It's not just fluff: mechanisms underlying the ability of *Saccharomyces cerevisiae* to build complex multicellular colonies

Zhihao Tan

A dissertation submitted in partial fulfillment

of the requirements for the degree of:

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Aimée Dudley, Chair

E. Peter Greenberg

Maitreya Dunham

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

It's not just fluff: mechanisms underlying the ability of *Saccharomyces cerevisiae* to build complex multicellular colonies

Zhihao Tan

Chair of the Supervisory Committee:

Aimée Dudley, PhD

Affiliate Faculty, Genome Sciences

Associate Investigator, Pacific Northwest Research Institute

Most colonies formed by *S.cerevisiae* are simple domes, not surprising considering the yeast is a unicellular organism. However, certain *S.cerevisiae* strains are able to form complex and intricately patterned colonies that involve the formation of multicellular structures. The colony patterns formed by these strains are highly reproducible, indicating that they result from a well defined developmental path. Understanding this remarkable ability could give us insights into important biological phenomena such as biofilm formation, biological shape and pattern determination and the genetic architecture underlying complex traits.

In this dissertation, the efforts taken to uncover the molecular mechanisms through which *S.cerevisiae* is able to achieve the formation of complex colony morphology are described. Through the thorough characterization of a novel switching phenomenon, we discover that the gain and loss of a single chromosome allows *S.cerevisiae* a quick, heritable and stable mechanism through which they are able to toggle their ability to form these complex colonies. This switch is made all the more remarkable with the realization that the phenotypic state of the colonies is able to confer fitness advantages in different conditions. The finding that the increased dosage of a single gene is sufficient for this switch then fuels an overexpression screen that uncovers novel suppressors of the trait. Through the further characterization of the transcriptome of several fluffy and smooth strains, we discover the importance of extracellular proteins for the formation of these colonies and how their expression is correlated with their molecular function.

In summary, the findings described here not only newly implicate several proteins in the modulation of the trait and highlight the distinct transcriptional regulation of the mechanistic effectors of the trait, but also provide further insight into how aneuploidy is able to modulate the phenotypic state and fitness of an organism.

Acknowledgements

Truth be told, it was only *after* I entered graduate school to study Molecular and Cellular Biology that I realized I did not know what graduate school actually entailed. Most of my previous education involved rote memorization of facts and figures, a very different set of skills in contrast to what I believe is actually needed in graduate school: among many others, the ability to consolidate information from diverse sources, the tenacity to persevere despite repeated failures and the application of a objective and critical mind to any set of data. I am thus filled with a slight sense of wonderment at having managed to survive graduate school, and though there is a slight hesitation in writing this section as I feel that there is little I have actually achieved or contributed, there are many people that I need to thank for having brought me this far.

First and foremost, my heartfelt appreciation goes towards my mentor, Aimée Dudley. Despite coming in starry-eyed and not knowing a scrap about systems biology or systems genetics (I still don't), she welcomed me with open arms into her lab. I believe she gave me just the right mix of guidance – close and tight mentoring at times when I was straying or floundering and needed it, freedom and trust at other times when I wanted to strike out on my own. Not only has she managed to put up with my many idiosyncrasies and shortcomings, she has done so staying completely positive and encouraging. This trust and faith that she has in me has given me confidence in my own abilities, and I leave graduate school better prepared for the path ahead because of everything she has done.

The rest of the Dudley lab, both present and past members, have also been amazing colleagues, their acceptance just as warm, and their encouragement and help always forthcoming. Michelle Hays has been the best friend that one can hope to have, and without her, half, or maybe even all of this dissertation might not be written; Cathy Ludlow has managed to keep the lab running in smooth order (really no small feat considering all the events that have taken place), allowing me to stay focused on my work; Gareth Cromie has always been a constant source of reason and calm, walking me back from the edge of many bioinformatics-induced cliffs; Amy Sirr has

provided unending optimism, her tag-team with Aimée will leave one dreaming of rainbow-filled waterfalls; Adrian Scott might be the coolest and most underrated person I have worked with; Eric Jeffery has provided top-notch and crucial support in my projects, his academic and not-entirely-academic discussions will be missed; Greg Cary will always be *the* senior graduate student in my heart, his words of wisdom helping me through graduate school; Cecilia Garmendia has made dreary days in lab that much more entertaining; Teresa Gilbert will always be fondly remembered. The team's maturity and commitment to scientific excellence will always be an inspiration.

The collaborators that I have worked with have also unselfishly brought their unique set of expertise into my projects: Alex Skupin, Patrick May and Jake Lin at the University of Luxembourg have shown me how world-class bioinformatics research should be done; Pekka Ruusuvuori at the Tampere University of Technology has shown such patience in our interactions, I feel disappointed at being unable to complete our collaboration on what would be a terribly exciting project.

I am proud to have a stellar set of committee members that have provided insightful advice and support despite their impossibly busy schedules: Pete Greenberg has gone above and beyond his role as GSR and provided invaluable mentoring; Maitreya Dunham and Ilya Shmulevich have have been enthusiastic and enjoyable people to bounce ideas with; Caroline Harwood has graciously stepped in when I needed her despite her countless commitments; Jennifer Nemhauser has been objectively critical but still fully encouraging in the early days. I also have to thank Uttam Surana at the Institute of Molecular and Cellular Biology in Singapore for keeping up with my progress and our wonderful and always entertaining scientific discussions on my trips home.

The Molecular and Cellular Biology department has been a great home for me. Directors Michael Emerman and Dave Raible were always ready to help and their commitment to and support of their students, especially me, is greatly appreciated. MaryEllin Robinson, Terry Duffey, Michele Karantsavelos, Milli Morris and Maria Sanders have been exceptional in their handling of administrative duties. I am thankful to the Agency for Science, Technology and Research in

Singapore for believing in me and funding my studies. The A*STAR scholarship officers, especially Jamie Tan, have made these 10 years go by as smoothly as possible.

Last but not least, I have to thank my family. It is only through their unwavering support that any of this has been possible.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1 - Background & Introduction

1.1 Defining the trait

Complex colony morphology of the yeast *Saccharomyces cerevisiae* is a trait that never fails to command one's attention, as it is so visually arresting. The endless supply of colony shapes and multicellular structures that are formed easily captures the imagination of anyone. Each genetic or environmental perturbation has the potential to give rise to a different colony morphology, and watching the temporal and spatial development of each new morphology is truly captivating (**Figure 1-1**).

FIGURE 1-1 - FLUFFY VOLCANO!

Perhaps what would be the most helpful would be to first define this trait. Investigators have used many terms to describe the colonies that are formed or the trait itself: formation of "mats"[1], fluffy

or smooth colonies[2-4], rough, complex or structured colony morphology[5-8] and biofilm colonies[5, 9-11]. The growth of these colonies has also been assessed in many different conditions: experiments describing mat formation are often performed on softer surfaces (0.3% agar vs. 2% agar for all other experiments)[1, 10]; the trait has been assessed on glycerol[2, 5, 9], sucrose[6], ethanol[8] or glucose[7, 11] as the carbon source. The variety of conditions that have been used might stem from whichever conditions investigators were able to obtain structured colonies on, as it has been seen that colonies from a single strain are able to develop very different shapes, including complete gain or loss of the trait, when grown on different nutrients and environments[3, 6, 7]. Despite the seeming lack of consensus of experimental conditions, the link that ties these studies together is the ability of the colonies to form complex, intricate multicellular structures and patterns that are beyond a simple hemispherical dome shaped colony that most laboratory strains form. While it might be expected that there would be condition-specific effects (i.e. phenotype x environment interactions) between studies, the strength of the phenotype and the ease of identification of changes in phenotype using visual inspection should mean that results from these different studies would still be well correlated with each other, and findings can be directly compared. Thus, while I have ensured that conditions in our own experiments are as controlled and precise as possible (described further in Chapter 2), I have chosen to use the terms fluffy and smooth colonies, complex colony morphology and biofilms interchangeably in this manuscript, sometimes when I feel a term is particularly apt or if a chapter has a particular focus.

However, with regard to the use of the term biofilms, there is an important distinction to be made. Some bacterial and fungal biofilms grow in environments that are entirely submersed, e.g. catheters[12], biofilm reactors[13], food processing units[14], while others are present at an air-solid interface, e.g. respiratory tract[15], oral cavity[16], wounds[17]. As we are assaying the growth of colonies on solid media, our findings will most likely be more applicable to biofilms formed at the air-solid interface and thus direct comparisons will be made to studies of this type. The studies of *S.cerevisiae* biofilms in liquid environments[18-20] might not be as relevant and different genes and pathways will probably be involved, e.g. while *FLO11,* encoding a cell-surface adhesin, is

necessary for complex colonies on solid media[1], *FLO11* is not necessary in the formation of liquid biofilms[20].

1.2 THE IMPORTANCE OF COMPLEX COLONY MORPHOLOGY

There are many areas for which the findings derived from the study of complex colony morphology could be particularly pertinent. I will give a brief introduction on how the study of this trait can help us in our understanding of biofilm formation, biological pattern development and the dissection of complex genetic traits.

BIOFILM FORMATION

Many opportunistic human pathogens form biofilms - highly structured, multicellular communities that are a key factor in persistent infection and antimicrobial drug resistance[21-23]. Biofilms are a major cause of medical-device associated infections[21, 22, 24], chronic infections of the oral[25], respiratory[15] and urinary tract surfaces[26], and their formation often leads to chronic non-healing of wounds[17, 25]. Their natural ability to adhere to many organic and inorganic surfaces and their inherent drug resistance make them exceptionally difficult to remove and are therefore a significant and pressing clinical problem[21, 22, 27]. The transition from a planktonic, unicellular lifestyle to a sessile, multicellular lifestyle requires the coordinated activation and repression of numerous genetic and biological programs. While some of these programs and pathways have been elucidated[23, 28], many still remain poorly characterized. A greater understanding of the molecular mechanisms that lead to biofilm formation will help to guide the development of drug therapies that specifically target biofilms, an area that is still greatly underdeveloped[27, 29].

While most *S.cerevisiae* laboratory strains form smooth, unstructured colonies when grown on solid media, strains that are able to form these highly structured fluffy colonies very often form colonies that resemble bacterial and fungal biofilms[3, 30]. The recognition that these fluffy colonies could be highly similar to biofilms led to the establishment of *S.cerevisiae* as a model to study biofilm formation[31]. Since then, investigators have found that these fluffy colonies do indeed

harbor properties that are not just characteristic but also deterministic of biofilms, e.g. presence of an extracellular matrix (ECM)[2, 9, 32], localized expression of drug efflux pumps[9], use of cell-cell communication[33], reduced penetration of chemicals into the colony interior[9] and increased adherence to inorganic surfaces[1]. The lack of pathogenicity of *S.cerevisiae* and the comprehensive resources that the *S.cerevisiae* community has at their disposal[34] thus make the study of complex colony morphology a great and appealing model from which general principles that guide biofilm development can be obtained.

PATTERN DEVELOPMENT

The development and variation of shapes and patterns in biology is one that is inextricably linked to developmental biology and studies of comparative and evolutionary morphogenesis. The nature of some of the signals and feedback that guide the differentiation of cells into spatially specialized tissues are common across nature, e.g. morphogen gradients guide the development of plants, *Drosophila* blastoderms and imaginal discs and vertebrate neural tube and limbs[35, 36]. This appearance of general principles that can guide our understanding of tissue and organismal development makes the use of model organisms like *S.cerevisiae* feasible. In addition, the advancement of developmental biology will be greatly enhanced by work involving computational modeling and *S.cerevisiae* has long been a favored organism to carry out theoretical work in.

Despite its use as a model organism for a long period of time, it has only been recently appreciated that smooth *S.cerevisiae* colonies are able to produce colonies that have clear spatial differentiation of cell types and gene expression: an ammonium transporter (*ATO1*) is only expressed on the peripheral layer of yeast colonies[37, 38]; sporulation in a yeast colony takes place in a defined pattern[39]; the upper and lower cells in a yeast colony have distinctly different metabolic states that mimic the difference between tumor and healthy metazoan cells[40]. While it is clear there is much left that we can learn even from smooth colonies, the multitude of shapes that fluffy *S.cerevisiae* colonies are able to form, the endless development trajectories that they can take, and the ease at which they are able to change their development pattern (e.g. due to a

change in temperature or nutrient concentrations)[6, 7] make them ideal systems in which studies of spatial organization and temporal development can be performed in.

COMPLEX GENETIC TRAITS

Unlike Mendelian traits that can be explained fully by the segregation of a single locus through a cross, complex genetic traits arise through the effects of multiple loci[41]. Depending on the trait, the complexity (e.g. unequal effect sizes) and number of loci that are implicated can vary; in some cases upwards of hundreds of loci can be implicated in the trait, e.g. human height, where at least 180 loci have been found[42]. Due to this large number of loci, these traits often display a continuous distribution of trait values in the population. However, depending on the trait, there could also be a dichotomous (binary) expression of the trait value; these complex binary traits could arise when a threshold character is set over an underlying continuous variable[43].

In many of these complex traits for which quantitative trait loci (QTLs) have been mapped, the loci that investigators find are unable to fully explain the variation seen in the data. This gives rise to the case of "missing heritability", which has been at the forefront of much discussion following the recent completion of multiple genome-wide association studies (GWAS)[44, 45]. Some of the proposed causes of this "missing heritability" are: insufficient sample size, presence of rare alleles that cannot be detected by current genotyping efforts (as the genetic variants are not present on the microarrays), chromosomal structural variants (i.e. copy number variation, translocations, etc.), population structure in the data that is unaccounted for and genetic interactions of 2 or more loci that were unable to be picked up due to insufficient statistical power[44, 45].

The use of *S.cerevisiae* to dissect complex traits helps to get around some of these problems. The rapid generation of a large pool of segregants from a single cross gives an investigator the large sample size needed for the statistical power to detect alleles of minor effect and possibly look for signals indicating interaction between 2 or more loci. A recent investigation that utilized 1008 segregants from a single yeast cross claimed to be able to explain almost all of the additive heritable variation for 46 traits[46]. The segregation patterns of complex colony morphology suggest

that it is also a highly complex genetic trait that can be studied in an effort to understand the genetic architecture underlying some of these complex traits (discussed more in Appendix C).

1.3 PROGRESS OF THE FIELD

The study of colony morphology in *S.cerevisiae* started at the turn of the millennium with the initial characterizations of a vineyard isolate that was naturally able to form fluffy colonies[30], and the realization that the lab strain Σ1278b was able to form elaborate multicellular structures on low-agar plates[1]. Even at this time, *FLO11* was already suggested as a gene that is necessary for the trait in Σ1278b[1], and all following studies thus far[6, 7], including those using different strains[4, 5, 33], have agreed with this finding and placed *FLO11* as a central protein in this trait. It is still unclear the exact mechanism through which *FLO11* contributes this this trait, though its documented role as a cell surface adhesion protein[47], its ability to be shed from the cell surface and thus likely contribute to the extracellular matrix[32, 48], and the intriguing possibility that it forms "velcrolike fibers" between cells[9] should all contribute in part.

Investigations following these initial studies aimed mostly to physiologically characterize these fluffy colonies. As mentioned above, one of the major findings was the identification of an extracellular matrix that surrounds the cells in these fluffy colonies[2]. This matrix contains a high-molecular-weight protein/protein complex of >200kDa that is both proteinaceous and heavily glycosylated. But while this matrix is easily extracted, it is resistant to the action of many glycosidases, leading to the inability to characterize it further[2]. This ECM could be responsible for the higher wet colony biomass that is seen in these fluffy colonies, a finding that indicates that the water storage capabilities are significantly different between fluffy and smooth colonies[4]. These fluffy colonies also exhibit colony-surface specific activity of multidrug-resistance transporters[9], reminiscent of the activation of similar drug efflux pumps seen in bacterial biofilms.

A few gene expression studies have also been performed in natural variants. These studies stem from the basis of a peculiar phenomenon, which is that strain variants that are naturally fluffy

sometimes “domesticate”, in which they lose their ability to form complex morphology at a particularly high rate[2]. These studies have shown that the transcriptomic and metabolic state between the fluffy strains and their smooth counterparts are highly different, with many carbohydrate, vitamin and amino acid metabolism/transport genes differentially expressed[2, 5].

More recently, many careful and detailed genetic screens have been performed in an attempt to map out the genes involved in colony morphology[6, 7, 10]. It should be noted that all these screens have been performed in a single strain background (Σ1278b). These studies have utilized transposon mediated mutagenesis[7, 49] or comprehensive deletion screens that cover nearly the entire genome[6, 7, 10] to look for genes contributing to the trait. Together, these studies show that hundreds of genes are involved in modulating this trait, and that this trait is finely tuned to sense changes in the environment, implicating the filamentous mitogen-activated protein kinase (MAPK)[6, 7, 49], Ras-cAMP-PKA[7, 49], target-of-rapamycin (TOR)[6], high osmolarity glycerol (HOG)[6] and RIM101[6, 49] pathways as playing important roles in modulating the trait.

