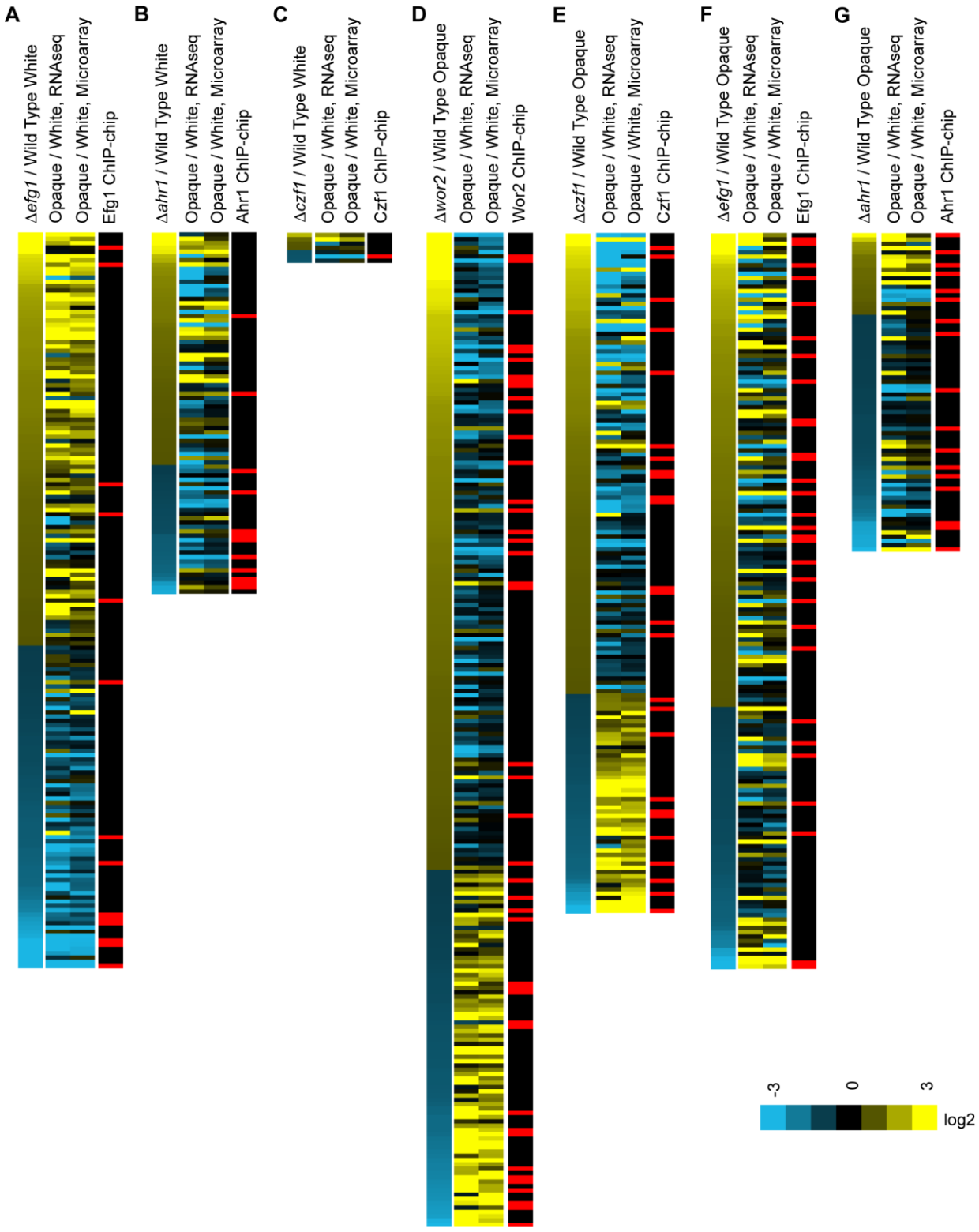
A
Δefg1 / Wild Type White
Opaque / White, RNAseq
Opaque / White, Microarray
Efg1 ChIP-chip
B
Δahr1 / Wild Type White
Opaque / White, RNAseq
Opaque / White, Microarray
Ahr1 ChIP-chip
C
Δczf1 / Wild Type White
Opaque / White, RNAseq
Opaque / White, Microarray
Czf1 ChIP-chip
D
Δwor2 / Wild Type Opaque
Opaque / White, RNAseq
Opaque / White, Microarray
Wor2 ChIP-chip
E
Δczf1 / Wild Type Opaque
Opaque / White, RNAseq
Opaque / White, Microarray
Czf1 ChIP-chip
F
Δefg1 / Wild Type Opaque
Opaque / White, RNAseq
Opaque / White, Microarray
Efg1 ChIP-chip
G
Δahr1 / Wild Type Opaque
Opaque / White, RNAseq
Opaque / White, Microarray
Ahr1 ChIP-chip
-3
0
3
log2

