Transcriptional Profiling of Biofilm Regulators Identified by an Overexpression Screen in *Saccharomyces cerevisiae*

Gareth A. Cromie*, Zhihao Tan†, Michelle Hays‡, Amy Sirr*, Eric W. Jeffery*, Aimée M. Dudley‡

* Pacific Northwest Research Institute, Seattle, Washington, USA

† Institute of Medical Biology, Agency for Science, Technology and Research, Singapore

‡ Molecular and Cellular Biology Program, University of Washington, Seattle, Washington, USA
Running Title: RNA-seq in smooth overexpression strains

Keywords: Biofilm, transcriptional regulation, overexpression, colony morphology

Corresponding author contact details:
Aimée M. Dudley
Pacific Northwest Research Institute
720 Broadway, Seattle, WA 98122
(206) 726-1237
aimee.dudley@gmail.com
Biofilm formation by microorganisms is a major cause of recurring infections and removal of biofilms has proven to be extremely difficult given their inherent drug resistance. Understanding the biological processes that underlie biofilm formation is thus extremely important and could lead to the development of more effective drug therapies, resulting in better infection outcomes. Using the yeast *S. cerevisiae* as a biofilm model, overexpression screens identified *DIG1*, *SFL1*, *HEK2*, *TOS8*, *SAN1* and *ROF1/YHR177W* as regulators of biofilm formation. Subsequent RNA-seq analysis of biofilm and non-biofilm forming strains revealed that all of the overexpression strains, other than *DIG1* and *TOS8*, were adopting a single differential expression profile, although induced to varying degrees. *TOS8* adopted a separate profile, while the expression profile of *DIG1* reflected the common pattern seen in most of the strains, plus substantial *DIG1*-specific expression changes. We interpret the existence of the common transcriptional pattern seen across multiple, unrelated overexpression strains as reflecting a transcriptional state, that the yeast cell can access through regulatory signaling mechanisms, allowing an adaptive morphological change between biofilm-forming and non-biofilm states.
Introduction

Many opportunistic human pathogens form highly structured, multicellular communities called biofilms, which are a key factor in persistent infections (COSTERTON et al. 1999; DONLAN AND COSTERTON 2002; FANNING AND MITCHELL 2012). Biofilms are a major cause of medical-device associated infections (COSTERTON et al. 1999; DONLAN AND COSTERTON 2002; DAROUICHE 2004), chronic non-healing of wounds (MANCL et al. 2013; SCALI AND KUNIMOTO 2013) and infections of the oral (MANCL et al. 2013), respiratory (KOBAYASHI 2005) and urinary tract surfaces (TENKE et al. 2012). The ability of biofilms to adhere to organic and inorganic surfaces as well as their increased drug resistance makes them a pressing clinical problem (COSTERTON et al. 1999; DONLAN AND COSTERTON 2002; RAMAGE et al. 2010). The transition from a planktonic, unicellular lifestyle to a sessile, multicellular lifestyle requires the coordinated activation and repression of numerous biological processes. While some of these pathways have been elucidated, many remain poorly characterized (FINKEL AND MITCHELL 2011; FANNING AND MITCHELL 2012). A better understanding of the molecular mechanisms required for biofilm formation is needed to guide the development of drug therapies that specifically target biofilms, an area that is largely underdeveloped (RAMAGE et al. 2010; PIERCE AND LOPEZ-RIBOT 2013).

The yeast Saccharomyces cerevisiae is an established model for the study of biofilm formation. While most S. cerevisiae strains form smooth, unstructured colonies on solid media, some strains are able to form highly structured “fluffy” colonies. Fluffy S. cerevisiae colonies not only visually resemble bacterial and fungal biofilms, but also share their structural and functional characteristics, including the presence of an extracellular matrix (KUTHAN et al. 2003; KARUNANITHI et al. 2010; VACHOVA et al. 2011), localized expression
of drug efflux pumps (VACHOVA et al. 2011), the use of intercellular communication (VOPALENSKA et al. 2010), reduced penetration of chemicals into the interior of the colony (VACHOVA et al. 2011), and increased adherence to inorganic surfaces (REYNOLDS AND FINK 2001).

A number of detailed genetic screens have been performed to elucidate the genes involved in colony morphology (GRANEK AND MAGWENE 2010; RYAN et al. 2012; VOORDECKERS et al. 2012; TAYLOR AND EHRENREICH 2015). Since the commonly used FY lab strain background (WINSTON et al. 1995) is unable to adopt the fluffy colony morphology, these studies have been performed with other strains that are able to form structured colonies. Results from these studies demonstrated that biofilm formation is finely tuned to the environment, with the ‘filamentation’ mitogen-activated protein kinase (MAPK) (GRANEK AND MAGWENE 2010; VOORDECKERS et al. 2012), Ras-cAMP-PKA (GRANEK AND MAGWENE 2010), target-of-rapamycin (TOR) (VOORDECKERS et al. 2012), and high osmolarity glycerol (HOG) (VOORDECKERS et al. 2012) signaling pathways all playing important roles in modulating colony morphology. Notably, the filamentation MAPK pathway regulates genes, such as the adhesin FLO11, that appear to be required for complex colony formation, while pseudohyphal growth itself is not essential for this phenotype (ST’OVICEK et al. 2010).

Another approach to understanding biofilm formation has been the comparison of wild yeast strains in their biofilm and non-biofilm forms. Because naturally-occurring fluffy wild strains can switch to a non-biofilm “smooth” colony-forming state at relatively high rates during the process of domestication in the laboratory (KUTHAN et al. 2003), it is possible to isolate smooth and fluffy forms of the same strain. These studies have uncovered
metabolic differences between fluffy strains and their smooth counterparts, including the
differential expression of genes involved in metabolism and the transport of
carbohydrates, vitamins and amino acids (Kuthan et al. 2003; Stovicek et al. 2014). A
fluffy strain (F45) studied in our laboratory also switched between fluffy and smooth
colony morphologies at rates much higher than that of spontaneous mutations (Tan et al.
2013). In this strain, the mechanism underlying the switch was the gain and loss of
individual chromosomes. Because increased copy number of only a subset of the sixteen
S. cerevisiae chromosomes could induce the phenotypic change, we hypothesized that
the switch was caused by dosage changes of specific genes on those chromosomes.
Consistent with this model, a genetic screen for individual genes on one of these
chromosomes (XVI) whose modest overexpression was sufficient to cause the fluffy to
smooth transition, identified DIG1 (Tan et al. 2013), which encodes a transcriptional
repressor in the filamenation MAPK pathway (Cook et al. 1996).