1.4 REFERENCES

1. Reynolds, T.B. & Fink, G.R. Bakers' yeast, a model for fungal biofilm formation. *Science* **291**, 878-881 (2001).
2. Kuthan, M. et al. Domestication of wild Saccharomyces cerevisiae is accompanied by changes in gene expression and colony morphology. *Mol Microbiol* **47**, 745-754 (2003).
3. Casalone, E., Barberio, C., Cappellini, L. & Polsinelli, M. Characterization of Saccharomyces cerevisiae natural populations for pseudohyphal growth and colony morphology. *Res Microbiol* **156**, 191-200 (2005).
4. St'ovicek, V., Vachova, L., Kuthan, M. & Palkova, Z. General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol* **47**, 1012-1022 (2010).
5. Stovicek, V., Vachova, L., Begany, M., Wilkinson, D. & Palkova, Z. Global changes in gene expression associated with phenotypic switching of wild yeast. *BMC Genomics* **15**, 136 (2014).
6. Voordeckers, K. et al. Identification of a complex genetic network underlying Saccharomyces cerevisiae colony morphology. *Mol Microbiol* **86**, 225-239 (2012).
7. Granek, J.A. & Magwene, P.M. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823 (2010).
8. Taylor, M.B. & Ehrenreich, I.M. Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet* **10**, e1004324 (2014).
9. Vachova, L. et al. Flo11p, drug efflux pumps, and the extracellular matrix cooperate to form biofilm yeast colonies. *J Cell Biol* **194**, 679-687 (2011).
10. Ryan, O. et al. Global gene deletion analysis exploring yeast filamentous growth. *Science* **337**, 1353-1356 (2012).
11. Granek, J.A., Murray, D., Kayrkci, O. & Magwene, P.M. The genetic architecture of biofilm formation in a clinical isolate of Saccharomyces cerevisiae. *Genetics* **193**, 587-600 (2013).
12. Al Mohajer, M. & Darouiche, R.O. Prevention and treatment of urinary catheter-associated infections. *Curr Infect Dis Rep* **15**, 116-123 (2013).
13. Martin, K.J. & Nerenberg, R. The membrane biofilm reactor (MBfR) for water and wastewater treatment: principles, applications, and recent developments. *Bioresour Technol* **122**, 83-94 (2012).
14. Van Houdt, R. & Michiels, C.W. Biofilm formation and the food industry, a focus on the bacterial outer surface. *J Appl Microbiol* **109**, 1117-1131 (2010).
15. Kobayashi, H. Airway biofilms: implications for pathogenesis and therapy of respiratory tract infections. *Treat Respir Med* **4**, 241-253 (2005).
16. Ren, Y., Jongsma, M.A., Mei, L., van der Mei, H.C. & Busscher, H.J. Orthodontic treatment with fixed appliances and biofilm formation-a potential public health threat? *Clin Oral Investig* (2014).
17. Scali, C. & Kunimoto, B. An update on chronic wounds and the role of biofilms. *J Cutan Med Surg* **17**, 371-376 (2013).
18. Zara, G., Budroni, M., Mannazzu, I. & Zara, S. Air-liquid biofilm formation is dependent on ammonium depletion in a Saccharomyces cerevisiae flor strain. *Yeast* **28**, 809-814 (2011).
19. Bojsen, R.K., Andersen, K.S. & Regenberg, B. Saccharomyces cerevisiae--a model to uncover molecular mechanisms for yeast biofilm biology. *FEMS Immunol Med Microbiol* **65**, 169-182 (2012).
20. Vandenbosch, D. et al. Genomewide screening for genes involved in biofilm formation and miconazole susceptibility in Saccharomyces cerevisiae. *FEMS Yeast Res* **13**, 720-730 (2013).
21. Costerton, J.W., Stewart, P.S. & Greenberg, E.P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318-1322 (1999).
22. Donlan, R.M. & Costerton, J.W. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev* **15**, 167-193 (2002).
23. Fanning, S. & Mitchell, A.P. Fungal biofilms. *PLoS Pathog* **8**, e1002585 (2012).
24. Darouiche, R.O. Treatment of infections associated with surgical implants. *N Engl J Med* **350**, 1422-1429 (2004).

25. Mancl, K.A., Kirsner, R.S. & Ajdic, D. Wound biofilms: lessons learned from oral biofilms. *Wound Repair Regen* **21**, 352-362 (2013).
26. Tenke, P. et al. Update on biofilm infections in the urinary tract. *World J Urol* **30**, 51-57 (2012).
27. Ramage, G., Culshaw, S., Jones, B. & Williams, C. Are we any closer to beating the biofilm: novel methods of biofilm control. *Curr Opin Infect Dis* **23**, 560-566 (2010).
28. Finkel, J.S. & Mitchell, A.P. Genetic control of Candida albicans biofilm development. *Nat Rev Microbiol* **9**, 109-118 (2011).
29. Pierce, C.G. & Lopez-Ribot, J.L. Candidiasis drug discovery and development: new approaches targeting virulence for discovering and identifying new drugs. *Expert Opin Drug Discov* **8**, 1117-1126 (2013).
30. Cavalieri, D., Townsend, J.P. & Hartl, D.L. Manifold anomalies in gene expression in a vineyard isolate of Saccharomyces cerevisiae revealed by DNA microarray analysis. *Proc Natl Acad Sci U S A* **97**, 12369-12374 (2000).
31. Reynolds, L.E. et al. Tumour angiogenesis is reduced in the Tc1 mouse model of Down's syndrome. *Nature* **465**, 813-817 (2010).
32. Karunanithi, S. et al. Shedding of the mucin-like flocculin Flo11p reveals a new aspect of fungal adhesion regulation. *Curr Biol* **20**, 1389-1395 (2010).
33. Vopalenska, I., St'ovicek, V., Janderova, B., Vachova, L. & Palkova, Z. Role of distinct dimorphic transitions in territory colonizing and formation of yeast colony architecture. *Environ Microbiol* **12**, 264-277 (2010).
34. Botstein, D. & Fink, G.R. Yeast: an experimental organism for 21st Century biology. *Genetics* **189**, 695-704 (2011).
35. Rogers, K.W. & Schier, A.F. Morphogen gradients: from generation to interpretation. *Annu Rev Cell Dev Biol* **27**, 377-407 (2011).
36. Kicheva, A., Cohen, M. & Briscoe, J. Developmental pattern formation: insights from physics and biology. *Science* **338**, 210-212 (2012).
37. Vachova, L. et al. Architecture of developing multicellular yeast colony: spatio-temporal expression of Ato1p ammonium exporter. *Environ Microbiol* **11**, 1866-1877 (2009).
38. Traven, A. et al. Transcriptional profiling of a yeast colony provides new insight into the heterogeneity of multicellular fungal communities. *PLoS One* **7**, e46243 (2012).
39. Piccirillo, S. & Honigberg, S.M. Sporulation patterning and invasive growth in wild and domesticated yeast colonies. *Res Microbiol* **161**, 390-398 (2010).
40. Cap, M., Stepanek, L., Harant, K., Vachova, L. & Palkova, Z. Cell differentiation within a yeast colony: metabolic and regulatory parallels with a tumor-affected organism. *Mol Cell* **46**, 436-448 (2012).
41. Hill, W.G. Understanding and using quantitative genetic variation. *Philos Trans R Soc Lond B Biol Sci* **365**, 73-85 (2010).
42. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838 (2010).
43. Yi, N. & Xu, S. Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* **155**, 1391-1403 (2000).
44. Manolio, T.A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).
45. Eichler, E.E. et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**, 446-450 (2010).
46. Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-237 (2013).
47. Lo, W.S. & Dranginis, A.M. FLO11, a yeast gene related to the STA genes, encodes a novel cell surface flocculin. *J Bacteriol* **178**, 7144-7151 (1996).
48. Sarode, N., Miracle, B., Peng, X., Ryan, O. & Reynolds, T.B. Vacuolar protein sorting genes regulate mat formation in Saccharomyces cerevisiae by Flo11p-dependent and -independent mechanisms. *Eukaryot Cell* **10**, 1516-1526 (2011).

49. Furukawa, K., Furukawa, T. & Hohmann, S. Efficient construction of homozygous diploid strains identifies genes required for the hyper-filamentous phenotype in Saccharomyces cerevisiae. *PLoS One* **6**, e26584 (2011).

Chapter 2 - WORKING WITH A HIGHLY VARIABLE PHENOTYPE

One can imagine that working with a trait like fluffy in which the phenotype can be affected by so many different variables could be challenging, and indeed it was. In this chapter I outline my initial efforts to control as many variables as possible, aiming to bring as much experimental rigor and reproducibility to my results as I could.

2.1 LAYING THE FOUNDATION FOR THE DETAILED, QUANTITATIVE ASSESSMENT OF FLUFFY

While strains forming fluffy colonies have been seen before in both natural variants[1, 2] and lab strains (e.g. Σ1278b and SK1)[3], what might be unappreciated even in labs that work with these strains is the seemingly endless possibility of shapes that can be formed with each new genotype. In addition to that, nutrient type and availability, and with that inter-colony spacing, also contribute to the development of the colonies. An example of this inter-strain variation in morphology development can be seen in **Figure 2-1**; the inter-colony spacing affecting the development of colonies can be seen in strain F07.

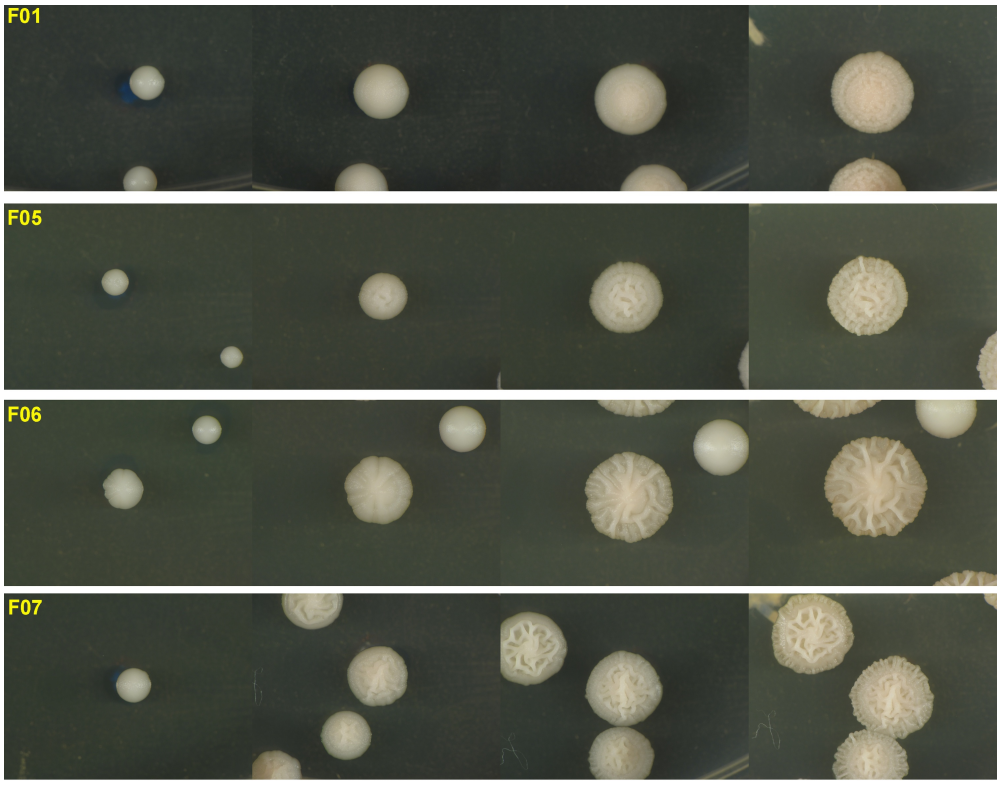


FIGURE 2-1 - INTER-STRAIN AND INTER-COLONY VARIATION IN COLONY DEVELOPMENT. EACH ROW DEPICTS THE DAILY DEVELOPMENT OF A DIFFERENT STRAIN. THE INTER-COLONY VARIATION IS PARTICULARLY NOTICEABLE IN THE 3RD IMAGE (FROM LEFT) FOR STRAIN F07.

To better control nutrient type and availability, an attempt was made to use Synthetic Complete (SC) media (constructed through precisely combining measured amounts of amino acids and vitamins) as the components would be clearly defined and highly reproducible, as compared to Yeast Peptone (YEP) media (constructed through the enzymatic digest of yeast cells), which possibly consists of variable amounts of components each time. However, the development of colonies was much slower on SC media, and structures were much less developed, making assessment of morphology difficult (**Figure 2-2**). The variation in morphology that arises from changing the carbon-source is also seen in our strains (**Figure 2-3)**. It is interesting to note that the effects differ with genotype, e.g. while ethanol seems to make F23 fluffier, it seems to make F45 less fluffy. A decision was made to assess the trait on YEP + glucose media, as this was the

most commonly used carbon source for *S.cerevisiae* studies, and still allowed the development of substantial complexity of colony morphology in our strains. Steps were taken to ensure experiments were as controlled as possible, e.g. plates for each experiment were from the same batch of media, media was filtered instead of autoclaved to prevent variable heat-degradation of nutrients associated with different autoclave runs, precise volumes of media were pipetted for each plate (25 ml for a 100mm-diameter Petri-dish, 30ml for an Omni-tray).

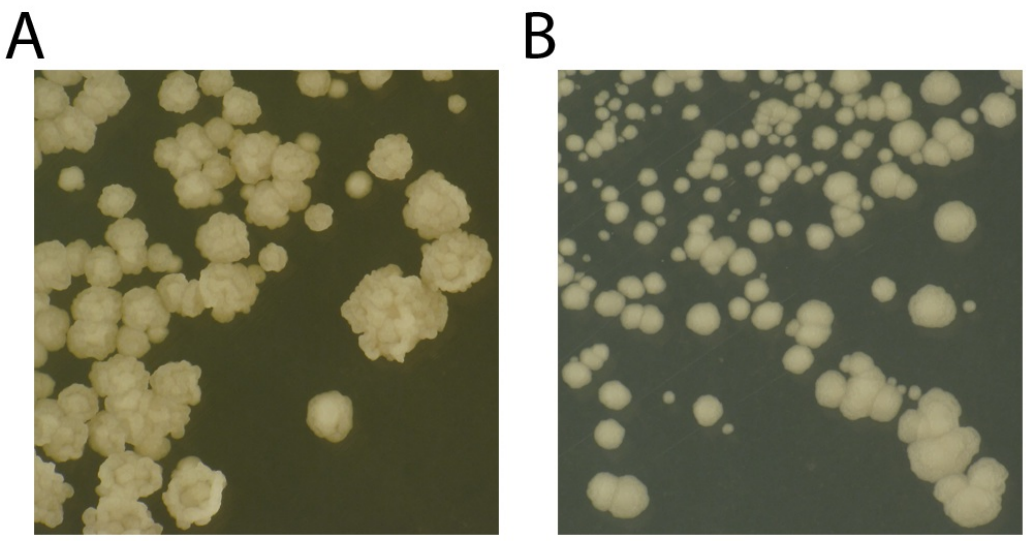


FIGURE 2-2 - COLONY MORPHOLOGY OF STRAIN F45 IS SIGNIFICANTLY REDUCED WHEN STRAINS ARE GROWN ON SC-URA MEDIA (B) AS COMPARED TO YPD MEDIA (A). IMAGES WERE TAKEN ON DAY 2 OF GROWTH

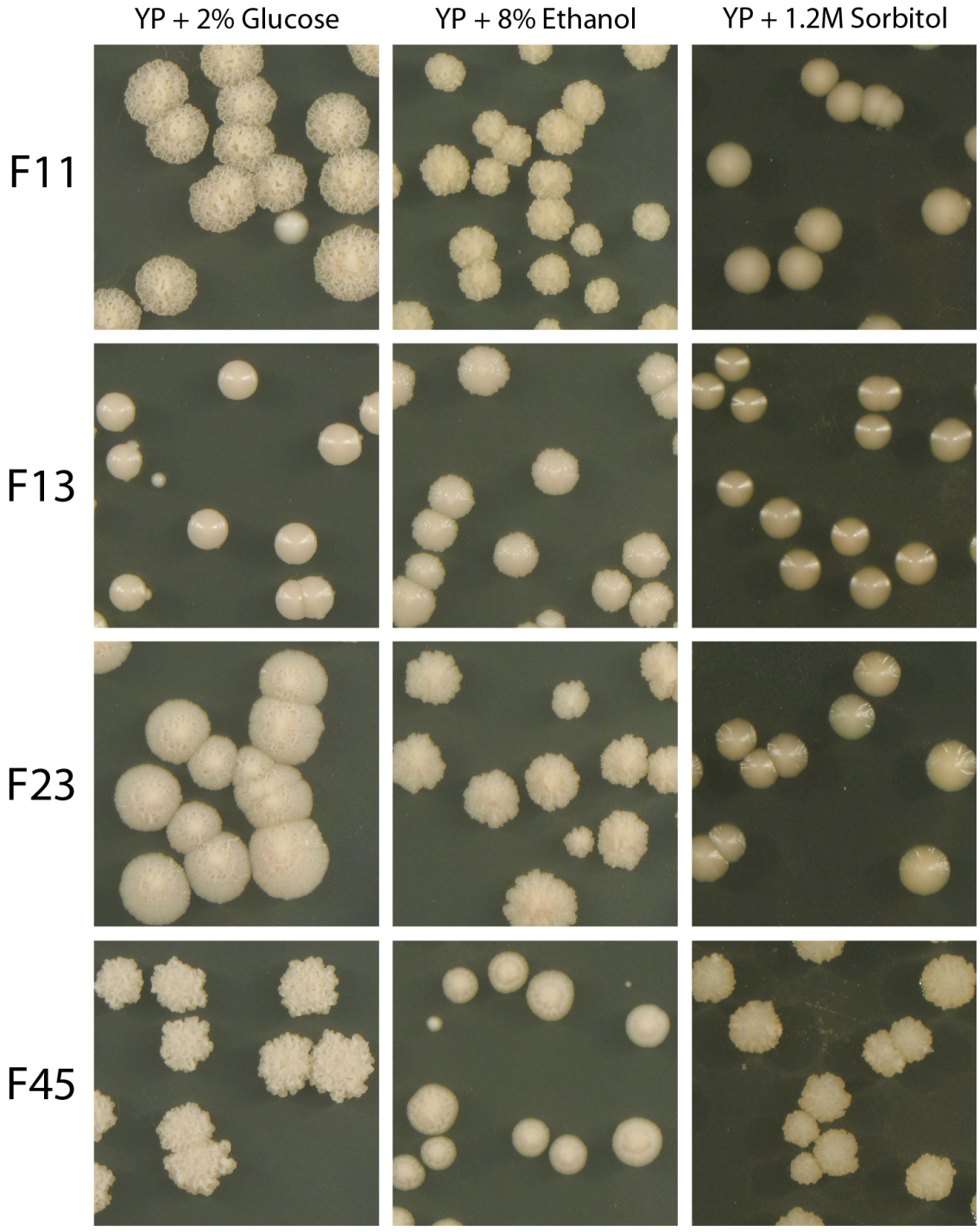


FIGURE 2-3 - EFFECTS OF MULTIPLE CARBON-SOURCES ON COLONY MORPHOLOGY

Next, I realized that spreading cells by glass beads and allowing colonies to grow up with random inter-colony spacing will introduce too much variability in the development of the colonies. To control for this, I decided to use a dissecting microscope that allowed the manipulation and precise placement of single cells. With micromanipulation, not only could the exact spacing between colonies be controlled, I could confirm that the colonies were in actuality developing from a single cell. This opened the ground for some of the phenotypic switching work that will be described in Chapter 4. Following some testing, I determined that the ideal spacing between colonies was 10mm on both the x- and y-axes, as this allowed enough space for proper growth of colonies up to day 5-6, and allowed enough replicate colonies to grow on the same plate to get a good assessment of the reproducibility of the morphology (**Figure 2-4A**).