Figure S1. Transcriptional profiling of regulator deletions in white or opaque cells cells. (A) *efg1* deletion in white cells (B) *ahr1* deletion in white cells (C) *czf1* deletion in white cells (D) *wor2* deletion in opaque cells (E) *czf1* deletion in opaque cells (F) *efg1* deletion in opaque cells (G) *ahr1* deletion in opaque cells. Transcriptional changes in the respective white or opaque deletion strains relative to the white or opaque parent strain (left lane). All genes differentially regulated at least twofold upon deletion of the given regulator are shown. Opaque or white enrichment of the same genes in a wild-type background (middle lanes). Efg1, Ahr1, Wor2 or Czf1 binding *in vivo* in white or opaque cells as determined by ChIP-chip is indicated in red in the rightmost lane. RNA-seq opaque or white enrichment values are taken from Tuch *et al*. (Tuch *et al.*, 2010) and microarray opaque or white enrichment values are taken from Lohse *et al.* (Lohse *et al.*, 2013); all other data is from this study.

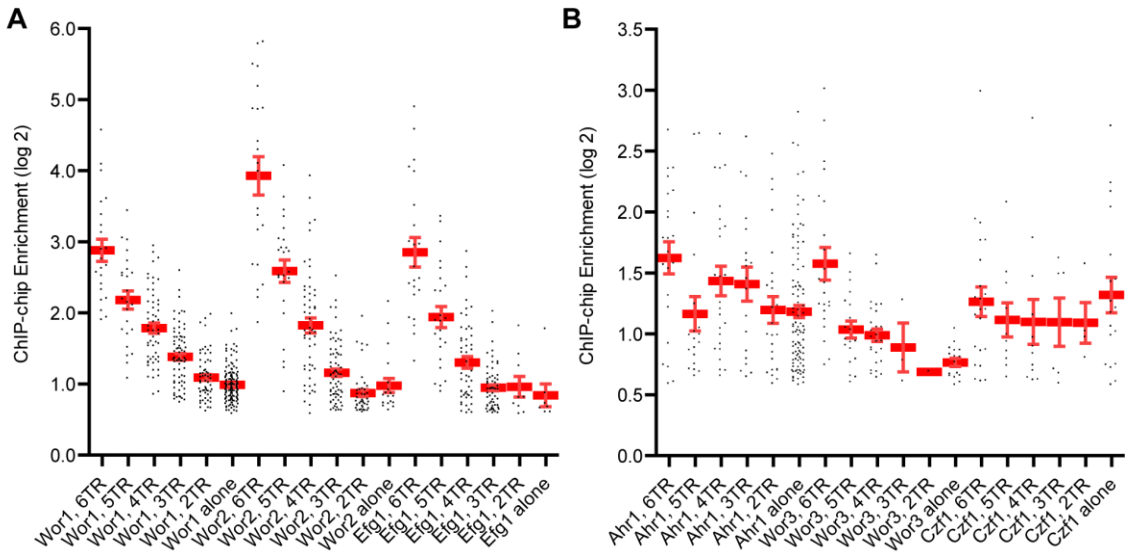

Figure S2. Comparison of binding enrichment and number of regulators bound in opaque cells. Enrichment for binding of Wor1, Efg1, Wor2, and Wor3 is greatest at intergenic regions with more core regulators bound. (A) Wor1, Efg1, Wor2, (B) Ahr1, Wor3, and Czf1. ChIP-chip enrichment values (log_2) are plotted on the y-axis, number of regulators bound at each intergenic region for each group are indicated on the x-axis. The mean of each set is indicated by the bold red line; error bars reflect the standard error of the mean.