Here, we extend our screen genome-wide and identify five additional genes whose
overexpression causes fluffy F45 colonies to become smooth. We then use RNA-seq to
characterize the transcriptional profiles of the smooth overexpression strains relative to
the original F45 strain and look for functional enrichment among the most strongly
induced and repressed genes. Several of our results support a model in which regulated
changes between a small number of transcriptional states underlie changes in colony
morphology. First, all six of the genes identified in our overexpression screens encode
regulatory proteins. Second, five of these genes, when overexpressed, induce a single
transcriptional response, along with a change in colony morphology, suggesting they
influence a common signaling network. Third, the transcriptional responses that we
observe include genes, such as *FLO11*, that are known to effect changes in colony morphology.

**Materials and Methods**

**Yeast Strains and Media**

Unless noted, standard media and methods were used for growth and genetic manipulation of yeast (*ROSE* 1990). The strains of *S. cerevisiae* used in this study are listed in Table 1.

**Table 1. Strains used in this study.**

<table>
<thead>
<tr>
<th>Strain Name</th>
<th>Progenitor</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>YO1853</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [pRS41K]</td>
<td>Strain (<em>TAN et al. 2013</em>), plasmid (<em>TAXIS AND KNOP 2006</em>)</td>
</tr>
<tr>
<td>YO1773</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [DIG1-pFA6a-KanMX4]</td>
<td>(<em>TAN et al. 2013</em>)</td>
</tr>
<tr>
<td>YO1829</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [SAN1-p5472]</td>
<td>(<em>HO et al. 2009</em>)</td>
</tr>
<tr>
<td>YO1832</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [TOS8-p5472]</td>
<td>(<em>HO et al. 2009</em>)</td>
</tr>
<tr>
<td>YO1835</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [ROF1-p5472]</td>
<td>(<em>HO et al. 2009</em>)</td>
</tr>
<tr>
<td>YO1845</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [SFL1-p5472]</td>
<td>(<em>HO et al. 2009</em>)</td>
</tr>
<tr>
<td>YO1849</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [HEK2-p5472]</td>
<td>(<em>HO et al. 2009</em>)</td>
</tr>
<tr>
<td>YO2111</td>
<td>F45</td>
<td>MATa·hoΔ0::hphMX6, SPS2·EGFP·natMX4, unmapped serine auxotroph, [SAN1(R280A)-pRS41k]</td>
<td>(<em>HO et al. 2009</em>)</td>
</tr>
<tr>
<td>YO780</td>
<td>F13</td>
<td>MATα·hoΔ0::hphMX6, SPS2·EGFP·natMX4</td>
<td>This study</td>
</tr>
</tbody>
</table>
Overexpression Screen

In order to perform the overexpression screen using the MoBY plasmid library (Ho et al. 2009), we first created pools of plasmids by pinning the master library from 96 well plates to selective (G418) LB Lennox Omnitrays and scraping the colonies from 10 such pinned plates into each pool pellet. Each pool pellet of approximately 3g of cells was split over four Qiafilter Plasmid Maxi Prep columns (Qiagen) per the manufacturer’s protocol.

With an assumed 1000 plasmids per pool, screening 4600 transformants was expected to give us a 99% chance of recovering each plasmid at least once. Based on this, we screened five bioassay trays (YPD with 2 % glucose and G418 for plasmid selection) per pool, plating 1000-1500 CFU per tray. Transformations were performed using standard LiAc/DMSO methods. Colony morphology was examined after 4 days of colony growth. Transformants with completely smooth or intermediate, i.e. more smooth than the F45 parent, phenotypes were picked. The frequency of smooth colonies following pool
transformation was low (<5%) in all the pools. Transformation using a negative vector control, AB352 (pFA6a, Addgene), did not produce any smooth colonies.

Plasmid DNA was prepared from the smooth strains by standard methods (HOFFMAN AND WINSTON 1987) and the identity of the overexpressed gene in each strain was then identified by PCR amplification and conventional ABI sequencing of the UPTAG and DNTAG barcodes (Ho et al. 2009). Strains with failed sequencing or inconsistent barcodes were removed from further consideration, leaving a set of 292 strains harboring 259 distinct plasmids.

Among the final set of plasmids, only 18 were identified more than once, leaving 241 singletons. The large proportion of single hits could have been due to a high false positive rate and/or failure to saturate the screen. To distinguish between these possibilities, we constructed a new smaller pool (a "singleton" pool) that contained 91 plasmids that came up in the list of single hits. The relative number of smooth vs fluffy colonies resulting from transformation with this singleton pool would help distinguish if the false positive rate was high or if the screen was undersaturated. The transformation plates were almost entirely fluffy with <1% smooths in ~3000 CFU. This suggested that the singletons observed in the screen were largely false positives, rather than reflecting screen undersaturation.

The 18 plasmids observed more than once in the initial screen were retested by individual transformations of each plasmid into F45, followed by screening for colony morphology. Plasmids that induced a phenotype change were further verified to ensure they contained the correct ORF, based on restriction digest pattern and size confirmation by PCR. Only five genes were verified as positive: SAN1, TOS8, YHR177W, HEK2 and SFL1. This
included all three of the genes isolated more than two times, but only two of the fifteen genes observed twice.

RNA Preparation and Sequencing

After four days of growth on YPD (with 2 % glucose plus G418 to maintain plasmid selection) plates at 30° C, whole colonies, arrayed in a “checkerboard” pattern (RUUSUVUORI et al. 2014), were harvested by scraping off the surface of the agar plate. To obtain sufficient amounts of RNA, 3-5 colonies were pooled for each sample, with four replicate samples of each strain.