While micromanipulation sufficed for most experiments, there was a requirement for a high-throughput method to lay down single cells for experiments that needed large numbers of biological replicates, e.g. experiments where colonies were harvested for RNA-seq. Together with Adrian Scott, we determined that it was possible to sort single cells using a FACS sorter (BD FACSAria II) onto an Omni-tray (Thermo Scientific). FACS sorting allowed for rapid placement of cells in any desired pattern, e.g. a 12 X 8 grid of 96 colonies with an inter-colony distance of 9mm; a “checker-board” layout of 48 cells with an inter-colony distance of 12.7mm (**Figure 2-4B**), etc.

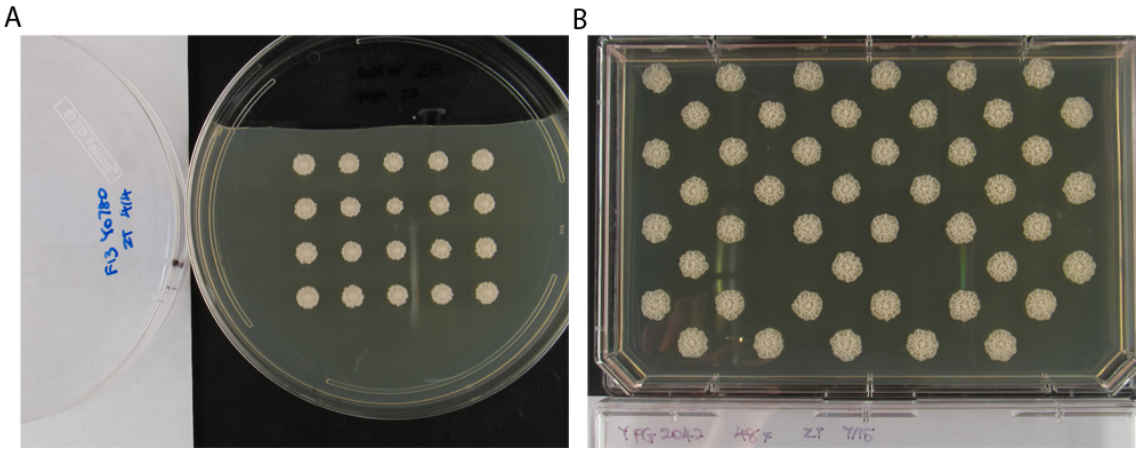


FIGURE 2-4 - REGULAR SPACING OF COLONIES ALLOWS FOR MORE CONTROLLED DEVELOPMENT. A) COLONIES MICROMANIPULATED TO A DISTANCE OF 10MM. B) COLONIES SORTED ONTO A "CHECKERBOARD" PATTERN WITH 12.7MM SPACING BETWEEN COLONIES.

With control of inter-colony spacing established, morphology variation between colonies is reduced, and the differences in morphology development due to carbon-sources and their concentrations can be much better appreciated. An example of a strain where these differences play a large role in development can be seen in **Figure 2-5**. While yet to be determined for every strain, as with previous studies, the strength of the phenotype usually increases with decreasing glucose concentrations[3, 4], while changing the carbon source to glycerol usually decreases the growth rate and completely changes the morphology. What can also be noted with this strain is the space that it needs to start becoming fluffy (**Figure 2-5,** compare 2% glucose, 96-well format to “checkerboard” format, where only the edges of the colonies in the checkerboard format are starting to turn fluffy). With this knowledge, the amount of glucose used or inter-colony spacing provided could be used in an experiment as a sort of dial to increase or decrease the fluffy morphology of strains.

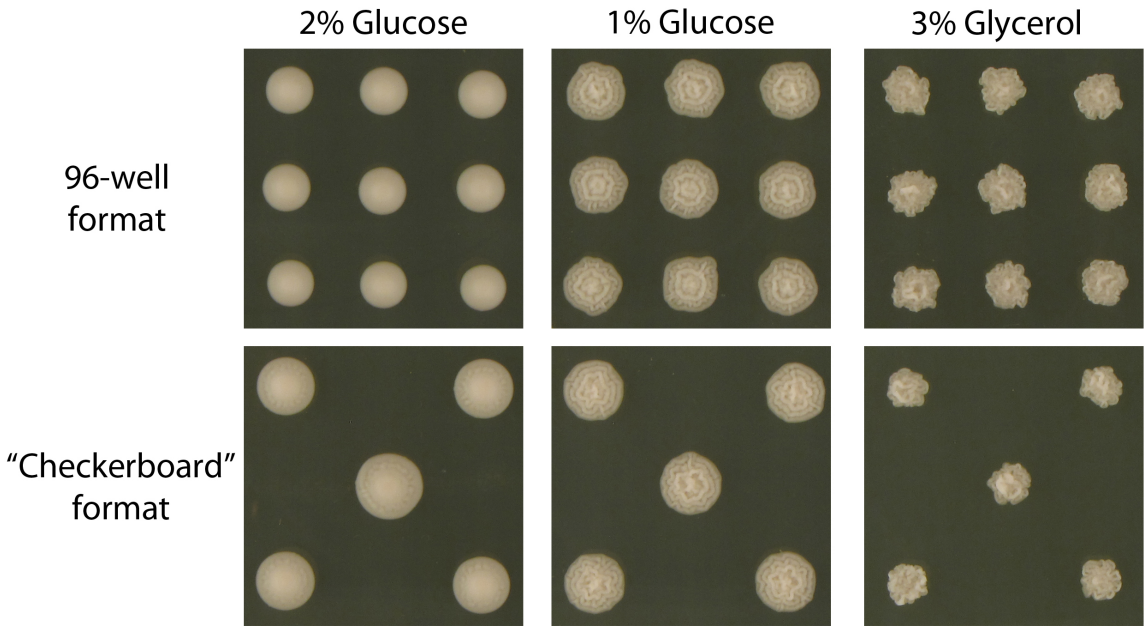


FIGURE 2-5 - NUTRIENT AND SPACING EFFECTS ON COLONY MORPHOLOGY OF A SINGLE STRAIN (YPG2042)

Another variable that should not be overlooked is temperature. The growth rate of strains will vary according to temperature, and if morphology is a function of growth rate vs. some other biological process, e.g. secretion of extracellular matrix material, unless the other process varies in tandem with the growth rate, the colony morphology will drastically change (**Figure 2-6**). Care should thus be taken for large scale experiments to ensure that imaging not involve removing strains from the incubator for long periods of time. An example of how this can confound an experiment is shown in **Figure 2-7**. In a large-scale experiment that was performed, due to the large number of plates that had to be imaged, plates were removed from the incubator for ~1-2 hours to a room where the temperature was lower (20-25°C). This very likely resulted in changes in morphology, with one strain showing particular temperature sensitivity and forming concentric rings that seemed to correspond to the number of times it was removed from the incubator (**Figure 2-7**).

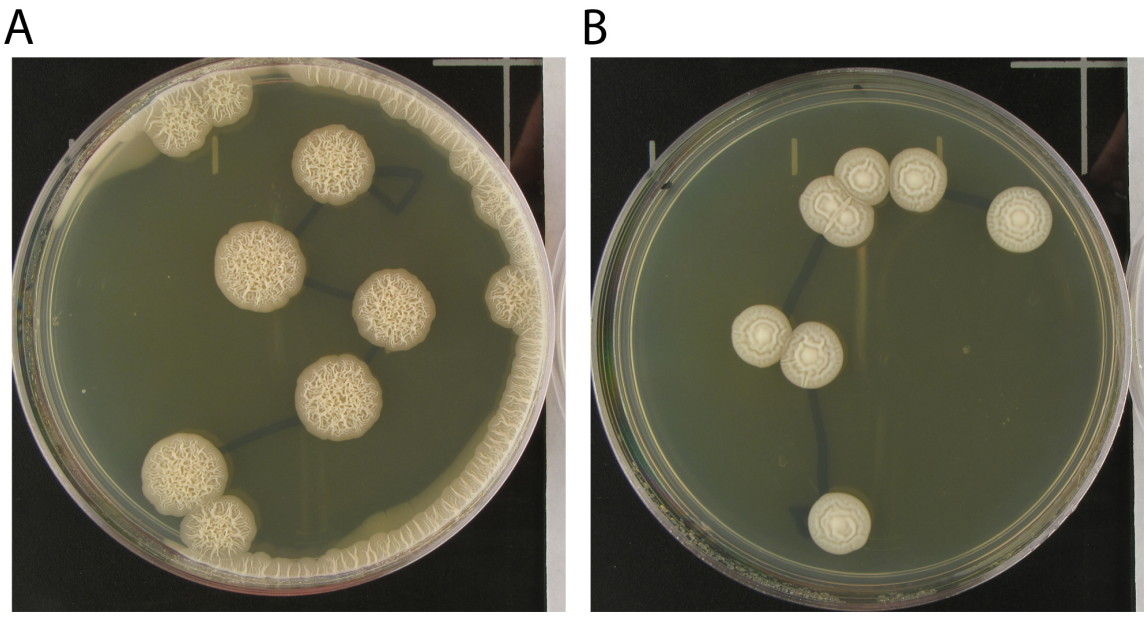


FIGURE 2-6 - EXAMPLE OF HOW TEMPERATURE CAN DRASTICALLY AFFECT COLONY MORPHOLOGY. A) CF60 (COHEN FLUFFY) COLONIES WERE ALLOWED TO GROW AT 30C FOR 7 DAYS. B) CF60 COLONIES WERE ALLOWED TO GROW AT 20C FOR 10 DAYS.

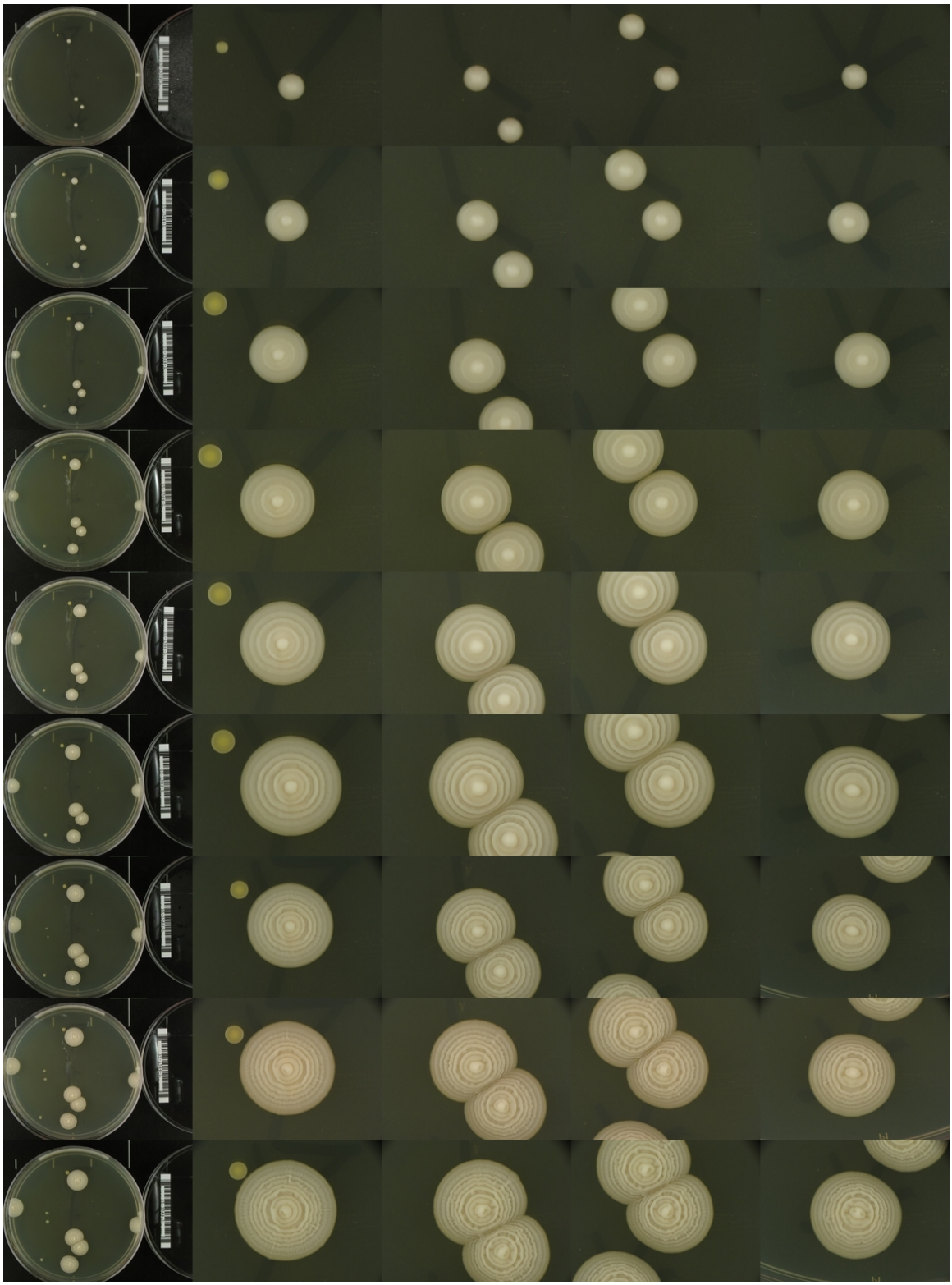


FIGURE 2-7 - MOVING PLATES IN AND OUT OF THE INCUBATOR CAN AFFECT COLONY MORPHOLOGY. NOTE THE REGULAR CONCENTRIC RINGS THAT ARE FORMING ON THE COLONY. STRAIN IS CF35 (COHEN FLUFFY).

With these variables controlled for, the reproducibility in the development of morphology is significantly raised, and one can start to really appreciate the consistency in colony patterns for many of these strains. Much smaller deviations in development can now be detected and reliably measured. This lays the foundation for more detailed measurements and analysis to be carried out, which will be discussed in Chapter 3.

2.2 REFERENCES

1. Cavalieri, D., Townsend, J.P. & Hartl, D.L. Manifold anomalies in gene expression in a vineyard isolate of Saccharomyces cerevisiae revealed by DNA microarray analysis. *Proc Natl Acad Sci U S A* **97**, 12369-12374 (2000).
2. Kuthan, M. et al. Domestication of wild Saccharomyces cerevisiae is accompanied by changes in gene expression and colony morphology. *Mol Microbiol* **47**, 745-754 (2003).
3. Granek, J.A. & Magwene, P.M. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823 (2010).
4. Voordeckers, K. et al. Identification of a complex genetic network underlying Saccharomyces cerevisiae colony morphology. *Mol Microbiol* **86**, 225-239 (2012).