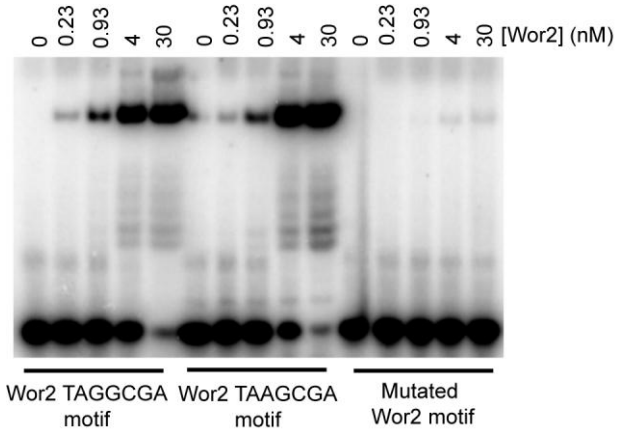

Figure S3. Validation of the Wor2 binding site. EMSA for DNA fragments containing two versions of the MITOMI Wor2 motif (TAGGCGA, TAAGCGA) or a mutated version of the motif performed using a 6-His-MBP-Wor2 (231-343aa) construct.

Supplemental Table Captions

Table S1. Breakdown of regulator binding patterns in white and opaque cells. (A) Breakdown of intergenic regions bound by Czf1, Efg1, Ahr1, or combinations of the three in white cells. (B) Breakdown of peak areas of Czf1, Efg1, and Ahr1 binding in white cells, based on 500bp called peaks. (C) Breakdown of Czf1 binding sites in opaque cells based on the characteristics of the Czf1 and Efg1 motifs (if any) at each site. (D) Transcriptional regulators bound by Czf1, Efg1, or Ahr1 in white cells, as well as previously published RNA-seq (Tuch *et al.*, 2010), and microarray analysis (Lohse *et al.*, 2013) of these genes in opaque versus white cells, values are on a log2 scale. (E) Breakdown of intergenic regions bound by Wor1, Wor2, Wor3, Czf1, Efg1, Ahr1, or combinations of the six in opaque cells. Wor1 and Wor3 binding data have been previously reported (Zordan *et al.*, 2007; Lohse *et al.*, 2013). (F) Breakdown of peak areas of Wor1, Wor2, Wor3, Czf1, Efg1, and Ahr1 binding in opaque cells, based on 500bp called peaks. Wor1 and Wor3 binding data have been previously reported (Zordan *et al.*, 2007; Lohse *et al.*, 2013). (G) Transcriptional regulators bound by Wor1, Wor2, Wor3, Czf1, Efg1, or Ahr1 in opaque cells, as well as previously published RNA-seq (Tuch *et al.*, 2010), and microarray analysis (Lohse *et al.*, 2013) of these genes in opaque versus white cells, values are on a log2 scale. Wor1 and Wor3 binding data have been previously reported (Zordan *et al.*, 2007; Lohse *et al.*, 2013).

Table S2. Further analysis of the observed regulator binding subsets. (A) Comparison of expected and observed levels of overlap between binding of Wor1, Wor2, Wor3, Czf1, Efg1, and Ahr1 in opaque cells. Values for broad categories (i.e. three or more

regulators bound) and specific combinations (i.e. Wor1+Wor2+Wor3) are shown. When calculating the number of expected binding events**,** the lower level events (i.e. where two regulators are bound also include the probability of the higher order events (i.e. with three or more regulators bound) that encompass the lower order event. For example, the “both Wor1 and Wor2 bound” set includes the probabilities for 15 higher order binding events like “Wor1, Wor2, and Wor3 bound” and “all 6 regulators bound”. As such, the expected numbers of intergenic regions (IR) bound that is labeled as “Basic” represent a conservative estimate for each method. Recalculation of the expected values with the higher order events removed has been included and is labeled as “Corrected”. We note that the corrected values do not differ significantly from the original calculations. The three methods for estimating the number of expected overlapping binding events are discussed in the Supplemental Experimental Procedures. (B) Examination of the ability of motifs for the six regulators to explain specific subsets of ChIP-chip binding events. The percent of locations in a given set passing a threshold (an instance of a given motif scoring at least as high as the score giving the indicated false positive percentage in a 500bp control set of locations) is indicated for each combination of locations and motifs. The percent of locations in an equivalently sized control set is also indicated for each motif. The motif enrichment at specific binding site analysis was limited to sites where only a single regulator was bound as well as combinations of regulators that were observed in at least five instances. The number of locations in each set is also indicated. (C) Statistical analysis of the number of white or opaque regulated genes adjacent to an intergenic region bound by a given number of regulators as compared to an equivalently sized control set