Following extraction by hot acid phenol (COLLART AND OLIVERO 2001), total RNA from the pooled colonies was quantified by Bioanalyzer (Agilent). 5 µg of total RNA for each sample was then processed using the Tru-Seq stranded mRNA kit (Illumina) following manufacturer instructions. Individual sequencing libraries were pooled and analyzed by paired-end, 51 nucleotide read sequencing in one lane of an Illumina HiSeq 2000.

Read-Pair Alignment

Read-pair alignment for RNA-seq data was carried out against the S288c reference (R64-1-1), with the FASTA and GFF files extended to include non-coding RNAs (ncRNAs) and genes present in F45, but absent in S288c, as described (CROMIE et al. 2017), using Bowtie2 (version 2.1.0) (LANGMEAD AND SALZBERG 2012) with the parameters [-N 1 -I 50 -X 450 -p 6 --reorder -x -S] and allowing 1 mismatch per read. For each strain, read alignments were converted to gene counts using featureCounts (version 1.4.0) in the
Subread package (Liao et al. 2014), with the parameters [-a -o -t gene –g ID –s 2 -T 1 -p -P -d 50 -D 450]. Reads were not filtered based on mapping quality, and thus we have been cautious in our interpretation of counts of genes that have paralogs with similar sequences, or which contain large regions of low sequence complexity. Read sequences and gene count tables are available from the Gene Expression Omnibus under accession GSE98079. Data for the control strain (empty vector) and the DIG1 overexpression strain have been previously published (Cromie et al. 2017) (GSE85843).

**Differential Expression Analysis**

Analysis of differential gene expression was carried out using edgeR [v. 3.6.8] (Robinson et al. 2010) based on the tables of raw counts produced by featureCounts (Materials and Methods). The table of counts was split into two sub-tables, with the first consisting of ORFs present in the S288c reference genome (genes with systematic names beginning with “Y”) and the second consisting of the novel F45 genes and ncRNAs. Library sizes were normalized using calcNormFactors, and dispersion parameters were estimated using the estimateGLMTrendedDisp and estimateGLMTagwiseDisp commands. To identify genes differentially expressed between overexpression strains and the F45 empty-vector control, we conducted pairwise testing using the glmFit and glmLRT commands, with a p-value cutoff of 0.01 (after Benjamini & Hochberg multiple hypothesis correction, i.e. false discovery rate). Only nuclear-encoded ORFs, present in the reference genome, with median basal expression of at least 1 read per million reads in the F45 empty-vector control were included when plotting log2 fold changes on
scatterplots, carrying out factor analysis, and calculating regression parameters or correlation coefficients.

Functional Enrichment of Gene Lists
Functional enrichment of *S. cerevisiae* gene lists was performed using g:Profiler (http://biit.cs.ut.ee/gprofiler/). Holm-Bonferroni corrected Enrichment p-values < 0.05 were accepted, with moderate hierarchical filtering.

Factor Analysis
Factor analysis was carried out using the fa command in the “psych” package of R (version 3.1.1, (TEAM 2014)), with a single factor specified and default parameters. Choice of a single factor was determined from scree and parallel analysis using the nScree command (package “nFactors”). Loadings for the single common factor were calculated using only nuclear encoded ORFs with median basal expression of at least 1 read per million reads in F45. The loadings were: DIG1=0.731, SAN1=0.883, TOS8=0.206, ROF1=0.812, SFL1=0.894 and HEK2=0.918. Significant non-zero values on the common factor were identified using the general linear model functionality of edgeR (ROBINSON *et al.* 2010) and applying these loadings (normalized to sum to 1) as the “contrast” terms in the glmLRT command with a p-value cutoff of 0.01 (after Benjamini and Hochberg multiple hypothesis correction). Significant non-zero values for the residual *DIG1* profile were identified as the residual from a linear regression between the *DIG1* profile and the common factor. The “contrast” terms for the common factor, multiplied by the regression coefficient, were subtracted from the contrast terms used to calculate the full *DIG1* profile.
to give the contrast terms producing the residual. The variance explained by the common factor for each overexpression profile was calculated as the variance of the overexpression profile times the communality value from the factor analysis. The communality values were: $DIG1=0.534$, $SAN1=0.780$, $TOS8=0.042$, $ROF1=0.659$, $SFL1=0.799$ and $HEK2=0.842$.

Availability of Data and Material

The datasets generated during and/or analyzed in the current study are available in the Gene Expression Omnibus (GEO) under GSE98079.
Results

A Genetic Screen for Modulators of Biofilm Formation: The fact that increased copy number of several chromosomes could modulate fluffy colony morphology (Tan et al. 2013) suggested that increased copy number of other genes might regulate the trait in the same way as DIG1. To identify such modulators, we performed a large-scale overexpression screen for genes that reduced or eliminated the complex structure of the fluffy colonies. Briefly, we transformed the fluffy, euploid version of F45 from our previous study (Tan et al. 2013) with the MoBY plasmid collection, a low copy-number (CEN) plasmid library containing 4981 individual ORFs under the transcriptional control of their native promoters (Ho et al. 2009). A set of 292 strains harboring 259 distinct plasmids were initially identified as having reduced colony morphology. Only 18 of these plasmids were identified more than once, suggesting either a high frequency of false positives or incomplete library coverage. Since retesting a subset of the singletons indicated that most were false positives (Materials and Methods) we proceeded with only the multiply-hit plasmids. The high frequency of false positives may result from the gain of additional copies of chromosomes, a phenomenon known to occur at high frequency in the F45 background and which often results in loss of colony morphology (Tan et al. 2013). Alternatively, other mechanisms such as prion-switching (Holmes et al. 2013) or high-frequency mutations (Halme et al. 2004) may be responsible for loss of colony morphology in the false positives.