Chapter 3 - THE DEVELOPMENT OF IMAGING PIPELINES AND COMPUTATIONAL TECHNIQUES TO ACQUIRE AND ANALYZE THE DEVELOPMENT OF COMPLEX COLONY MORPHOLOGY

This chapter is based on the following published paper:

Pekka Ruusuvuori[1,2], Jake Lin[1,2,3], Adrian C. Scott[4], Zhihao Tan[4,6], Saija Sorsa[1], Aleksi Kallio[5], Matti Nykter[5], Olli Yli-Harja[1,2], Ilya Shmulevich[1,2], Aimée M. Dudley[4,6]. Quantitative analysis of colony morphology in yeast. *Biotechniques.* 2014 Jan; 56(1):18-27

1. Department of Signal Processing, Tampere University of Technology, Tampere Finland
2. Institute for Systems Biology, Seattle, WA
3. Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg
4. Pacific Northwest Diabetes Research Institute, Seattle, WA
5. Institute of Biomedical Technology, University of Tampere, Tampere, Finland
6. Molecular and Cellular Biology Program, University of Washington, Seattle, WA

ACS, ZT, and AMD designed the experiments. ACS and ZT performed the experiments. AMD supervised the experimental work. PR, JL, and IS designed software and computational analysis. JL, PR, SS, and AK wrote software. PR and JL performed the computational analysis. MN, OYH, and IS supervised the software development and computational work. PR, ACS, ZT, AMD, JL, MN, and IS wrote the paper. All authors read and approved the manuscript.

Acknowledgements go to Cecilia Garmendia-Torres and Alexander Skupin for helpful discussions, and Tapio Manninen for advice with data analysis. This work was funded by a

National Institutes of Health Award (P50 GM076547/Center for Systems Biology) and a strategic partnership between the Institute for Systems Biology and the University of Luxembourg; Pekka Ruusuvuori is funded by Academy of Finland (project #140052); Zhihao Tan is funded by the Agency for Science, Technology, and Research (Singapore).

3.1 ABSTRACT

Phenotypic classification of complex colony morphology has traditionally relied on qualitative scoring systems that limit detailed phenotypic comparisons between strains. Automated imaging and quantitative analysis have the potential to improve the speed and accuracy of experiments designed to study the genetic and molecular networks underlying different morphological traits. In this chapter, the development of a platform that uses automated image analysis and pattern recognition to quantify phenotypic signatures of yeast colonies will be described. This strategy enables quantitative analysis of individual colonies, measured at a single time point or over a series of time-lapse images, as well as the classification of distinct colony shapes based on image-derived features. Phenotypic changes in colony morphology can be expressed as changes in feature space trajectories over time, thereby enabling the visualization and quantitative analysis of morphological development. To facilitate data exploration, results are plotted dynamically through an interactive web application that integrates the raw and processed images across all time points, allowing exploration of the image-based features and principal components associated with morphological development. The web application YIMAA is available at http://yimaa.cs.tut.fi.

3.2 INTRODUCTION

While studies aimed at characterizing the variation in colony morphology in *S. cerevisiae* have been as objective as possible, qualitative classification schemes, such as having a single investigator categorize colonies by eye, are still widely used[1-3]. Image analysis tools have been applied to the automated analysis of yeast colonies. The image analysis platform ImageJ[4] offers tools for processing quantifying colony images[5], and the image analysis tool CellProfiler[6] has been used to segment colonies on agar plates and group them based on shape, size and color information. Methods and software for quantifying colony growth combined with statistical analysis have also been presented[7, 8]. Other model organisms have also been subjected to

quantitative, image-based characterization and morphological classification. For example, image analysis has been applied to the automated screening of a variety of phenotypes (including morphology) in *Caenorhabditis elegans*[9], and recently an application case similar to ours was applied to the study of filamentous fungi using a set of over 30 morphological features[10].

Here, we describe an automated image analysis pipeline (**Figure 3-1**) that facilitates the quantitative study of colony morphology dynamics in large, time-lapse datasets. We start with automated image processing and then extract a large, generic set of quantitative descriptors. The combination of high-dimensional feature representation together with a sparse, supervised, logistic regression-based classification model is a powerful platform for the analysis of colony morphology. We have also built a web-based application to facilitate the intuitive exploration of the original raw and segmented time series images, the results of Principal Component Analysis (PCA), and hundreds of individual quantitative features. We test the accuracy of our method by using it to computationally distinguish the complex ("fluffy") and unstructured ("smooth") colony phenotypes[11, 12] based on image data from both single time points and fine resolution time lapses.

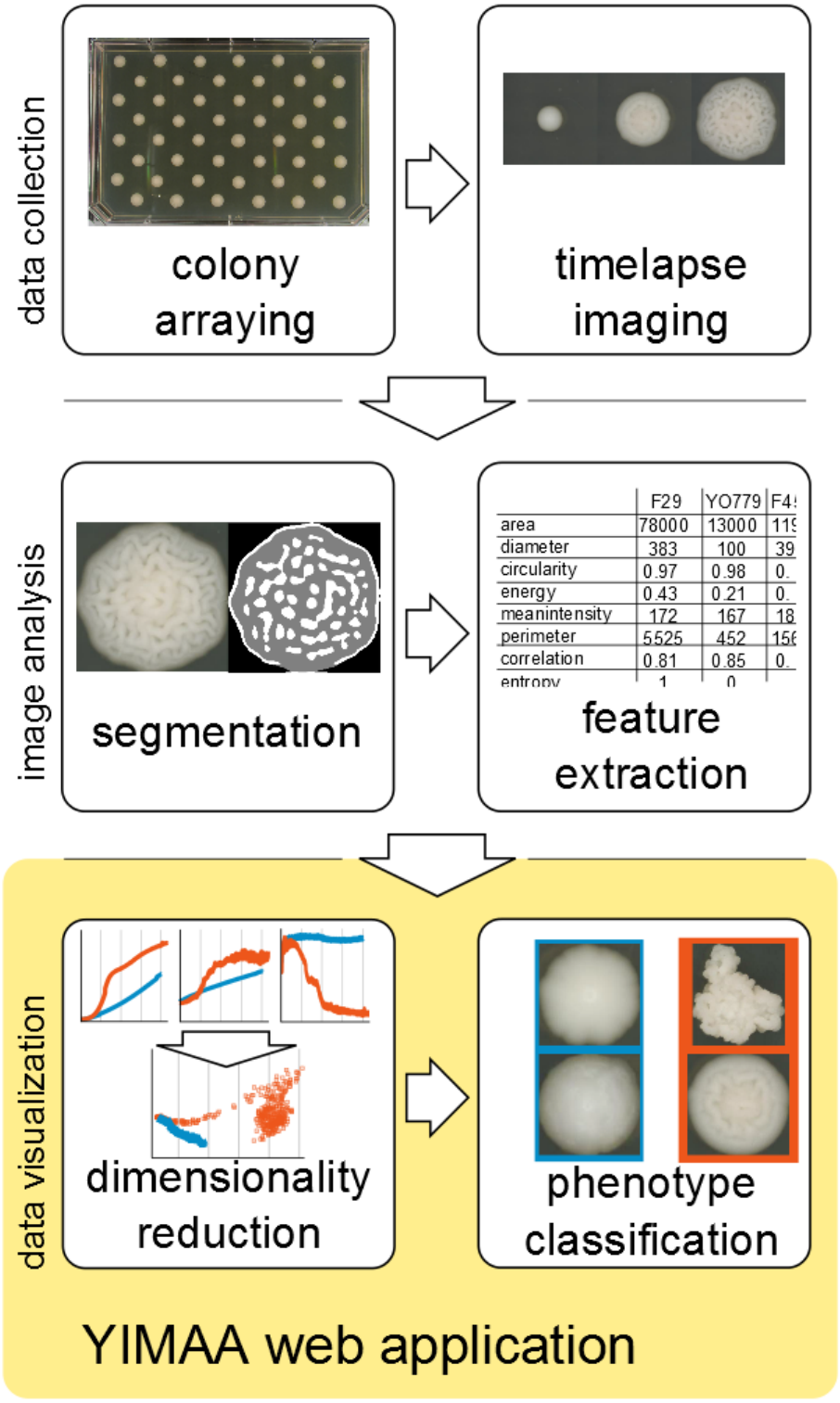


FIGURE 3-1 - THE COMPONENTS OF THE PLATFORM FOR AUTOMATED, QUANTITATIVE ANALYSIS OF YEAST COLONIES.

3.3 METHODS

YEAST STRAINS AND GROWTH CONDITIONS

Standard media and methods were used for the growth and genetic manipulation of *S. cerevisiae*[13]. All colonies were grown and imaged in a 30°C warm room on YPD (2% glucose) agar plates. Strains used in this study are described in **Table 3-1**.

TABLE 3-1 - *S. CEREVISIAE* STRAINS USED IN THIS STUDY

Name	Genotype	Source
FY4	MATa, Prototroph	F. Winston[14]
F7 (YPG385)	MATα *hoΔ::HphMX6, SPS2:EGFP:KanMX4*	This study*
F11 (YPG407)	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
F18 (YPG490)	MATα *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
F25 (YPG542)	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
F29 (YPG586)	MATα *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
F31 (YPG583)	MATα *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
F45 (YPG725)	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*, unmapped serine auxotrophy	A. Dudley[12]*
F47 (YPG746)	MATa *hoΔ::HphMX6, SPS2:EGFP:KanMX4*	This study*
F49 (YPG755)	MATa *hoΔ::HphMX6, SPS2:EGFP:KanMX4*	This study*
YPG339	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
YPG344	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
YPG348	MATa *hoΔ::HphMX6, SPS2:EGFP:KanMX4*	This study*

YPG352	MATα *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
YPG356	MATa *hoΔ::HphMX6, SPS2:EGFP:KanMX4*	This study*
YPG360	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*	This study*
YO779	MATa *hoΔ::HphMX6, SPS2:EGFP:NatMX4*, unmapped serine auxotrophy, ρ- or ρ^0	This study; a respiratory deficient isolate of F45

* Haploid segregant of a cross between UC5[15] and DBVPG1853[16]

COLONY IMAGING

Colonies used to distinguish the fluffy and smooth phenotype based on a single time point were generated by manually micromanipulating individual cells into a gridded pattern separated by 10 mm in both the x and y directions. Colonies were imaged after 5 days of growth using a PowerShot SX10IS camera outfitted with a DCR-250 macro lens (Raynox).

Colonies used for automated, time-lapse imaging were generated by depositing single cells 12.7 mm apart in a "checkerboard" pattern with a FACSAria II cell sorter (BD Biosciences). These colonies were imaged every 14 minutes for 5 days using a 5d Mark II camera outfitted with a MP-E 65mm 1-5x macro lens (Canon). The camera was attached to a custom built 2-axis gantry that moves the camera over the entire set of plates. Camera settings were held constant at an exposure time of 0.2 seconds and aperture of *f*/16. White balance was set using a grey card. Focus was held constant.

GENERATING QUANTITATIVE COLONY PHENOTYPE SIGNATURES USING IMAGE DERIVED FEATURES

The first step in our automated pipeline involves segmenting the colony area as the region of interest and extracting features that describe the colony shape, size, intensity, fractal, and texture. We segment using a straightforward, intensity-based global thresholding operation[17] and

then apply an additional size constraint to prevent detecting excessively small or large objects, which can arise from debris on the plate or camera lens flare. We also perform image border clearing to remove false segmentations that occur when colonies located close to plate borders have refraction from the edge of the plate incorrectly assigned to the colony. This first set of segmentation masks (**Figure 3-2A**) are used for the first round of feature extraction. The shape and size categories include basic descriptors for object morphology (e.g. area, convex area and roundness). Intensity-based features provide quantitative measures of the intensity distribution (e.g. intensity percentiles and deviation), whereas the texture features (e.g. intensity deviations in local area, texture features from grey-level co-occurrence matrices[18], histogram of oriented gradients[19] and local binary patterns[20]) take the spatial information into account.

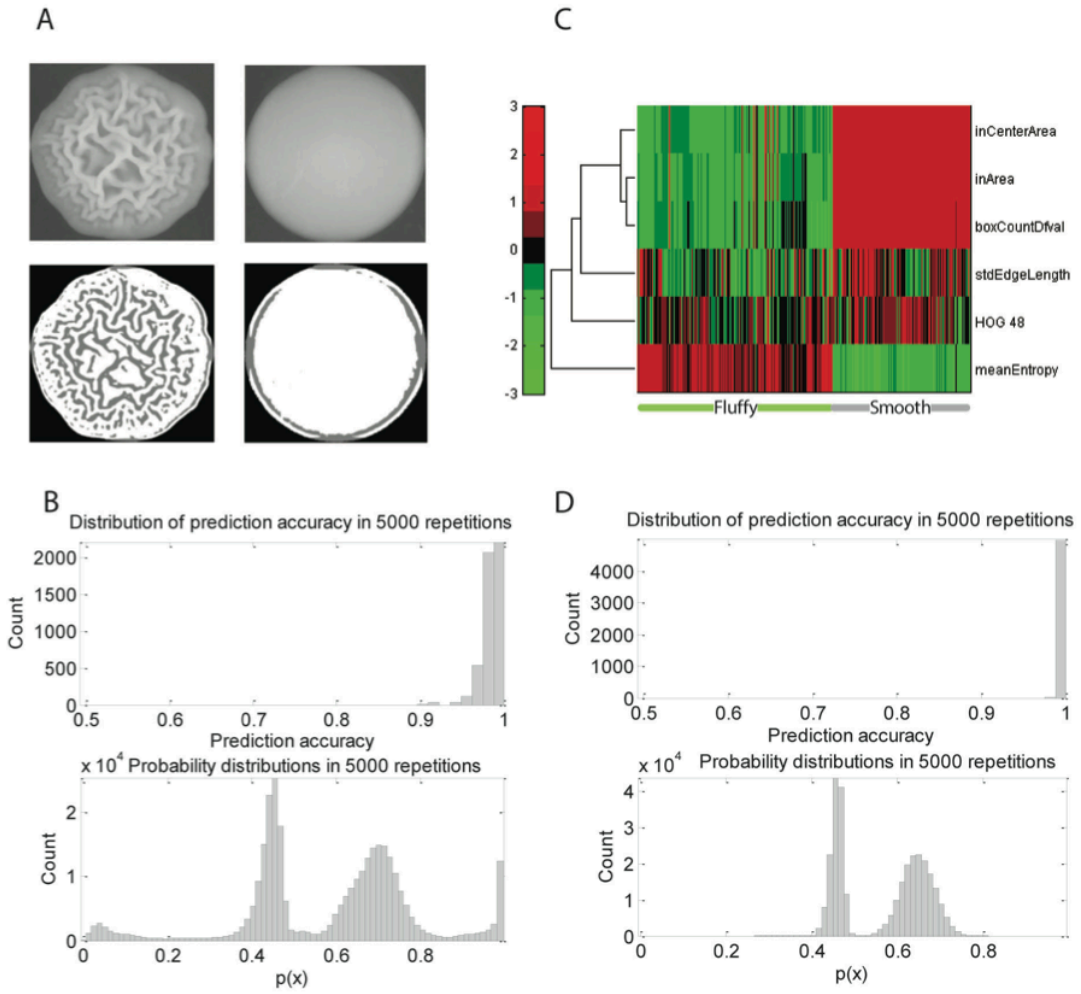


FIGURE 3-2 - PHENOTYPE ANALYSIS OF COLONIES FROM STATIC IMAGES. A) EXAMPLE IMAGES OF FLUFFY AND SMOOTH PHENOTYPES AND THE CORRESPONDING SEGMENTATION RESULTS. B) CLASSIFICATION ACCURACIES (TOP) AND PROBABILITY VALUES FOR CLASS REPRESENTING THE COMPLEX PHENOTYPES (BOTTOM) DURING THE 5000 REPETITIONS. C) HIERARCHICAL CLUSTERING OF THE SELECTED FEATURE SUBSPACE SHOWS HOW THE FEATURES CHOSEN BY THE LOGISTIC REGRESSION CLASSIFIER SEPARATE THE PHENOTYPES AND HOW THE COLONIES WITHIN PHENOTYPE SHOW SIMILAR FEATURE PATTERNS. D) CLASSIFICATION ACCURACIES (TOP) AND PROBABILITY VALUES FOR CLASS REPRESENTING THE COMPLEX PHENOTYPES (BOTTOM) DURING 5000 REPETITIONS OF HOLD-OUT ERROR ESTIMATION AFTER EXCLUDING RESPIRATORY DEFICIENT MUTANTS.

The next step involves an additional round of segmentation to detect shapes inside the colonies, visible as intensity changes in 2D projection images, and the extraction of a different set of features from the segmented images. For this segmentation we use a Difference of Gaussians

segmentation[21], where the difference of two low-pass filtered versions of the original image (highly blurred and slightly blurred) is thresholded. The two low-pass filters serve as a band-pass filter and the resulting binary image contains areas where intensity changes exist, but in which sharp variation, such as noise, is suppressed. Ideally, the resulting segmentation mask would be empty for a smooth colony and capture the colony shape for a fluffy colony. The features extracted from these second segmentation masks include descriptors containing information about the shapes detected inside the colony (e.g. area of the mask relative to the colony size, mask area in the center and border of the colony, number of objects in the mask, object sizes and deviations).

The combined feature set serves as a quantitative signature of colony phenotype, with colonies derived from the same strain or belonging to the same phenotypic class sharing similar characteristics among many of the features (**Figure 3-2D**). A detailed description of all 427 features is given in the Supplemental Materials. The feature list can be extended or trimmed without changes to the subsequent classification process.

SUPERVISED COLONY PHENOTYPE CLASSIFICATION

To transform these quantitative features into biologically meaningful phenotype information, we used a supervised classification strategy. To circumvent the need to specify the features used, we chose a classifier model with built-in feature selection, specifically the l_1 regularized logistic regression[22, 23], which produces sparse solutions and thus includes only a subset of the features in the model.