of intergenic regions. Control sets were created by taking a subset of intergenic regions, starting from the longest, that had the same mean length as the specific set we were looking at. Intergenic regions in the location set being examined were not excluded from the control set. With defined experimental and control sets, we then determined the total number of locations in each set and the number of locations that passed a specific criteria (e.g. twofold differential regulation upon switching to opaque as determined by RNA-seq). We used both the previously published RNA-seq data (Tuch *et al.*, 2010) as well as recently published microarray data (Lohse *et al.*, 2013) for comparing regulator binding to white-opaque regulation. Statistical significance was determined using the Hypergeometric Distribution Test in Excel. (D) Statistical analysis of the number of genes adjacent to an intergenic region bound by a specific regulators that were mis-regulated at least twofold when that regulator was deleted. We used the previously published Wor3 deletion microarray data (Lohse *et al.*, 2013) as well as the newly generated deletion microarray data for Efg1, Ahr1, Czf1, and Wor2 in opaque cells. Control sets were created by taking a subset of intergenic regions, starting from the longest, that had the same mean length as the specific set in question. Intergenic regions in the location set being examined were not excluded from the control set. Analysis was performed for datasets where the deleted gene and its intergenic region was included (top) or excluded (bottom) from consideration. Statistical significance was determined using the Hypergeometric Distribution Test in Excel. (E) White-opaque switch frequencies for deletion of *BRG1*, *CRZ2*, or *AAF1* as well as the equivalent control strain.

Table S3. (A) List of strains used in this study. (B) List of plasmids used in this study. (C) List of oligonucleotide sequences used in this study, including ones for the Wor1-, Wor2-, Efg1-, and Czf1-specific libraries of oligonucleotides containing systematic substitutions of all possible nucleotides at each position within the Wor1, Wor2, Efg1, and Czf1 *cis*-regulatory sequences. Also includes list of all 740 oligonucleotide sequences for the revised MITOMI 2.0 Random 8mer Library used in this study. (D) List of "red-flagged" locations in ChIP-chip data. Called peaks were filtered by subtraction of this set of likely artifactual peaks, based on the fact that these loci showed variable but substantial enrichment in many deletion (control) ChIP-chip experiments.

Supplemental File Captions

File S1. Compilation of microarray, RNA-seq, and ChIP-chip data presented in this study. From left to right in the Excel spreadsheet, columns are as follows. (A) Orf19 number designation is based on the Candida Genome Database (CGD). (B) Gene name, where applicable. (C) CGD description of the gene. (D) Whether the gene is a transcriptional regulator, based on Homann *et al.* (Homann *et al.*, 2009) represents yes. (E) Whether the upstream region for the gene is bound by Ahr1, Czf1, or Efg1 in white cells; "1" represents yes. (F) Whether the upstream region for the gene is bound by Ahr1, Czf1, Efg1, Wor1, Wor2, or Wor3 in opaque cells; "1" represents yes. (G) Whether the gene was excluded from our analysis based on a lack of observed transcription in previously published RNA-seq experiments (Tuch *et al.*, 2010); "1" represents exclusion. (H) Maximum Czf1 enrichment in the upstream region for the gene in a white cell, values are on a log2 scale. (I) Maximum Efg1 enrichment in the upstream region for the gene in a white cell; values are on a log2 scale. (J) Maximum Ahr1 enrichment in the upstream region for the gene in a white cell; values are on a log2 scale. (K) Maximum Wor1 enrichment in the upstream region for the gene in an opaque cell, based on reanalysis of previously published data (Zordan *et al.*, 2007); values are on a log2 scale. (L) Maximum Wor2 enrichment in the upstream region for the gene in an opaque cell; values are on a log2 scale. (M) Maximum Czf1 enrichment in the upstream region for the gene in an opaque cell; values are on a log2 scale. (N) Maximum Efg1 enrichment in the upstream region for the gene in an opaque cell; values are on a log2 scale. (O) Maximum Wor3 enrichment in the upstream region for the gene in an opaque cell; values are on a log2 scale, data from Lohse *et al*. (Lohse *et al.*, 2013). (P)