After individual re-transformation into F45 and assays of colony morphology, five genes were verified: SAN1, TOS8, YHR177W, HEK2 and SFL1, which we compared to DIG1, identified in our previous screen (Tan et al. 2013). The strength of the reduction in colony
structure varied from very strong (*DIG1*, *SFL1*, *HEK2*), through intermediate (*SAN1*, *ROF1*), to weak (*TOS8*). These results (Figure 1A) suggested that modest increases in the copy number of these genes, is sufficient to reduce complex colony morphology to varying degrees (Materials and Methods). Because of the weak phenotypes associated with overexpressing some of these genes, we further tested their role in colony morphology by deleting them in a different strain background, F13 (Figure 1B). F13 is a strain from the same cross that produced F45 (*TAN et al. 2013*) and forms colonies that are smooth in the center and structured on the periphery. Our results from F45 suggested that the genes we identified in the overexpression screen(s) suppress colony structure, and consistent with this, deletion of these genes in F13 increased the degree of colony structure in that genetic background (Figure 1B).

Three of the five genes identified in our screen have known links to complex colony phenotypes. *SFL1* encodes a transcriptional repressor of the cell-surface flocculation genes, which are important for complex colony formation (*LO AND DRANGINIS 1998; GUO et al. 2000; CONLAN AND TZAMARIAS 2001; HALME et al. 2004; GRANEK AND MAGWENE 2010*) (*VOORDECKERS et al. 2012*). *YHR177W* is a paralog of *MIT1*, which encodes a transcriptional regulator of pseudohyphal growth (*CAIN et al. 2012*), and is an ortholog of *WOR1*, a master regulator of the white-opaque phenotypic switch in *Candida albicans* (*ZORDAN et al. 2006*). Interestingly, while *YHR177W* has been shown to encode a protein having DNA-binding properties (*CAIN et al. 2012*), its effect on colony morphology in the Σ1278b strain background was unclear, with deletion reducing complex colony morphology in one study (*FURUKAWA et al. 2011*) but having little effect in another (*CAIN et al. 2012*). Owing to its identification in our screen, we have given *YHR177W* the name
ROF1 (Regulator Of Fluffy1). HEK2 encodes an RNA-binding protein which regulates the floculin FLO11, first via regulation of the ASH1 mRNA transcript, which specifies differential gene expression in mother vs daughter cells, and also through a second, post-transcriptional mechanism (IRIE et al. 2002; WOLF et al. 2010).

To the best of our knowledge, the remaining two genes from our screen have not been previously shown to influence colony morphology. SAN1 encodes a ubiquitin-ligase involved in targeting aberrant nuclear proteins to the proteasome for degradation (DASGUPTA et al. 2004; GARDNER et al. 2005). TOS8 encodes a homeodomain-containing transcription factor whose targets include a statistically significant enrichment for genes involved in bud growth (HORAK et al. 2002).

For one of these genes, SAN1, reagents were available to test whether the catalytic activity of the encoded protein played a role in its effect on colony morphology. To test whether the phenotypic effect of SAN1 overexpression was dependent on the function of San1 as a ubiquitin ligase, we overexpressed an allele (san1-R280A, kind gift of Dr. Richard Gardner) with a single amino acid change in the RING domain that inactivates the protein’s ubiquitin ligase activity without altering protein structure (FREDRICKSON et al. 2013). In contrast to overexpression of fully functional SAN1, overexpression of the catalytically dead san1-R280A allele had no phenotypic effect, with F45 colonies remaining as fluffy as the vector alone control (Figure 1C). Therefore, the effect of SAN1 copy number appears to be mediated through its ubiquitin ligase activity.

A Common Transcriptional Profile Characterizes Most of the Overexpression Strains: Because four of the six genes (including DIG1) identified in our overexpression
screen are known or putative transcription factors, differences in RNA expression levels
could help explain their influence on colony morphology. To investigate this, we began by
comparing the mRNA expression patterns of each of the smooth overexpression strains
to the fluffy progenitor (F45). Briefly, we isolated total RNA from single colonies growing
on solid medium and prepared libraries to measure stranded mRNA using the Illumina
TruSeq method (Materials and Methods). Each smooth overexpression strain showed
a substantial number of genes that were significantly differentially expressed relative to
the fluffy F45 vector-alone control (multiple hypothesis corrected p<0.01: \( DIG1=1976; \)
\( HEK2=2668; SAN1=1630; SFL1=2385; TOS8=722; ROF1=1567 \)) (Table S1). Notably,
significant differential expression was seen in the \( SAN1 \) and \( HEK2 \) overexpression
strains, despite the fact that these genes do not encode transcription factors.

Comparing the differential expression profiles of these strains identified substantial
correlations between the \( ROF1, HEK2, SFL1, SAN1 \) and, to a lesser extent, \( DIG1 \)
overexpression patterns, but little correlation between the strain overexpressing \( TOS8 \)
and any of the others (Table 2). In particular, the differential expression profiles of the
\( HEK2 \) and \( SFL1 \) strains (R=0.86) and the \( SAN1 \) and \( ROF1 \) strains (R=0.84) were very
similar, despite the fact that genes in each pair encode one transcription factor and one
protein with a different function. One possible explanation for these correlated pairs is
that overexpression of the non-transcription factor leads to increased expression of the
transcription factor, resulting in the same transcriptional profile as when the transcription
factor is directly overexpressed from the plasmid. However, we found that for both pairs
of genes with highly similar expression profiles (\( HEK2-SFL1 \) and \( SAN1-ROF1 \)), the
expression level of one gene in the pair was essentially unperturbed by the overexpression of the other (Table 3).

To further explore the observed correlations between the RNA expression profiles of many of our overexpression strains, we carried out common factor analysis (Materials and Methods). Factor analysis seeks to explain correlations between observed variables (e.g. the individual global expression patterns of our overexpression strains) in terms of a potentially smaller number of unobserved variables called factors or latent variables. Each observed variable is explained as a linear combination of these latent factors plus residual effects unique to the observed variable (including noise).