In logistic regression based classification, a feature vector $\boldsymbol{x}$ can be classified based on the conditional probability of belonging to the fluffy class given by the logistic regression algorithm as follows:

$$\mathrm{p}(\boldsymbol{x}) = \frac{1}{1 + \exp(\beta_0 + \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x})}$$

where $p(\boldsymbol{x})$ is the probability for the positive class given the feature vector $\boldsymbol{x}$ (i.e. $p(\boldsymbol{x}) = P(\text{"fluffy"}|X = \boldsymbol{x})$), and the parameters β_0 and $\boldsymbol{\beta}$ are estimated by maximizing the l_1 penalized log-likelihood

$$\sum_{\boldsymbol{x} \in F} \log \mathrm{p}(\boldsymbol{x}) \;\; + \sum_{\boldsymbol{x} \in S} \log\big(1 - \mathrm{p}(\boldsymbol{x})\big) \;\; - \lambda \|\boldsymbol{\beta}\|_1$$

where *F* denotes the fluffy class training samples, *S* is the smooth (non-fluffy) class training set, and λ is the parameter regularizing the sparsity of the solution. In practice, the solution is typically very sparse leading to computationally efficient models[24], with only a small subset of features receiving a non-zero weight in vector $\boldsymbol{\beta}$. Further, the use of logistic regression enables the extension to multi-class cases with more than two different strains or phenotypes.

QUANTITATIVE ANALYSIS OF COLONY SPATIOTEMPORAL DYNAMICS

Time-lapse image sequences are processed frame by frame as individual colony images once the colonies are large enough to be visible in the image (approximately one day of growth). The most obvious effect of colony growth is colony size, which also affects the quantification process. All features are extracted in the same manner from both small and large colonies. Feature trajectories are visualized by reducing the dimensionality with principal component analysis. Finally, a spatiotemporal profile of the yeast colony's development is built in which the spatial locations of the colony shapes are visualized over time by taking a cumulative sum of the colony shape segmentation masks.

WEB APPLICATION FOR DATA AND RESULT BROWSING

We have developed the Yeast Image Analysis (YIMAA) web application that serves as an interface for the original and binary segmentation images together with the time-lapsed plotting of quantitative phenotypic results. YIMAA is built using the open source components Highcharts.js,

jQuery, and jQuery plugins. The design of YIMAA focuses on interactivity and integration of images with dynamic time series plotting. Quantitative results are retrieved using AJAX. Image data are stored as assets organized by experiment and fetched on demand. The YIMAA web application is available at http://yimaa.cs.tut.fi. The source code for the project, including the implementation of the image analysis pipeline can be found at http://code.google.com/p/yimaa/.

3.4 RESULTS

Our study aimed to develop a generalized method for quantitatively representing the properties of microbial colonies. For this purpose, we selected a general feature set that is not tailored to a single strain or classification task. Extracting a large set of image-derived features that measure different characteristics of the colony also helps ensure that changes in the experiment or objects being studied, e.g. different magnifications, illumination settings, or strains, do not require significant alterations to the computational framework. Such generalization will facilitate its use in a variety of applications.

Our own research on yeast colony morphology has two experimental designs in which this general framework could be applied. First, the classification of colonies into “smooth” and “fluffy” classes at a single time point, which was performed manually in our previous work[12], could be performed more objectively and in higher throughput using image-derived features. Second, an automated image analysis pipeline could be used to extract quantitative features for many individual colonies as they grow and change shape over a series of time-lapsed images. In this framework, features extracted from the images form a vector of numerical values for each colony, where an element of the vector represents a feature value at the time point sampled. Both descriptions of colony morphology could be used to inform the genetic analysis of a relatively large number of yeast strains under a variety of environmental conditions.

To assess the discriminating power of our morphological signatures, we first tested whether the method could distinguish the smooth and fluffy morphologies using static images acquired at a

single time point (**Figure 3-2**). Smooth (YPG339, YPG 344, YPG348, YPG352, YPG356 and YPG360) and fluffy (F7, F11, F18, F25, F29, F31, F45, F47 and F49) yeast strains (**Table 3-1**) were grown on solid YPD medium. Twenty replicates (colonies) of each strain were photographed daily, and day five was selected as the static time point. Colonies that failed to grow were removed from subsequent analysis, yielding a dataset of 251 colony images. This dataset was analyzed and uploaded to the YIMAA web application. Representative images are shown in **Figure 3-2A**, with a fluffy colony in the upper left and a smooth colony in the upper right. The ternary-valued segmentation images (below the colony images) illustrate the region-of-interest identified by two rounds of segmentation, with the grey area corresponding to the intra-colony shapes. Quantitative features were then extracted from the images and normalized to zero-mean and unit variance.

We determined the average classification accuracy (98.79%) by performing a 4-fold cross validation for 5,000 repetitions with Monte Carlo random sampling on the 251 colony images described above. The upper panel of **Figure 3-2B** illustrates the distribution of classification accuracies for the validation partitions in the 5,000 loop trials. The lower panel of **Figure 3-2B** shows the distribution of probability values (also obtained from the 5,000 cross validation repetitions), where the probability of a sample $\boldsymbol{x}$ belonging to the "fluffy" class, p($\boldsymbol{x}$), is given by the logistic regression classifier. Classification is performed by dividing the probability space into two classes. In practice, p($\boldsymbol{x}$) < 0.5 corresponds to a smooth classification. Since the classifier is learned using ¾ of the samples chosen randomly at each repetition, the actual classification model varies between the trials and the values of model weight vector $\boldsymbol{\beta}$ change within the validation loop. To analyze the model behavior and learn which features are most informative, we collected the model parameter values in all 5000 trials. As expected, only a small number of features were used in the classifier model during the cross validation, with six features receiving a non-zero weight value in the model weight vector $\boldsymbol{\beta}$.

Next, we hierarchical clustered (in feature space) the colony image samples using the subset of six features shown to contribute to the classifier model during cross validation. The clustering

(**Figure 3-2C**) showed a clear separation between the fluffy and smooth strains, and the heat map reveals that colonies with the same phenotype share similar feature values. The selection counts confirm that, as expected based on the applied regularization, the logistic regression classifier produced a sparse model using only a small subset of the features. Thus, the classification results obtained with the regularized logistic regression classifier show that the features comprising phenotypic signatures can be used as a basis of classifying complex phenotypes in an automated manner when training samples are available.

Interestingly, the histogram of probability values in **Figure 3-2B** appeared to consist of two main distributions (large peaks on both the smooth and fluffy side) with additional, smaller peaks on each side. Such behavior suggested the existence of phenotypic subclasses or outlier samples. To explore this possibility, we analyzed the images that comprised these small peaks manually and discovered that they corresponded to cases of respiratory deficient mutants (RDM) that had arisen spontaneously from the corresponding "parental" strain. Since the ability to respire drastically affects colony size as well as the ability to form fluffy colonies[12], we removed all images from RDM samples. Repeating the classification procedure described above on the remaining 238 images resulted in a near perfect average classification accuracy (**Figure 3-2D**), with only 5 false predictions out of 300,000 classifications during cross validation. These probability distributions included only two modes, and together with the improved classification accuracy, suggested that the respiratory deficient mutants were indeed not covered by the two-class model. Finally, we tested whether the logistic regression classification framework could be used to define a third class consisting of respiratory deficient mutants (13 samples). With a limited sample size, we chose a simple leave-one-out cross validation, yielding 96.41% overall accuracy, with all fluffy and smooth samples classified correctly but only 4/13 RDM samples classified correctly. Thus, in this dataset considering the RDM samples separately gives improved classification accuracy for the fluffy and smooth phenotypes, but evaluating the applicability of the proposed framework for automated classification of RDM samples would require a larger dataset.

To test the ability of the method to analyze the spatiotemporal dynamics of colonies as they grow and change shape, we acquired a set of 18 time-lapse image sequences of four different strains (FY4, F29, F45 and YO779), where each sequence contained between one and three colonies. Features were then extracted over the course of the time-lapse, providing a quantitative representation (in feature space) of the morphological dynamics of colonies over time (**Figure 3-3A**). Examples of fluffy and smooth colonies at different times during development are shown in **Figure 3-3B**. We also generated “strain summaries” for each strain at each time point by taking the median value for each feature across all replicates. Both the feature profiles of each individual replicate (colony) and these strain summaries were then analyzed by principal component analysis allowing the trajectories in feature space as the colony develops to be visualized in reduced dimensions (**Figure 3-3C**). The time-lapse results (**Figure 3-3**) demonstrate that the feature dynamics quantified for fluffy and smooth colonies differ in the two example features, and that the PCA plots reveals different feature trajectories for different phenotypes.

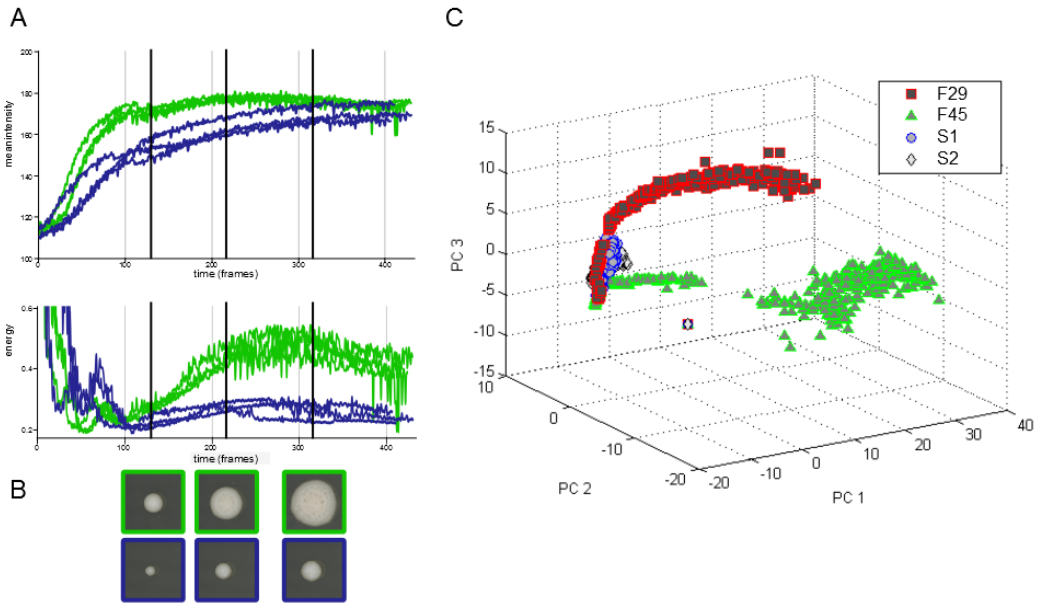


FIGURE 3-3 - ANALYSIS OF SPATIOTEMPORAL DYNAMICS OF YEAST COLONIES EXTRACTED FROM TIME COURSE DATA. A) EXAMPLES OF FEATURES 'MEAN INTENSITY' AND 'ENERGY' DURING A TIME-LAPSE MEASUREMENT; GREEN LINES ARE FOR THREE REPLICATES OF THE COMPLEX F29 STRAIN AND BLUE FOR THE SMOOTH STRAIN YO779. B) SNAPSHOTS OF COLONIES OF F29 (TOP) AND YO779 (BOTTOM) AT THREE TIMES DURING DEVELOPMENT (INDICATED BY BOLD VERTICAL LINES IN A). C) DIMENSIONALITY REDUCTION OF TIME-LAPSE FEATURE TRAJECTORIES USING PRINCIPAL COMPONENT ANALYSIS. THE TRAJECTORIES SHOWN ARE FOR STRAIN SUMMARIES, WHICH ARE OBTAINED BY TAKING MEDIAN ACROSS ALL INDIVIDUAL COLONIES AT EACH TIME POINT. STRAINS F29 AND F45 ARE FLUFFY, WHILE STRAINS S1 (FY4, TABLE 3-1) AND S2 (YO779, TABLE 3-1) ARE SMOOTH.

In addition to the image analysis software, we also developed a web application (YIMAA) that allows investigators to easily explore the results of the quantitative analysis alongside the raw input images from their experiment. The default page plots the PCA analysis results for an example from this study (strain F29). Users can also select multiple strains from the drop down list and their PCA results are plotted instantly. The plot can be animated to display points in order

across the time series allowing the user to explore the PCA values over time. This animation has “pause” and “play” functions. As the plotting advances, the gallery container shows the raw and segmented image of the most recently plotted point. YIMAA can also plot a time series of any of the several hundred individual features captured by the image analysis pipeline, and clicking on any time point brings up the associated images. Within the gallery panels, choosing a second strain permits side-by-side image comparison. A user guide and screen shots of the YIMAA web application can be found in the supplemental materials of the paper this chapter is based on.

3.5 CONCLUSIONS

A platform for the quantitative analysis of yeast colony morphology has been developed and its use for visualizing changes in colony morphology in feature space demonstrated. These quantitative colony morphology signatures can be used for supervised classification of colony phenotypes. These methods add statistical rigor to the analysis of colony morphology and will enable the use of a variety of computational tools, such as the classification and visualization tools described here, for the automated analysis of colony shapes. The automated aspect of the software can also enable studies at scales not possible for manual scoring, i.e. extremely large numbers of images. Finally, a web application has been built for easy and rapid sharing of results. This integrative environment for data exploration can be extended to other large-scale image analysis projects and to other colony-forming microorganisms.

3.7 REFERENCES

1. Granek, J.A. & Magwene, P.M. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823 (2010).
2. St'ovicek, V., Vachova, L., Kuthan, M. & Palkova, Z. General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol* **47**, 1012-1022 (2010).
3. Voordeckers, K. et al. Identification of a complex genetic network underlying Saccharomyces cerevisiae colony morphology. *Mol Microbiol* **86**, 225-239 (2012).
4. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
5. Dymond, J.S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).
6. Lamprecht, M.R., Sabatini, D.M. & Carpenter, A.E. CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques* **42**, 71-75 (2007).
7. Memarian, N. et al. Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC Bioinformatics* **8**, 117 (2007).
8. Dittmar, J.C., Reid, R.J. & Rothstein, R. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics* **11**, 353 (2010).
9. Wahlby, C. et al. An image analysis toolbox for high-throughput C. elegans assays. *Nat Methods* **9**, 714-716 (2012).
10. Posch, A.E., Spadiut, O. & Herwig, C. A novel method for fast and statistically verified morphological characterization of filamentous fungi. *Fungal Genet Biol* **49**, 499-510 (2012).
11. Kuthan, M. et al. Domestication of wild Saccharomyces cerevisiae is accompanied by changes in gene expression and colony morphology. *Mol Microbiol* **47**, 745-754 (2003).
12. Tan, Z. et al. Aneuploidy underlies a multicellular phenotypic switch. *Proc Natl Acad Sci U S A* **110**, 12367-12372 (2013).
13. Rose, M.D., Winston, F.M., Hieter, P. & Cold Spring Harbor Laboratory. Methods in yeast genetics : a laboratory course manual. (Cold Spring Harbor Laboratory Press, [Cold Spring Harbor, N.Y.]; 1990).
14. Winston, F., Dollard, C. & Ricupero-Hovasse, S.L. Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. *Yeast* **11**, 53-55 (1995).
15. Fay, J.C. & Benavides, J.A. Evidence for domesticated and wild populations of Saccharomyces cerevisiae. *PLoS Genet* **1**, 66-71 (2005).
16. Liti, G., Barton, D.B. & Louis, E.J. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics* **174**, 839-850 (2006).
17. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62-66 (1979).
18. Haralick, R.M., Shanmugam, K. & Dinstein, I. Textural Features for Image Classification *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610-621 (1973).
19. Ludwig, O., Delgado, D., Gonçalves, V. & Nunes, U. in 12th International IEEE Conference on Intelligent Transportation Systems, ITSC '09 1-62009).
20. Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971-987 (2002).
21. Russ, J.C. The image processing handbook, Edn. 6th. (CRC Press, Boca Raton; 2011).
22. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267-288 (1996).
23. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
24. Manninen, T., Huttunen, H., Ruusuvuori, P. & Nykter, M. Leukemia prediction using sparse logistic regression. *PLoS One* **8(8): e72932. doi:10.1371/journal.pone.0072932** (2013).

Chapter 4 - ANEUPLOIDY PROVIDES A MECHANISM FOR SWITCHING BETWEEN FLUFFY AND SMOOTH GROWTH STATES

This chapter is based on the published paper:

Zhihao Tan[1,2], Michelle Hays[1], Gareth A. Cromie[1], Eric W. Jeffery[1], Adrian C. Scott[1], Vida Ahyong[3], Amy Sirr[1], Alex Skupin[1,4] and Aimée M. Dudley[1,2]. Aneuploidy underlies a multicellular phenotypic switch. *Proc Natl Acad Sci U S A*. 2013 Jul 23; 110(30):12367-72

Boldface indicates equal contribution

1. Institute for Systems Biology, Seattle, WA
2. Molecular and Cellular Biology Program, University of Washington, Seattle, WA
3. Tetrad Program, University of California, San Francisco, San Francisco, CA
4. Luxembourg Center for Systems Biomedicine, University of Luxembourg, Luxembourg

ZT, MH, GAC, EWJ, A. Skupin, and AMD designed research; ZT, MH, EWJ, ACS, VA, and A. Sirr performed research; GAC, ACS, VA, and A. Sirr contributed new reagents/analytic tools; ZT, MH, GAC, ACS, and A. Skupin analyzed data; and ZT, MH, and AMD wrote the paper.