Maximum Ahr1 enrichment in the upstream region for the gene in an opaque cell; values are on a log2 scale. (Q) Whether RNA-seq data for a gene is considered trustworthy. (R) Previously published RNA-seq of opaque versus white cells (Tuch *et al.*, 2010); values are on a log2 scale. (S) Previously published microarray analysis of opaque versus white cells (Lohse *et al.*, 2013); values are on a log2 scale. (T) Microarray analysis of a white *czf1* deletion strain versus wild-type white cells; values are on a log2 scale. (U) Microarray analysis of a white *efg1* deletion strain versus wild-type white cells; values are on a log2 scale. (V) Microarray analysis of a white *wor3* deletion strain versus wild-type white cells; values are on a log2 scale, data from Lohse *et al*. (Lohse *et al.*, 2013). (W) Microarray analysis of a white *ahr1* deletion strain versus wild-type white cells; values are on a log2 scale. (X) Microarray analysis of an opaque *wor2* deletion strain with ectopically expressed Wor1 versus wildtype opaque cells with ectopically expressed Wor1; values are on a log2 scale. (Y) Microarray analysis of an opaque *czf1* deletion strain versus wildtype opaque cells; values are on a log2 scale. (Z) Microarray analysis of an opaque *efg1* deletion strain versus wildtype opaque cells; values are on a log2 scale. (AA) Microarray analysis of an opaque *wor3* deletion strain versus wildtype opaque cells; values are on a log2 scale, data from Lohse *et al*. (Lohse *et al.*, 2013). (AB) Microarray analysis of an opaque *ahr1* deletion strain versus wildtype opaque cells; values are on a log2 scale.

File S2. MochiView image plots of 15kb regions centered on the sets of Ahr1, Czf1, and Efg1 binding sites in white cells. Plots produced using the SnapShot Function in MochiView v1.46 (Homann and Johnson, 2010).

File S3. Position Specific Weight Matrices (PSWM) for Ahr1, Czf1, Efg1, and Wor1 as well as Position Specific Affinity Matrices (PSAM) for selected Wor1, Wor2, Czf1, and Efg1 Motifs. This file contains PSWMs for the Ahr1, Czf1, Efg1, and Wor1 motifs developed using ChIP-chip binding data for these regulators. It also contains PSAMs developed from the Random 8mer Library experiments for the Efg1 motifs presented in Fig. 4B and File S4, the Czf1 motifs presented in Fig. 4C and File S4, the Wor1 motifs presented in Fig. 4E and File S4, and the Wor2 motifs presented in Fig. 4F and File S4. This file also includes the summaries of the Matrix Reduce results for all motifs developed for Wor1, Wor2, Czf1, and Efg1 in the Random 8mer Library experiments. The previously published Wor3 motif (Lohse *et al.*, 2013) is also included as are concentration-dependent binding curve experiment derived PSAMs shown in File S4.

File S4. Raw and additional data for Efg1, Czf1, Wor1, and Wor2 MITOMI experiments assessing binding of these regulators to a pseudorandom 8mer oligonucleotide library as well as the series of concentration-dependent binding curves for a set of all possible single systematic mutations within the Efg1, Czf1, Wor1, and Wor2 *cis*-regulatory sequences.

File S5. MochiView image plots of 15kb regions centered on the sets of Ahr1, Czf1, Efg1, and Wor2 binding sites in opaque cells. Plots produced using the SnapShot Function in MochiView v1.46 (Homann and Johnson, 2010).

File S6. MochiView image plots of 15kb regions centered on the sets of Wor1, and Wor3 binding sites in opaque cells. Plots produced using the SnapShot Function in MochiView v1.46 (Homann and Johnson, 2010). Data for Wor1 and Wor3 have been previously published (Zordan *et al.*, 2007; Lohse *et al.*, 2013).