Applying common factor analysis to our transcriptional profiles identified a single strongly significant common factor in the data (Figure S1). This factor explained most of the variance in the ROF1, HEK2, SFL1 and SAN1 strains and left a similar level of residual variance in each strain, suggesting this residual variance might largely reflect a common level of noise rather than substantial strain-specific effects on expression (Figure 2). The common factor also explained approximately half of the variance in the DIG1 strain (variance explained=53%), leaving a substantially higher degree of residual variance for DIG1 than for the ROF1, HEK2, SFL1 and SAN1 overexpression strains (Figure 2). As expected from the lack of significant correlation between the TOS8 and other profiles, the latent factor explained almost none of the TOS8 differential expression pattern (variance explained=4%). Taken together, these data suggest that overexpression of ROF1, HEK2, SFL1 and SAN1 induces a single expression profile (the common factor), although to differing extents, as the slopes of linear regressions between the common factor and the
transcriptional profiles of these strains varied from 0.24 for ROF1 to 0.39 for HEK2 (Figure 3). In contrast, the effect of DIG1 overexpression appears to reflect a combination of the common factor and substantial differential expression specific to the DIG1 strain (Figure 4A), while overexpression of TOS8 induces a completely distinct profile (Figure 4B). Thus, it appears that a very limited number of transcriptional states are observed in the six overexpression strains: a profile unique to the TOS8 strain, a single transcriptional state induced to varying extents by the ROF1, HEK2, SFL1 and SAN1 strains, and a profile unique to the DIG1 strain, which appears to reflect the common factor along with additional DIG1-specific transcriptional effects.

Potential Role of Smooth Versus Fluffy Colony Environment on Gene Expression:
Because overexpression of all of the genes identified in our genetic screen elicited a common phenotypic response, colonies that are less fluffy than their F45 progenitor, one possible explanation for their similar expression profiles could be a common response to the smooth versus fluffy colony state or “environment”. For example, complex colonies have a larger surface area to volume ratio than smooth colonies and contain internal cell-free spaces (STOVICEK et al. 2012), differences that could alter nutrient and oxygen availability for cells in the two colony-types. To test this possibility, we compared the RNA expression profiles of the overexpression strains to our previously published gene expression data for F45 colonies with deletions of two genes, CIS3 and FLO11, which cause F45 colonies to become smooth (although maintaining irregular edges) (CROMIE et al. 2017). Our previous results suggested that the CIS3 deletion (cis3Δ) provided the cleanest readout of the smooth colony environment, and that this environment does not
substantially alter gene expression. Consistent with this result, correlations between the
*cis3Δ* and our overexpression profiles were very poor (*Table 4*), suggesting that the
smooth colony environment does not appear to drive the differential gene expression
patterns seen in our overexpression strains. We did observe some modest correlations
between *flo11Δ* and our overexpression profiles (*Table 4*). However, *flo11Δ* appears to
have effects on the expression of genes involved in colony morphology signaling
pathways, rather than just reflecting gene expression associated with the smooth colony
state (CROMIE *et al.* 2017). These effects are likely to explain the modest correlations
between *flo11Δ* and our overexpression profiles.

11

**Shared Features of the Common Factor and TOS8 Transcription Profiles:** Analysis
of our overexpression strains identified three independent transcriptional profiles: the
common factor, the *TOS8* profile, and the *DIG1* residual profile. Both the common factor
and the *TOS8* overexpression profile are associated with a reduction in structured colony
morphology. *DIG1* overexpression is also associated with this morphological change but,
because the *DIG1* overexpression profile consists of the common factor combined with a
*DIG1*-specific residual expression profile, the relationship between the *DIG1* residual
profile and colony morphology was not as clear. Therefore, to identify transcriptional
changes commonly associated with the fluffy to smooth transition, we compared the
*TOS8* overexpression profile and the common factor.

First, we functionally characterized the two expression profiles, individually, by looking for
gene ontology (GO) term, transcription factor target and metabolic/regulatory pathway
enrichment among the sets of genes significantly (p<0.01, after multiple hypothesis correction) and strongly (>1.5-fold) induced and repressed in each profile (Supplemental Tables 2 and 3). This analysis identified several overlapping features of GO term and pathway enrichment. In both profiles, cell periphery genes were over-represented among induced genes, and genes encoding anchored components of the membrane and reproductive genes were over-represented among repressed genes. In both the TOS8 and common factor profiles, MAPK signalling genes were significantly over-represented among genes repressed in the smooth colonies, as were genes involved in mating. One of the pathways known to induce complex colony morphology is the filamentation MAPK cascade, and the components of this pathway overlap to a large extent with the mating/pheromone-response pathway (Roberts and Fink 1994; Chou et al. 2006; Cullen and Sprague 2012).

We hypothesized that the genes consistently upregulated or consistently downregulated in both the common factor and the TOS8 overexpression profile might include genes involved in the common change in colony morphology associated with both profiles. Consistent with this hypothesis, FLO11 expression is repressed both in the common factor and in the TOS8 profile (Figure 4B). Similarly, although there was a very poor correlation between the common factor and the TOS8 profile across all genes (R²=0.04) (Figure 4B), genes that were significantly (p<0.01, after multiple hypothesis correction) differentially expressed in both profiles tended to show a consistent direction of effect (Fisher’s exact test, 2-tailed:p<2.2e-16). Among 400 such genes, 153 were induced and 173 were repressed in both profiles. In contrast, only 36 genes were induced in the
common factor and repressed by TOS8 overexpression, and only 38 were repressed in the common factor and induced by TOS8 overexpression.