Acknowledgements go to members of the Dudley laboratory, Judy Berman, Ilya Shmulevich, Amir Sherman, Maitreya Dunham, and Sui Huang for helpful discussions and critical reading of the manuscript; Kirk Anders and Greg Prelich for providing plasmids. This work is funded by a strategic partnership between the Institute for Systems Biology and the University of Luxembourg. Zhihao Tan is funded by the Agency for Science, Technology, and Research, Singapore.

4.1 ABSTRACT

In the course of manipulating many of the fluffy strains present in our collection, we identified strains that are able to reversibly toggle between the fluffy and smooth colony-forming states. The high rates of switching and reversibility of phenotype hinted at an underlying cause that was not due to spontaneous mutations. This chapter describes the efforts that were taken to uncover the mechanism for the observed phenotypic switching. Using a combination of flow cytometry and high-throughput restriction-site associated DNA tag sequencing, we first show that this switch is correlated with a change in chromosomal copy number. Furthermore, the gain of a single chromosome is sufficient to switch a strain from the fluffy to the smooth state, and its subsequent loss to revert the strain back to the fluffy state. We further uncover that copy number imbalance of six of the 16 *S.cerevisiae* chromosomes and even a single gene can modulate the switch. Our results thus support the hypothesis that the state switch is produced by dosage-sensitive genes, rather than a general response to altered DNA content. These findings add a complex, multicellular phenotype to the list of molecular and cellular traits known to be altered by aneuploidy and suggest that chromosome missegregation can provide a quick, heritable, and reversible mechanism by which organisms can toggle between phenotypes.

4.2 INTRODUCTION

Phenotypic switching allows cells or organisms to reversibly toggle between multiple, heritable states[1-3]. These phenotypic states can confer distinct properties to cells that may be advantageous in some environments, e.g. increased resistance to heat[4], faster proliferation rates in the presence of excess nutrients[5], or altered interaction with host immune systems[6]. Although some phenotypic switches can be induced by a change in environment[7], others may arise stochastically, possibly as a bet-hedging strategy that increases the probability that a subpopulation of genetically isogenic cells survive a wider range of environmental perturbations[4, 8, 9].

The ability to switch phenotypes could prove especially useful for microorganisms that are unable to physically relocate to an environment of choice and need to weather environmental fluctuations. Indeed, microbes are known to adopt multiple phenotypic states, and though the nature of these states can vary widely, they are all characterized by rates of switching too high to be that of spontaneous mutation[2, 3]. Several interesting examples have been discovered in opportunistic fungal pathogens. Perhaps the best-characterized state switch in fungi is the white-opaque switch of the *Candida albicans*[10], which is regulated by transcriptional feedback loops[11] and is associated with differential virulence[12], host colonization[13] and mating competency[14]. *C. albicans* can also differentiate a subpopulation of "persister" cells that, like their bacterial counterparts, exhibit increased resistance to antimicrobial drugs[15]. *Candida glabrata* exhibits at least two spontaneous and reversible phenotypic switches: a "core switching system" and an "irregular wrinkle switching system"[16] which exhibit differential tissue colonization and clearing rates in vivo[17]. *Cryptococcus neoformans* switches between smooth, mucoid, pseudohyphal, and wrinkled colonies: state changes that are associated with biochemically distinct capsular polysaccharides[18], changes in karyotype[18], and differential fungal burden in infected host organs[19]. Because these different states may exhibit differential virulence, cell surface properties, or resistance to host immune defenses[2, 3], understanding the mechanisms underlying phenotypic switching could prove essential for the development of effective therapeutics against microbial pathogens.

Some strains of *S. cerevisiae* also exhibit a striking phenotypic switch. Although most laboratory strains produce relatively unstructured "smooth" colonies, some natural isolates of *S. cerevisiae* produce colonies with complex multicellular features, such as folds, crevices, and channels that form as the colony grows on solid media[20, 21]. These "fluffy" colonies possess properties similar to those of microbial biofilms, including the secretion and maintenance of an extracellular matrix[21-23], localized expression of drug efflux pumps[23], increased adherence[24], and the use of cell–cell communication[25]. Fluffy colony formation requires the function and coordination of numerous pathways that underlie the trait, and the deletion of key factors produces a smooth colony phenotype[26-28]. Fluffy colonies may benefit from both rapid acquisition of larger territories as well

as an increased capacity for nutrient and water transport[21]. Interestingly, some fluffy strains can switch to the smooth morphology at a rate significantly higher than that expected for spontaneous mutations[21, 29].

4.3 METHODS

YEAST STRAINS, MEDIA, AND MANIPULATION

Unless noted, standard media and methods were used for growth and genetic manipulation of yeast[30]. Strains used in this study are listed in **Table 4-1**. Colony morphology was assayed on YPD (2% glucose) plates. Respiratory competency was assayed by growth on YPG (3% glycerol) plates. The conditional centromere experiments used minimal medium (SC-Ura-His, Sunrise Science Products) to enrich for disomes and 5-FOA resistance (SC, 1 g/L 5-FOA, Zymo Research) to select euploid revertants. To ensure consistency in colony growth, all images of yeast colonies presented are from colonies arrayed into an ordered grid either by micromanipulation with 10-mm colony separation or FACS sorting at 12.7mm spacing.

TABLE 4-1 - *S. CEREVISIAE* STRAINS IMAGED IN THIS STUDY. WITH THE EXCEPTION OF YO795, YO796, YO880 AND YO881, ALL STRAINS ARE DERIVATIVES OF F45.

Strain Name	Genotype	Karyotype
F45	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
YO486[a]	MATa *SPS2:EGFP:kanMX4, hoΔ::HphMX6*	Disomy I, naturally occuring
YO502[b]	MATα *SPS2:EGFP:NatMX4*, *hoΔ::HphMX6*	Euploid haploid
YO785	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Disomy XVI, naturally occurring

YO885[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid, obtained from YO785
YO903	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Disomy XV, naturally occurring
YO904	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Disomy III, naturally occurring
YO958	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Disomy X, naturally occurring
YO963	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Disomy V, naturally occurring
YO983	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid, obtained from YO903
YO987	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid, obtained from YO904
YO1015	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, ura3Δ0, his3Δ*::KanMX4*, *cen16Δ::*P_{GAL1}*-CEN3-ura3::HIS3*, serine auxotroph	Euploid haploid
YO1064	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen16Δ::* P_{GAL1}*-CEN3-ura3::HIS3/cen16Δ::* P_{GAL1}*-CEN3-URA3*, serine auxotroph	Disomy XVI, conditional centromere
YO1099	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen16Δ::* P_{GAL1}*-CEN3-ura3::HIS3*, serine auxotroph	Euploid haploid, “2nd Gen”
YO1451	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid, obtained from YO958
YO1474	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine	Euploid haploid, obtained

	auxotroph	from YO963
YO1496	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen1Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen1Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy I, conditional centromere
YO1497	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen4Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen4Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy IV conditional centromere, + VII naturally occurring
YO1505	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen7Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen7Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy VII, conditional centromere
YO1510	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen8Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen8Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy VIII, conditional centromere
YO1532	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen9Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen9Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy IX, conditional centromere
YO1533	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen9Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen9Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy IX conditional centromere, + III naturally occurring
YO1542	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen12Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen12Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy XII conditional centromere, + XV naturally occurring
YO1543	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen12Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen12Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy XII conditional centromere, XV gained then lost naturally

YO1546	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen13Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen13Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy XIII, conditional centromere
YO1548	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen14Δ:: P_{GAL1}-CEN3-ura3::HIS3/cen14Δ:: P_{GAL1}-CEN3-URA3*, serine auxotroph	Disomy XIV, conditional centromere
YO795[c]	*SPS2:EGFP:NatMX4, hoΔ::HphMX6, gal7Δ0::kanMx4*	Euploid haploid
YO796[d]	*SPS2:EGFP:NatMX4,* hoΔ::dsdAMX4, *gal7Δ0::kanMx4*	Euploid haploid
YO880[e]	MATα *SPS2:EGFP:NatMX4*, *hoΔ::HphMX6*, *gal7Δ::KanMX4* (derived from YO795 x YO796 progeny)	Disomy XVI, naturally occurring
YO881[f]	MATα *SPS2:EGFP:NatMX4 hoΔ::HphMX6 gal7Δ::KanMX4* (derived from YO880)	Euploid haploid
YO898[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
YO959[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
YO965[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, respiratory deficient	Euploid haploid
YO971[g]	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, ura3Δ0, serine auxotroph	Euploid haploid
YO972[g]	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, ura3Δ0, serine auxotroph	Euploid haploid
YO977[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid

YO979[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
YO981[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
YO1101[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, *ura3Δ0*, *his3Δ::KanMX4*, *cen16Δ:: P_{GAL1}-CEN3-ura3::HIS3*, serine auxotroph	Euploid haploid
YO1540[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
YO1770	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *pFA6a-KanMX4* vector (AB352), biological replicate of YO1771	Euploid haploid
YO1771	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *pFA6a-KanMX4* vector(AB352), biological replicate of YO1770	Euploid haploid
YO1772	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *DIG1-pFA6a-KanMX4* vector (AB340), biological replicate of YO1773	Euploid haploid
YO1773	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *DIG1-pFA6a-KanMX4* vector (AB340), biological replicate of YO1772	Euploid haploid
F45_YSEQ127[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
F45_YSEQ128[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid
F45_YSEQ129[g]	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph	Euploid haploid

[a]Derived from UC5 (Sake)[31]
[b]Derived from DBVPG1853 (White Tecc)[32]
[c]Derived from EC-33 ("Evolution Canyon")[33]
[d]Derived from YPS163 (Oak tree)[34]
[e]Haploid progeny derived from a cross between YO795 and YO796
[f]Derived from YO880
[g]Reference Euploid Panel

Isolation of Smooth Colony Strains

To isolate strains with a smooth colony phenotype, cells from a fluffy colony were grown overnight at 30°C in liquid YPD (2% glucose). Cells were then plated onto YPD solid agar at a density of 200 cells per plate. Colonies were grown for 3 days at 30°C and then screened for the smooth colony morphology.

Imaging

All images were taken with a Canon PowerShot SX10IS, F8.0, 1/8s shutter speed, ISO80 under consistent lighting conditions, imaging distances and zoom. Images were cropped and scaled with Adobe Photoshop CS5.1 with no further modifications to the images.

Molecular Karyotyping by RAD-seq Coverage

The molecular karyotype of each strain was determined by multiplexed RAD-seq[35] of yeast genomic DNA as described[36]. For each strain, the coverage ratio of each marker is compared with the haploid reference and plotted against a DNA marker index ordered by genomic position (chromosomes I to XVI). Marker-to-marker variation was normalized using a panel of euploid strains. Raw RAD-seq reads of strains for which ploidy was analyzed have been deposited in the National Center for Biotechnology Information Sequence Read Archive (www.ncbi.nlm.nih.gov/sra/) with accession no. ERP002462.

CONDITIONAL CENTROMERES

Replacement of the native centromeres of chromosomes with a conditional centromere was adapted from a protocol using a well-characterized construct[37] that has been modified further[38]. Briefly, galactose induction was performed by first growing the strains in YP-Raffinose for 48 hours at 30°C to saturation and diluting 1:2,000 for growth overnight at 30°C. Cultures were then diluted to 5×10^6 cells per ml, split into two aliquots, and galactose (1.5% final concentration) was added to one. Cultures were monitored by cell count until cell density had increased 2.5-fold (~3–4 h). 1×10^7 cells were then plated onto SC-His-Ura plates to select for yeast disomic for the chromosome that contains the conditional centromere. In our hands, ~$1–3 \times 10^3$ colonies were recovered on each selection plate. Additionally, we see a high frequency of respiratory deficient colonies on our selection plates. Because respiratory competency is generally required for the fluffy phenotype, we checked all colonies obtained for growth on glycerol plates and only analyzed respiratory competent strains. For 5-FOA selection, 10^7 cells from colonies growing up on SC-His-Ura plates were resuspended in PBS and plated onto 5-FOA plates using glass beads. Plates were allowed to grow for 3 days before colonies were selected for further analysis. An average of ~$1–2 \times 10^2$ colonies were recovered on each selection plate.

PLOIDY DETERMINATION BY FLOW CYTOMETRY

Flow cytometry was used to verify the haploid DNA content of each strain for which an image is shown in the manuscript. Protocols for ploidy determination by DNA content using flow cytometry were adapted from published methods[39, 40]. To rule out the possibility of self-diploidization, the DNA content of known haploid (BY4741) and diploid (BY4743) strains were compared to F45 background strains. All F45 background strains used in morphology assays showed peaks corresponding to haploid DNA content. The relatively small 1N peak and peak near 750 in our F45 background strains are very likely due to cell clumping, with this effect being stronger in the more flocculent F45 fluffy strain (**Figure 4-1**). This effect is similar to published observations in

other non-laboratory strains[41]. The histograms shown are representative of gated cell populations, and the scatter plots show the gates applied.

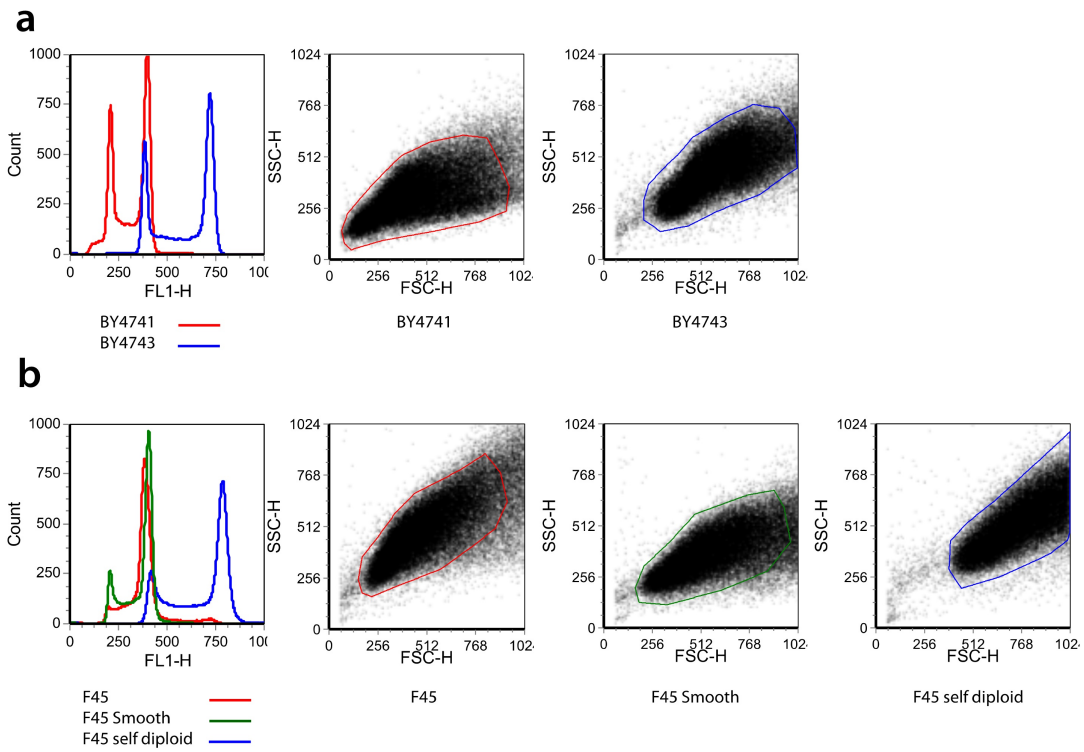


FIGURE 4-1 - DNA CONTENT ASSESSED BY FLOW CYTOMETRY. A) HISTOGRAM AND APPLIED GATES FOR LABORATORY STRAIN CONTROLS BY4741 (HAPLOID, RED) AND BY4743 (DIPLOID, BLUE) B) HISTOGRAM AND APPLIED GATES FOR F45 (RED), F45 SMOOTH (YO785, GREEN) AND AN EXAMPLE OF AN F45 SELF-DIPLOIDIZED STRAIN (BLUE)

DIG1 PLASMID CONSTRUCTION

Plasmid pAB340 was constructed by cloning *DIG1* and the surrounding intergenic region into a centromere plasmid with a G418 resistance marker. *DIG1* and its native promoter were cloned from BY4741 using the following primers into vector pFA6a-*KanMX4*[42] utilizing Clontech In-Fusion HD cloning kit following the manufacturer's protocol.

DIG1_infusionconstr3_F: 5'- GAATTCATCGATGATGCTCTTTTAAATTCTTCTGTTTG-3'

DIG1_infusionconstr3_R: 5'- ACTAGTGGATCTGATCAATAACAAGGAGGGAAGACCA-3'

The empty pFA6a-*KanMX4* vector (AB352) was used as the control plasmid for comparative growth and phenotype experiments.

MOLECULAR KARYOTYPING BY RAD-SEQ COVERAGE

PROCESSING RAW SEQUENCING DATA

For each lane of sequencing, the raw reads were separated into bins based on their 4-base strain-specific barcode (5' end of each read). Within each strain-specific bin the barcodes were removed and reads were then aligned to the S288C reference genome sequence (SGD R64-1-1_20110203) using BWA (v0.5.8)[43] with up to 6 mismatches allowed. The 5' start positions of all reads with a Phred-scaled mapping alignment quality of at least 20 were counted, resulting in a set of marker positions for each strain along with the number of reads aligning to each of those positions.