The genes commonly induced in both the TOS8 overexpression and common factor profiles were enriched for several GO terms including “transmembrane transport” (GO:0055085; p=8.48e-4), “cell periphery” (GO:0071944; p=1.06e-3) and “extracellular region” (GO:0005576; p=1.26e-3) (Table S4). Similarly, enrichment for several GO terms was seen in the set of genes commonly repressed in both the TOS8 overexpression and common factor profiles, including “conjugation with cellular fusion” (GO:0000747; p=2.10e-6), “cell periphery” (GO:0071944; p=1.25e-7) and “anchored component of membrane” (GO:0031225; p=4.87e-4) (Table S4). Repressed genes were also enriched for targets of Tec1 (p=3.35e-6), a transcriptional inducer of complex colony morphology that acts in the filamentation MAPK pathway (CHOU et al. 2006) (Table S4). TEC1 itself was also one of the genes commonly repressed, as were the flocculin genes FLO10 and FLO11 that are known to be effectors of complex colony phenotypes (LO AND DRANGINIS 1998; GUO et al. 2000; HALME et al. 2004; GRANEK AND MAGWENE 2010; VOORDECKERS et al. 2012). These results support the hypothesis that a set of genes important for the fluffy-smooth morphological switch is consistently repressed or induced in both the common factor and TOS8 overexpression profiles.

Central Role of the Filamentation MAPK Pathway in the Change in Colony Morphology: The DIG1 strain has the strongest smooth colony phenotype of any of our overexpression strains (Figure 1A). The transcriptional profile of this strain consists of the common factor plus a residual DIG1-specific expression pattern suggesting that the
residual profile also contributes to the smooth phenotype. Dig1 is a repressor of the Ste12-Tec1 transcriptional complex (Cook et al. 1996; Chou et al. 2006), and genes in “MAPK signalling pathway - yeast” (KEGG:04011; p=3.81e-3) and targets of Tec1 (p=1.85e-5) were over-represented among the genes repressed in the DIG1-residual profile (Table S5). Therefore, the residual DIG1 profile appears to reflect repression of the filamentation MAPK pathway.

As discussed above, many of the genes repressed in both the common factor and the TOS8 profile are genes in the filamentation MAPK pathway. Consistent with this, many of these genes (80/173) were also downregulated in the DIG1 residual profile, a highly significant enrichment (Fisher’s exact test, 1-tailed: p<2.2e-16). This further supports the hypothesis that a substantial component of the overlap between the TOS8 and common factor profiles reflects repression of the filamentation MAPK pathway. In the DIG1 overexpression strain these genes are repressed as part of the common factor and then further repressed by specific effects of DIG1 overexpression (i.e. the DIG1 residual profile). This double signalling through the filamentation MAPK pathway may explain why the DIG1 overexpression strain has such a strong smooth phenotype (Figure 1A).

Effects of Overexpression Strains on Non-Coding RNAs and Non-Reference Gene Transcripts: Finally, we extended our analysis to the expression levels of a large number of non-coding RNAs (ncRNAs) and a small number of genes present in strain F45, but absent from the S288c reference genome (Cromie et al. 2017) (Materials and Methods). Similar to reference mRNAs, substantial differential expression was observed among these transcripts in all overexpression strains (Table S6). The same pattern of
correlations between overexpression strains, that was seen with reference genes (Table 2) was also observed with ncRNAs and non-reference mRNAs (Table 5) (Figure S2).

This implies that the three underlying transcriptional profiles that we identified from analysis of reference genes extend to include effects on the expression of substantial numbers of non-reference genes and non-coding transcripts.
Discussion

Morphology Change as a Regulated Switch Between Transcriptional States:

Several mechanisms have been shown to govern the stable switching between structured/biofilm and non-structured colony morphologies in S. cerevisiae. These include high frequency mutations (HALME et al. 2004) and prion-based mechanisms (HOLMES et al. 2013) as well as aneuploidy, specifically the gain and loss of additional copies of whole chromosomes (TAN et al. 2013). Notably, several of these mechanisms are transcriptional in nature. In the F45 strain used in this study, overexpression of the Dig1 transcriptional repressor contributes to the fluffy-smooth switch induced by a chromosome XVI disome (TAN et al. 2013). In another case, high frequency mutations that inactivate the Ira1 and Ira2 Ras-activating proteins (HALME et al. 2004), cause changes in colony morphology through the Ras-cAMP-PKA pathway which regulates the Sfl1 and Flo8 transcription factors (RUPP et al. 1999). Similarly, switching between the prion and non-prion forms of the Mot3 transcription factor produces a colony morphology switch (HOLMES et al. 2013).

One of the best characterized phenotypic switches regulated by a transcriptional circuit in fungi is the white to opaque switch in C. albicans (reviewed in (HUANG 2012; SOLL 2014)). The opaque state is necessary for mating, while white cells can form structured biofilms that are highly impermeable to drugs and components of the immune system. Each of the two states appears to be adaptive for survival in different host niches, with white cells being favored in systemic infection models and opaque cells showing improved ability to colonize the skin. A transcriptional circuit with two stable states
underlies the white and opaque phenotypes, with white being the default state. Because switching between states leads to the differential expression of a large number of genes, each state is associated with a specific expression profile. Interestingly, WOR1 (a homolog of S. cerevisiae ROF1) is the master regulator of the white-opaque switch and ectopic expression of WOR1 can drive an entire population of white cells to become opaque (HUANG et al. 2006).

Similar to the effect of overexpressing WOR1 in C. albicans white-opaque switching, we propose that overexpression of most of the genes identified in our study produces a common change in phenotype and gene expression that replicates an adaptive, regulated switching mechanism. That is, some aspect(s) of the regulated switching between the fluffy and smooth transcriptional states produces an adaptive change in colony morphology and the expression profiles that we observe reflect transcriptional states that S. cerevisiae may access in response environmental conditions favoring the biofilm or non-biofilm states of the microbial community. We previously demonstrated that the smooth and fluffy (biofilm) states of the F45 background used in this study are each adaptive in different environmental conditions (TAN et al. 2013).