MARKER SELECTION

Next, a series of filtering steps were performed to select an optimal set of markers to represent the DNA copy number across the genome. First, markers with a median of 0 or 1 counts over all strains sequenced for this study (~400) were removed. We believe that the majority of these markers represent sequencing or DNA fragmentation errors. Second, the consistency of the markers was compared to the expected length of the *Mfe*I-*Mbo*I DNA fragment, based on the S288C reference sequence, i.e. each marker should represent a set of reads directed to the *Mfe*I end of a unique *Mfe*I-*Mbo*I restriction fragment.

For each marker within each strain, the proportion of all reads aligning to that marker was then calculated, as follows:

$$Marker\ proportion\ = \frac{Marker\ count}{Total\ read\ counts\ for\ strain}$$

The consistency of each marker was then expressed as the Coefficient of Variance (CV) of those proportions across all strains and plotted against the predicted marker fragment length (**Figure 4-2**).

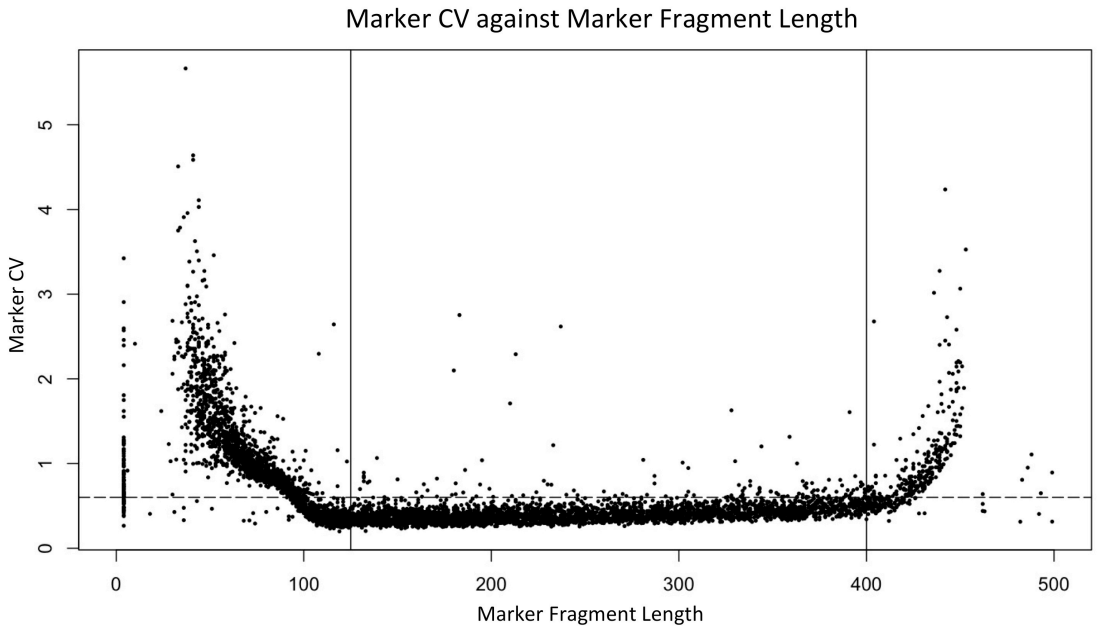


FIGURE 4-2 - MARKER PERFORMANCE BASED ON THE PREDICTED MARKER FRAGMENT LENGTH. EACH POINT REPRESENTS A MARKER. LINES ARE FILTERS THAT WERE APPLIED TO THE DATA, MARKERS WHOSE FRAGMENT LENGTHS AND CV FALL IN THE LOWER MIDDLE SECTION WERE USED FOR SUBSEQUENT ANALYSIS.

As can be seen from the figure, the CV of markers arising from fragments well within the expected gel-selected size range (~75-425 bp) is low, but rises rapidly outside this range. In addition, a small number of markers within the selected size range still have inconsistent behavior (high CV). We therefore decided to set 2 filters, 125bp <= Marker fragment length <= 400bp, and Marker CV <= 0.6, to select for markers that we believe should behave in a reliable manner.

$$Markers_{Length\ and\ C.V.filtered} =$$

$$Marker\ if\ (125bp \leq Marker\ Length \leq 400bp \cap C.V. \leq 0.6)$$

This reduces the markers available for the subsequent analysis by about 30%, but still leaves more than 3000 markers spanning the genome.

INDIVIDUAL STRAIN CALCULATIONS RELATIVE TO THE EUPLOID REFERENCE PANEL

To normalize marker-specific coverage effects, a reference euploid panel (R.E.P.) of 13 sequenced strains was selected. For each marker in each query strain the read coverage (proportion) was normalized using the mean read coverage (proportion) from this panel. Assessment of the copy number of each chromosome in each query strain was then carried out as follows. First, for each marker in the length and CV filtered set, the proportion of all reads aligning to the marker was calculated and then normalized by dividing by the mean marker proportion from the reference euploid panel:

$$Marker\ ratio = \frac{Marker\ proportion}{Mean\ marker\ proportion\ of\ R.E.P.}$$

As this ratio will vary slightly depending on ploidy (aneuploids versus euploids or other aneuploids), it was then normalized to set the median marker ratio to 1:

$$Marker\ ratio_{Query\ strain\ normalized} = \frac{Marker\ ratio}{Median\ marker\ ratio\ for\ query\ strain}$$

The resulting normalized marker ratios were then plotted on a $\log_2$ scale against the marker number, ordered according to the genomic position, using the R statistical software environment (http://www.r-project.org/). Each point represents a marker, and chromosomes are given alternating colors. To keep a constant scale, the y-axis was limited from the values of -1 to 2. The ratios of an extremely small number of markers lie outside this range, but do not affect the interpretation of the figures. These plots, in conjunction with flow cytometry analysis, were then used to assess the number of copies of each chromosome in each strain.

GROWTH ASSAYS IN LIQUID AND SOLID MEDIA

The growth behaviors of fluffy and smooth strains were compared in liquid versus solid media. Liquid growth for F45, YO785, YO1770/YO1771 (2 independent transformants of F45 containing the empty vector pFA6a-kanMX4), and YO1772/YO1773 (2 independent transformants of F45 containing the *DIG1* CEN plasmid pAB340) was measured using a Tecan Sunrise instrument with Magellan 6 software. Overnight cultures of strains were sonicated to separate large cell clumps and were counted using a hemocytometer. Cultures were seeded in treated 96-well plates at a density of 10^5 cells/ml in a well volume of 150µl. Plates were incubated with continuous shaking and OD_{600} readings were taken every 15 minutes until saturation. F45 was compared to the chromosome XVI disome (YO785) in YPD while the overexpressed *DIG1* smooth strains (YO1772/YO1773) were compared with the F45 plus empty

vector (YO1770/YO1771) in YPD + G418 media. Despite agitation YO1770 and YO1771 clumped at the edges of the wells, so a final OD_{600} measurement was taken at the end of the growth following resuspension. Multiple replicates of each strain (25-30 wells) were measured across 2 runs. Outlying wells were excluded from subsequent analysis. Additionally, liquid growth OD_{600} to cell count comparison was established on a per-strain basis at 1:4, 1:16, 1:64, and 1:256 dilutions. From the resulting growth curves (**Figure 4-3A, B**) we extracted the instantaneous growth rates (**Figure 4-4C, D**) by assuming exponential growth for all four strains for each time step.

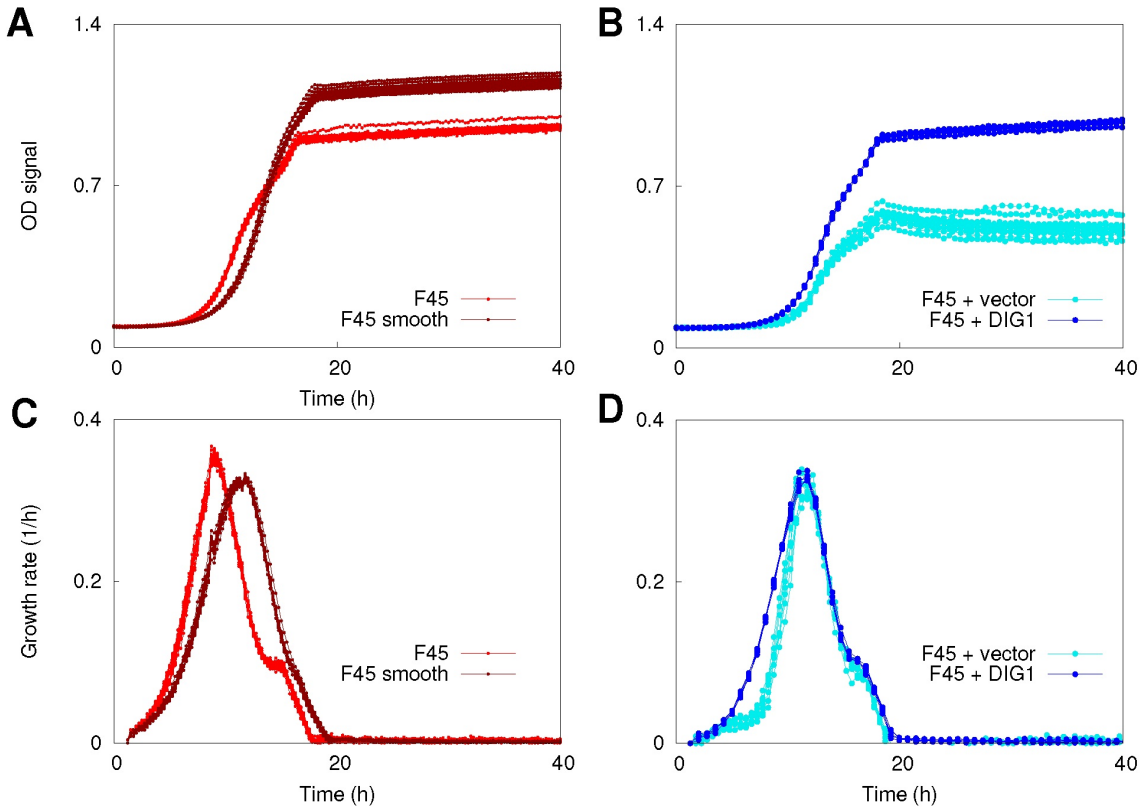


FIGURE 4-3 - GROWTH RATES OF STRAINS IN LIQUID MEDIA. OD (A, B) AND GROWTH RATES (C, D) ARE PLOTTED AGAINST TIME FOR F45 AND F45 SMOOTH/YO785 (A, C); F45 + VECTOR/YO1770 AND F45 + DIG1/YO1772 (C, D).

For the solid media growth time courses, single cells were spotted to solid media using a BD FACSAria II in a “checkerboard” layout of 48 cells and 12.7 mm spacing between cells. Five plates of each strain were prepared in this fashion. After spotting, one plate from each strain was placed face-up without lid, and a sheet of 0.25" thick clear acrylic was weighted down over these four plates to prevent contamination and desiccation. These plates were imaged every 15 minutes from day 1 through day 4 of growth. Five images were taken at each time point, and the five images were averaged to decrease noise. The F45 and YO785 plates were imaged side-by-side with one camera, and YO1770 and YO1772 plates were imaged side-by-side with a second camera. Images of each plate were segmented using ImageJ (http://rsb.info.nih.gov/ij/) to extract colony area from each image (**Figure 4-4**). The rate of growth in each replicate was calculated by dividing the difference in area between successive timepoints by the elapsed time. These rates were then smoothed using an 11-element moving average window, and the maximum rates for all replicates were calculated and averaged. To obtain cell counts per colony, representative colonies (20 per strain) were scraped from the plates and resuspended in 1ml of PBS at the end of day 4. OD_{600} of these cell suspensions were measured using a Tecan Sunrise at dilutions of 0, 1:4, 1:16, 1:64, and 1:256. Cells from two colonies per strain were also counted (at each dilution) using a hemocytometer to establish the relationship between OD_{600} and cell number for each strain and to obtain total cells per colony,

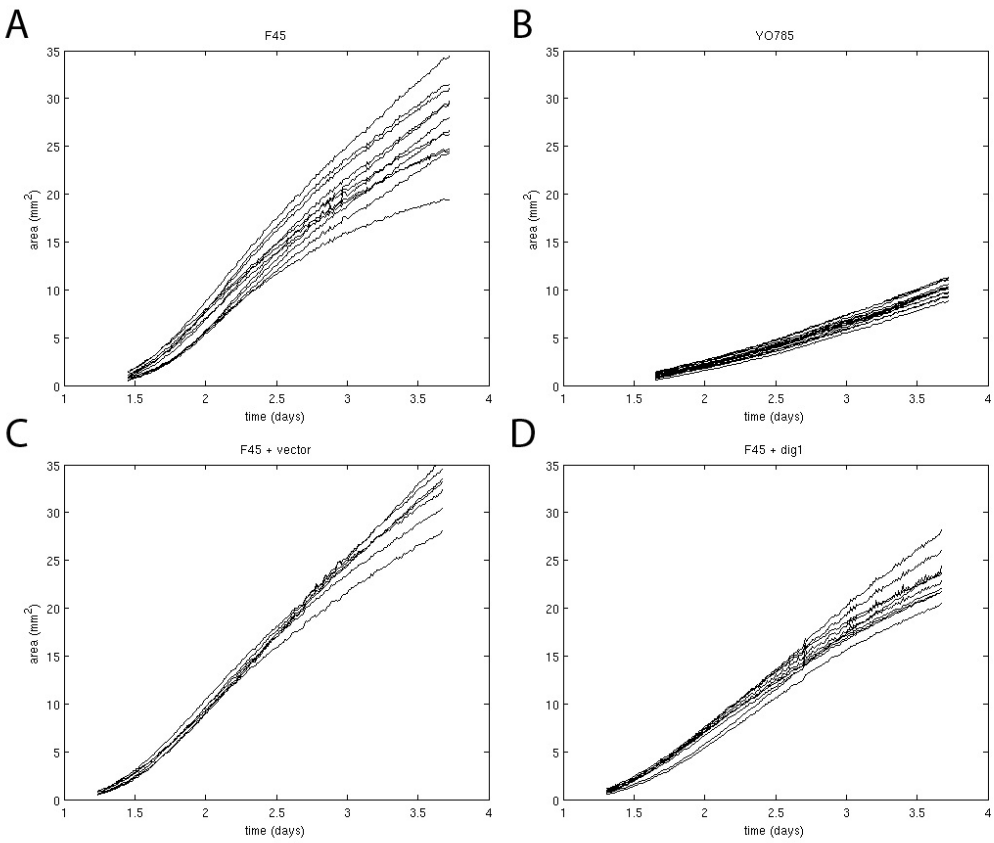


FIGURE 4-4 - GROWTH RATES OF STRAINS ON SOLID MEDIA. COLONY AREA (MM2) PLOTTED AGAINST TIME (IN DAYS) FOR F45 (A), F45 SMOOTH/YO785 (B), F45 + VECTOR/YO1770 (C), F45 + DIG1/YO1772 (D).

CURING PRIONS FROM STRAINS

To cure our strains of prions in two different ways, we grew our strains in liquid YPD media containing 1 mM Guanidine Hydrochloride (GuHCl) for 24 hours and on YPD + 4mM GuHCl agar plates for 3 days (30). Following GuHCl treatment, cells were streaked or plated out onto YPD agar plates for assessment of colony morphology (**Figure 4-5**).

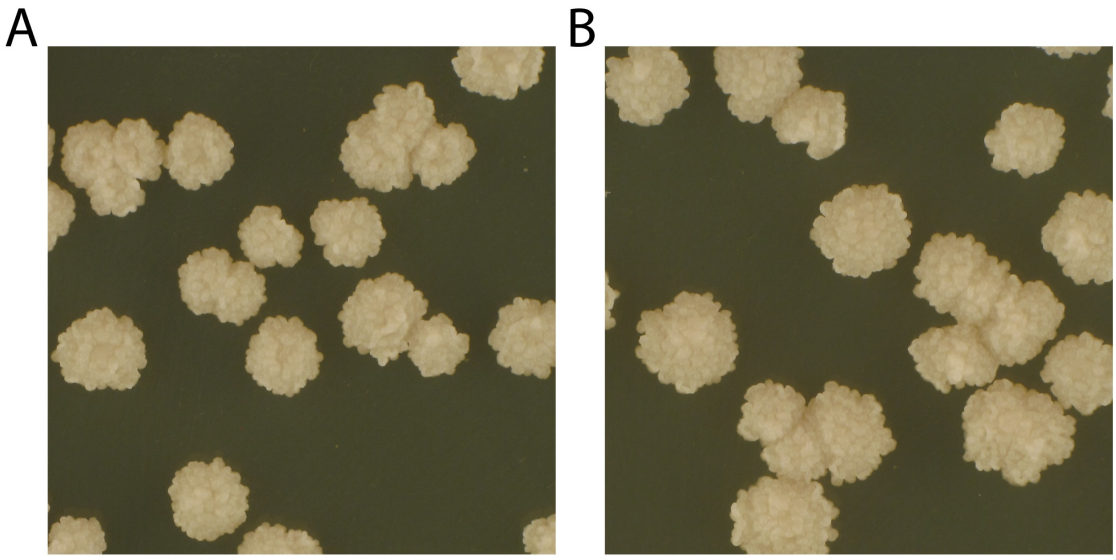


FIGURE 4-5 - PHENOTYPIC SWITCHING FOR OUR STRAIN IS NOT DUE TO PRIONS. A) F45 COLONIES THAT HAVE BEEN CURED OF PRIONS BY GROWTH IN YPD + 1MM GUANIDINE HYDROCHLORIDE FOR 24 HOURS. B) F45 COLONIES THAT HAVE NOT BEEN CURED OF PRIONS. IMAGES ARE TAKEN ON DAY 3 OF GROWTH. SINCE THE COLONY MORPHOLOGIES OF BOTH TREATED AND UNTREATED COLONIES ARE INDISTINGUISHABLE, PRIONS ARE UNLIKELY TO BE THE MECHANISM UNDERLYING THE SWITCH FOR OUR STRAIN.