The regulated transcriptional switching model makes several key predictions. First, regulatory genes should exist that control the process and perturbation of these genes could trigger the phenotypic switch between the fluffy and smooth colony morphologies, mimicking signaling through those genes. Second, perturbation of regulatory genes acting in the same pathway should produce the same transcriptional state. Third, the transcriptional response should alter the activity of downstream pathways that effect the morphological change. Our data are consistent with each of these expectations.
**Regulatory Repressors of Biofilm Formation:** The six genes that caused fluffy to smooth transitions in our overexpression screens all encode known or predicted regulatory proteins, with four being transcription factors. In contrast, we did not identify any genes that appear to have mechanistic roles in colony morphology, such as cell surface proteins, despite the fact that deleting such genes individually (e.g. cis3Δ and flo11Δ) is sufficient to cause a fluffy-to-smooth transition in the same genetic background (Cromie et al. 2017). Furthermore, the genes that we identified appear to be components of (or to interact with) regulatory pathways that repress the complex colony phenotype. In fact, DIG1 and SFL1 encode known transcriptional repressors of genes needed for complex colony formation, and act in the filamentation MAPK and Ras-cAMP-PKA signaling pathways, respectively (Conlan and Tzamarias 2001; Chou et al. 2006; Granek and Magwene 2010).

Our study also identified a phenotype for genes that were previously characterized only at the level of biochemical activity (known or predicted). This was true for the putative transcription factor Rof1, which we named based on its phenotype in this study, and the ubiquitin ligase San1, neither of which had been linked to colony morphology prior to our study. Previous characterization of San1 had identified a number of target proteins (Gardner et al. 2005; Rosenbaum et al. 2011), and the phenotype identified here may help identify specific pathways that are regulated by this activity. This may also be true for HEK2, which encodes a protein that has been shown to bind to a large number of mRNAs (Hogan et al. 2008).
Convergence on a Common Transcriptional State: We expected that perturbation of genes operating in the same regulatory pathway would produce the same transcriptional response. Overexpression of four of the six genes that we identified (SFL1, SAN1, ROF1 and HEK2) produced a single transcriptional state, the common factor, consistent with them operating in a single transcriptional regulatory pathway. This common factor is also a large component of the transcriptional profile observed when a fifth gene, DIG1, is overexpressed.

The “single regulatory pathway” identified by these genes is probably better understood as a signaling network, as it encompasses both the cAMP-PKA pathway (including SFL1) and the filamentation MAPK pathway (including DIG1) (Vinod et al. 2008). Crosstalk between these pathways is known to exist (Mosch et al. 1999; Borneman et al. 2006) allowing signal integration to occur (Cullen and Sprague 2012). Such a signaling network could integrate a range of environmental signals to produce a transcriptional response, and ultimately a phenotypic change. This model can explain how increasing the expression of genes that do not directly regulate gene expression, such as HEK2 and SAN1, can produce a strong transcriptional response, and one shared with transcription factors such as SFL1 and ROF1. In this case, SFL1 and ROF1 represent (complex morphology repressing) transcriptional nodes in the signaling network, while HEK2 and SAN1 might regulate elements of the signaling pathways at the RNA- and protein-abundance levels. For these four genes, the signal produced by overexpression would, after signal integration across the network, have the same final effect in terms which transcriptional state the cell/colony adopts. That is, the transcriptional state corresponding to the common factor may be one that emerges from any one of a range of perturbations.
In contrast to the other genes in our study, overexpression of *TOS8* did not induce the common-factor transcriptional state. This result suggests that Tos8 operates in a different regulatory pathway than the products of the other genes. This, in turn, suggests that the *TOS8* overexpression profile may represent a distinct transcriptional state that cells can adopt under environmental conditions that differ from those producing the common profile.

**Genes Important for the Smooth-Fluffy Morphological Switch:** The common-factor and *TOS8* transcriptional profiles are essentially uncorrelated and only a small proportion of genes are significantly differentially expressed in both. However, among the overlap between these two profiles, there is a statistically significant overrepresentation of genes showing differential expression with the same direction of effect, i.e. induced or repressed in both profiles, compared to genes that have opposite directions of effect. This result suggests the existence of a differentially expressed subset of genes that are involved in the common fluffy-smooth morphological change associated with both profiles. Because the overexpression strains appear to be operating through gene-expression regulatory mechanisms, the commonly differentially expressed genes may represent modules of co-regulated genes, that respond as a group to environmental signals.

A high proportion of the genes commonly repressed in both the *TOS8* and common factor profiles, were also significantly repressed in the residual *DIG1*-expression profile, i.e. the gene expression changes seen in the *DIG1* overexpression strain, after accounting for the effect of the common factor. It therefore appears that *DIG1* and the filamentation
MAPK pathway are particularly important in control of the critical gene module(s), repressing genes whose expression is required for fluffy colony formation.
Acknowledgements

We thank Richard Gardner for helpful discussion and for providing the san1-R280A allele. This work was funded by a grant from the National Institutes of Health (P50 GM076547) and a strategic partnership between the Institute for Systems Biology and the University of Luxembourg. Z.T. was funded by the Agency for Science, Technology and Research, Singapore.
Table 2. Correlation (R) between differential gene expression profiles (relative to F45) of overexpression strains.

<table>
<thead>
<tr>
<th></th>
<th>DIG1</th>
<th>SFL1</th>
<th>HEK2</th>
<th>SAN1</th>
<th>ROF1</th>
<th>TOS8</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIG1</td>
<td>0.70</td>
<td>0.65</td>
<td>0.66</td>
<td>0.49</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>SFL1</td>
<td>0.70</td>
<td>0.86</td>
<td>0.74</td>
<td>0.67</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>HEK2</td>
<td>0.65</td>
<td>0.86</td>
<td>0.79</td>
<td>0.72</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>SAN1</td>
<td>0.66</td>
<td>0.74</td>
<td>0.79</td>
<td>0.84</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>ROF1</td>
<td>0.49</td>
<td>0.67</td>
<td>0.72</td>
<td>0.84</td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td>TOS8</td>
<td>0.42</td>
<td>0.15</td>
<td>0.09</td>
<td>0.24</td>
<td></td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 3. Expression change for each of the genes from our screen in each overexpression strain, relative to F45 (log₂ fold change).

<table>
<thead>
<tr>
<th>Gene</th>
<th>DIG1</th>
<th>SFL1</th>
<th>HEK2</th>
<th>SAN1</th>
<th>ROF1</th>
<th>TOS8</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIG1</td>
<td>2.11</td>
<td>-0.11</td>
<td>0.16</td>
<td>-0.03</td>
<td>-0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>SFL1</td>
<td>-0.14</td>
<td>2.23</td>
<td>0.10</td>
<td>-0.10</td>
<td>-0.09</td>
<td>0.24</td>
</tr>
<tr>
<td>HEK2</td>
<td>-0.07</td>
<td>0.06</td>
<td>1.63</td>
<td>-0.07</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>SAN1</td>
<td>-0.17</td>
<td>0.07</td>
<td>0.19</td>
<td>1.32</td>
<td>-0.05</td>
<td>-0.08</td>
</tr>
<tr>
<td>ROF1</td>
<td>-0.06</td>
<td>0.23</td>
<td>0.21</td>
<td>0.05</td>
<td>1.58</td>
<td>0.14</td>
</tr>
<tr>
<td>TOS8</td>
<td>-0.04</td>
<td>0.54</td>
<td>0.15</td>
<td>-0.14</td>
<td>-0.02</td>
<td>2.31</td>
</tr>
</tbody>
</table>

Table 4. Correlation (R) between differential gene expression profiles (relative to F45) of overexpression strains and those of CIS3 and FLO11 deletions.

<table>
<thead>
<tr>
<th></th>
<th>FLO11</th>
<th>CIS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIG1</td>
<td>0.36</td>
<td>-0.08</td>
</tr>
<tr>
<td>SFL1</td>
<td>0.42</td>
<td>-0.07</td>
</tr>
<tr>
<td>HEK2</td>
<td>0.44</td>
<td>-0.05</td>
</tr>
<tr>
<td>SAN1</td>
<td>0.29</td>
<td>-0.04</td>
</tr>
<tr>
<td>ROF1</td>
<td>0.20</td>
<td>-0.01</td>
</tr>
<tr>
<td>TOS8</td>
<td>0.01</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Table 5. Correlation (R) between differential expression profiles (relative to F45) of overexpression strains for ncRNAs and non-reference mRNAs.

<table>
<thead>
<tr>
<th></th>
<th>DIG1</th>
<th>SFL1</th>
<th>HEK2</th>
<th>SAN1</th>
<th>ROF1</th>
<th>TOS8</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIG1</td>
<td>0.72</td>
<td></td>
<td>0.75</td>
<td>0.62</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td>SFL1</td>
<td></td>
<td>0.72</td>
<td>0.87</td>
<td>0.65</td>
<td>0.59</td>
<td>0.33</td>
</tr>
<tr>
<td>HEK2</td>
<td>0.75</td>
<td>0.87</td>
<td></td>
<td>0.71</td>
<td>0.62</td>
<td>0.29</td>
</tr>
<tr>
<td>SAN1</td>
<td>0.62</td>
<td>0.65</td>
<td>0.71</td>
<td></td>
<td>0.70</td>
<td>0.38</td>
</tr>
<tr>
<td>ROF1</td>
<td>0.49</td>
<td>0.60</td>
<td>0.62</td>
<td>0.70</td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>TOS8</td>
<td>0.48</td>
<td>0.33</td>
<td>0.29</td>
<td>0.38</td>
<td>0.37</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Effects of overexpressing genes identified in screen for modulators of the fluffy phenotype. A). Increase in copy number of DIG1, SFL1, HEK2, ROF1/YHR177W, SAN1 and TOS8 leads to a reduction in fluffy morphology in strain F45, grown on YPD with 2% glucose. B) Deletion of DIG1, SFL1, HEK2, ROF1/YHR177W, SAN1 and TOS8 leads to an increase in fluffy morphology in strain F13, grown on YPD with 1% glucose. C) When the catalytically dead san1-R280A allele is overexpressed in F45, colonies remain fluffy.
Figure 2. Initial variance in each overexpression strain (variance of log$_2$ fold-change relative to F45) and residual variance unexplained by common factor (using the communality values from factor analysis).
Figure 3. *SFL1*, *HEK2*, *SAN1* and *ROF1* overexpression profiles (relative to F45) vs the common factor, showing linear regression line and slope.
Figure 4. A). DIG1 overexpression profile (relative to F45) vs the common factor, with linear regression line and slope. B). TOS8 overexpression profile (relative to F45) vs the common factor.
**Figure S1.** Scree plot used to assess number of factors for factor analysis. Support for the first factor is strong, but support for the second is very weak, so only a single factor was used in the factor analysis.

**Figure S2.** Comparison of correlations between the differential expression profiles of overexpression strains for reference mRNAs vs non-reference mRNAs and ncRNAs. Expressed in terms of coefficient of determination ($R^2$).
Table S1. Log₂ fold changes and multiple hypothesis-corrected p-values for each overexpression strain, relative to F45.

Table S2. GO category, pathway and transcription factor enrichment analysis for genes strongly (>1.5-fold) and significantly (p<0.01 after multiple hypothesis correction) induced or repressed, relative to F45, in the common factor.

Table S3. GO category, pathway and transcription factor enrichment analysis for genes strongly (>1.5-fold) and significantly (p<0.01 after multiple hypothesis correction) induced or repressed, relative to F45, by TOS8 overexpression.

Table S4. GO category, pathway and transcription factor enrichment analysis for genes significantly (p<0.01 after multiple hypothesis correction) induced or repressed, relative to F45, in both the common factor and TOS8 overexpression profiles.

Table S5. GO category, pathway and transcription factor enrichment analysis for genes strongly (>1.5-fold) and significantly (p<0.01 after multiple hypothesis correction) induced or repressed, relative to F45, in the DIG1-residual profile.

Table S6. Log₂ fold changes for non-coding transcripts and non-reference genes, relative to F45, and multiple hypothesis-corrected p-values for each overexpression strain, relative to F45.
References


Irie, K., T. Tadauchi, P. A. Takizawa, R. D. Vale, K. Matsumoto et al., 2002 The Khd1 protein, which has three KH RNA-binding motifs, is required for proper localization of ASH1 mRNA in yeast. EMBO J 21: 1158-1167.