4.4 RESULTS

BIDIRECTIONAL TOGGLING BETWEEN PHENOTYPIC STATES

While studying colony morphology in the haploid progeny of a cross between a Japanese sake strain and an Ethiopian white tecc strain (**Table 4-1**), we observed a high frequency (10^{-3}) of colonies that switched from a fluffy to a smooth morphology. The smooth state of these isolates was clonally heritable and stable through multiple cell divisions, as evidenced by the formation of smooth colonies from single cells isolated from 4-day-old smooth colonies (**Figure 4-6A**). However, when grown for longer periods of time (10-16 days), “blebs” appeared on the colony surface of otherwise smooth colonies (**Figure 4-6B**). We hypothesized that these patches of cells had adopted characteristics distinct from the neighboring cells, and indeed, a portion of single

cells isolated from these patches had regained the ability to form fluffy colonies with shapes indistinguishable from that of the original strain (**Figure 4-6C**). We therefore set out to determine the mechanism underlying this bidirectional switch.

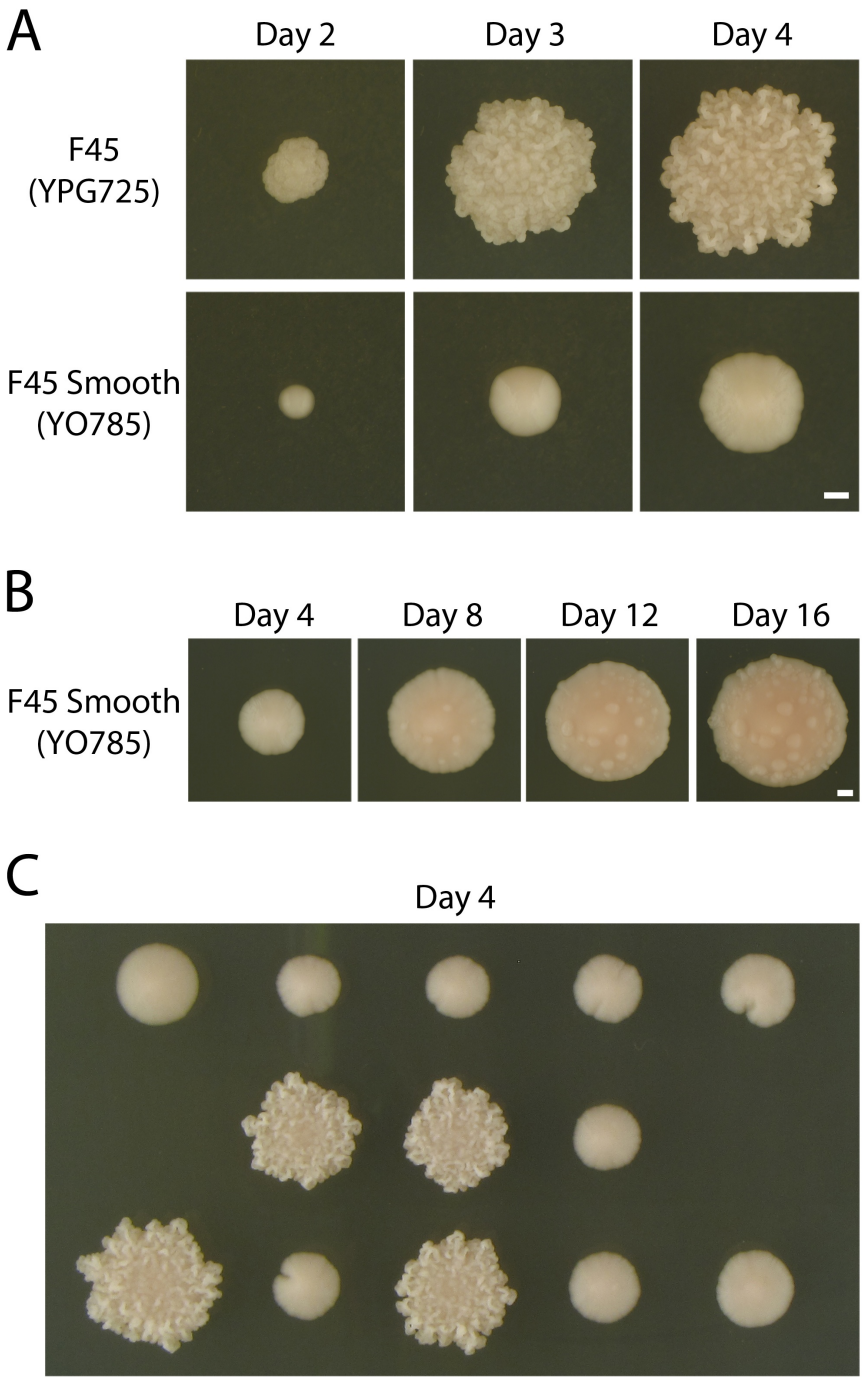


FIGURE 4-6 - TOGGLING BETWEEN THE FLUFFY AND SMOOTH STATES. A) MORPHOLOGY DEVELOPMENT OF A COLONY GROWING FROM A SINGLE CELL OF THE F45 FLUFFY STRAIN AND A SMOOTH VARIANT (F45 SMOOTH). B) DEVELOPMENT OF "BLEBS" ON THE SURFACE OF A F45 SMOOTH COLONY. C) SINGLE CELLS ISOLATED BY MICROMANIPULATION FROM THE "BLEBS"ON DAY 16 YIELDED ENTIRELY FLUFFY OR SMOOTH COLONIES. SCALE BARS ARE 1MM.

ANEUPLOIDY IS SUFFICIENT FOR THE PHENOTYPIC SWITCH

Because transition to a prion state is an epigenetic phenomenon that can affect colony morphology[29], we tested for the effect of prions by passaging the original F45 fluffy strain on guanidine hydrochloride, a method known to cure prions in *S. cerevisiae*[44], and then re-assayed colony morphology on our assay medium (YPD). The results (**Figure 4-5**) showed that the fluffy morphology of F45 was independent of the prion status, and therefore prions were unlikely to be the molecular mechanism underlying the switch in our strain background.

Because changes in DNA copy number can also occur reversibly and at relatively high frequencies, we next tested our strains for aneuploidy (chromosome numbers that deviate from multiples of the haploid chromosome count). Using a RAD-seq strategy that directs genome sequencing to specific restriction sites[35] thereby sequencing the same ~1% of each strain in a highly multiplexed fashion, we determined the karyotype of our strains based on ~3,000 markers across the genome. Together with DNA content estimates determined by flow cytometry (**Figure 4-3**), RAD-seq showed that while the original fluffy strain (F45) was euploid, transition to the smooth state (F45 Smooth) was accompanied by the gain of an extra copy of chromosome XVI (**Figure 4-7A**). Remarkably, the additional copy of chromosome XVI was lost when the strain transitioned back to the fluffy state (**Figure 4-7A**, F45 "2nd gen"), exhibiting a morphology indistinguishable from that of the original strain.

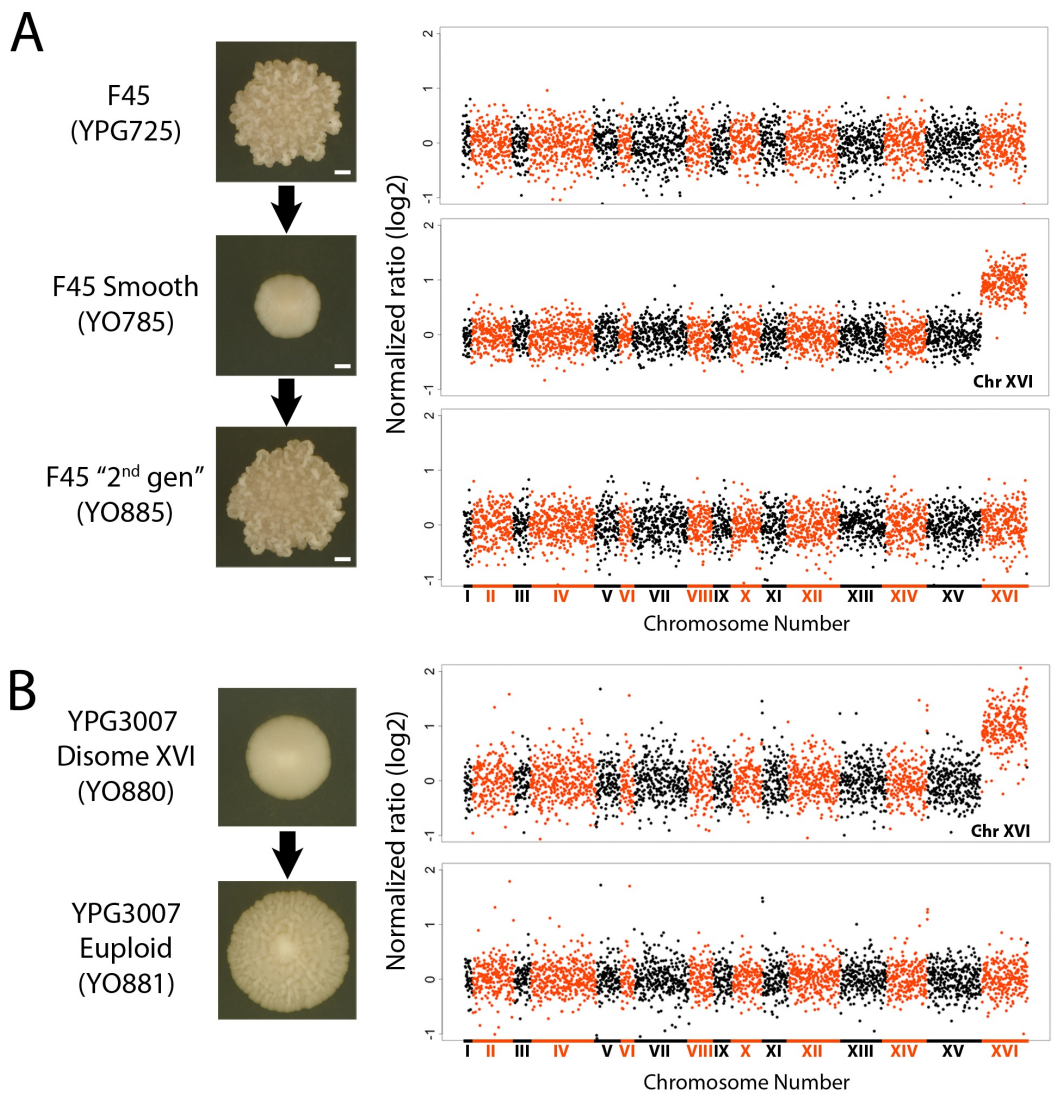


FIGURE 4-7 - CHANGE IN MORPHOLOGY AFFECTED BY CHROMOSOME XVI COPY NUMBER. A) PHENOTYPIC TOGGLING OF NATURALLY DERIVED STRAINS. PLOTS OF RAD-SEQ DATA SHOW THE PRESENCE OF AN EXTRA CHROMOSOME XVI IN F45 SMOOTH AND ITS SUBSEQUENT LOSS IN THE FLUFFY REVERTANT (F45 "2ND GEN"). CHROMOSOMES ARE ALTERNATIVELY COLORED BLACK AND RED. B) COLONY MORPHOLOGY IN YO880, A STRAIN UNRELATED TO F45. THE LOSS OF AN EXTRA CHROMOSOME XVI ALSO CORRELATES WITH THE GAIN IN COMPLEX COLONY MORPHOLOGY IN THIS STRAIN BACKGROUND.

To test whether this phenomenon was specific to F45, we chose a strain (YO880) from an independent genetic background that also harbored an extra copy of chromosome XVI. YO880 is

a haploid strain derived from a cross between strains originating from “Evolution Canyon” in Israel and an oak tree in North America. Similar to F45, the version of YO880 containing an extra copy of chromosome XVI was smooth (**Figure 4-7B**), but upon extended growth produced patches of fluffy colony producing cells. RAD-seq confirmed that these fluffy isolates had also lost the extra copy of chromosome XVI (Figure 4-7B). Thus, in two independent genetic backgrounds, smooth and fluffy colony morphologies were correlated with chromosome XVI copy number.

To determine whether chromosome XVI disomy was sufficient for the phenotypic switch, we engineered the original strain (F45) to enrich for chromosome XVI disomes, independent of their phenotype (**Figure 4-8**). This system[38] modifies the centromere of a target chromosome so that it can be temporarily inactivated (by exposure to galactose), thereby increasing the frequency of chromosome-specific nondisjunction[37]. Galactose-induced nondisjunction events are selected using a marker that can exist in one of two states, an intact *HIS3* or an intact *URA3*[38]. Only cells that have obtained two copies of the target chromosome can excise *HIS3* from one copy and maintain *URA3* on the other, allowing them to grow on the selective medium (SC –His –Ura). In our strain background, ~10^3 out of 10^7 galactose-induced cells produced His^+ Ura^+ colonies. Since colony morphology was difficult to score on the selective medium, we randomly chose 48 colonies and assayed their morphology on rich medium (YPD) (**Figure 4-9**). All of these colonies were smooth, and RAD-seq of a representative subset (n=22) confirmed the presence of an additional copy of chromosome XVI (**Figure 4-10**). To test whether restoration of the euploid state was sufficient to switch colonies back to the fluffy morphology, we used minimal medium containing 5-Fluoroorotic Acid (5-FOA) to select for the loss of the *URA3*-containing copy of chromosome XVI (**Figure 4-11**). In agreement with the original phenotypic selection, RAD-seq confirmed that all randomly selected colonies with adequate sequence coverage (n=12) were euploid and exhibited the fluffy colony morphology (**Figure 4-8** and **Figure 4-10**). Thus, the gain and loss of chromosome XVI is sufficient to reversibly toggle cells between the fluffy and smooth states.

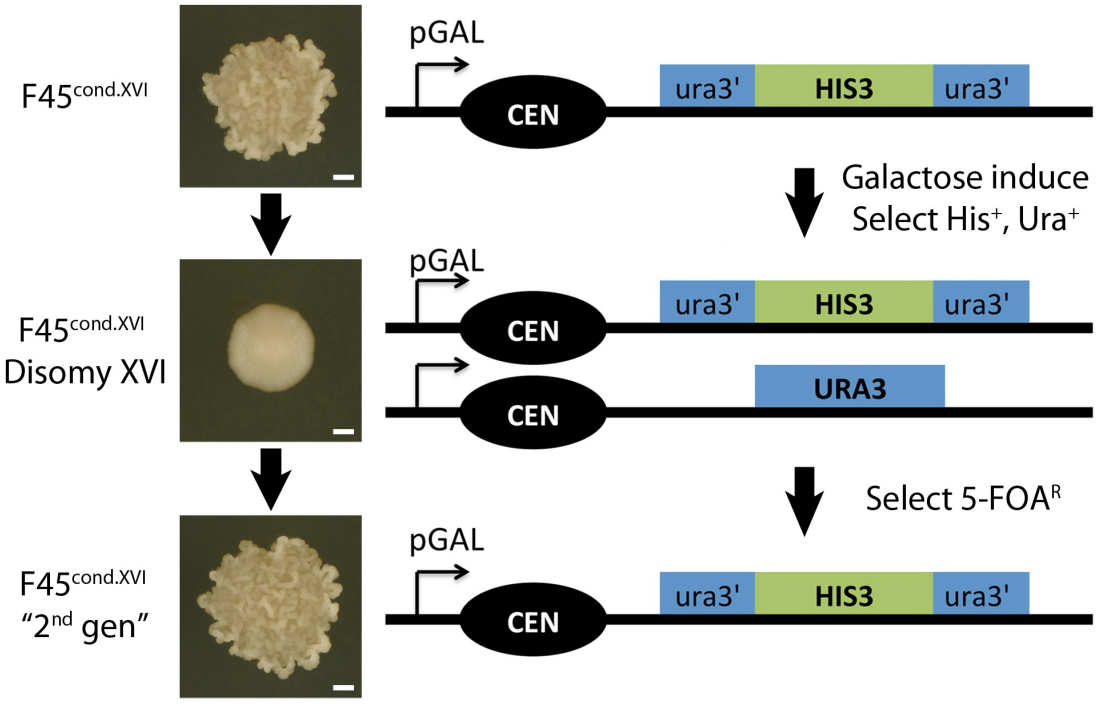


FIGURE 4-8 - PHENOTYPIC TOGGLING OF THE STRAIN CONTAINING A CONDITIONAL CENTROMERE ON CHROMOSOME XVI ($F45^{COND.XVI}$). THE PROTOCOL USED TO SELECT CHROMOSOME GAIN AND LOSS IS DEPICTED (RIGHT). SCALE BAR IS 1 MM. IMAGES WERE TAKEN ON DAY 4 OF COLONY GROWTH.

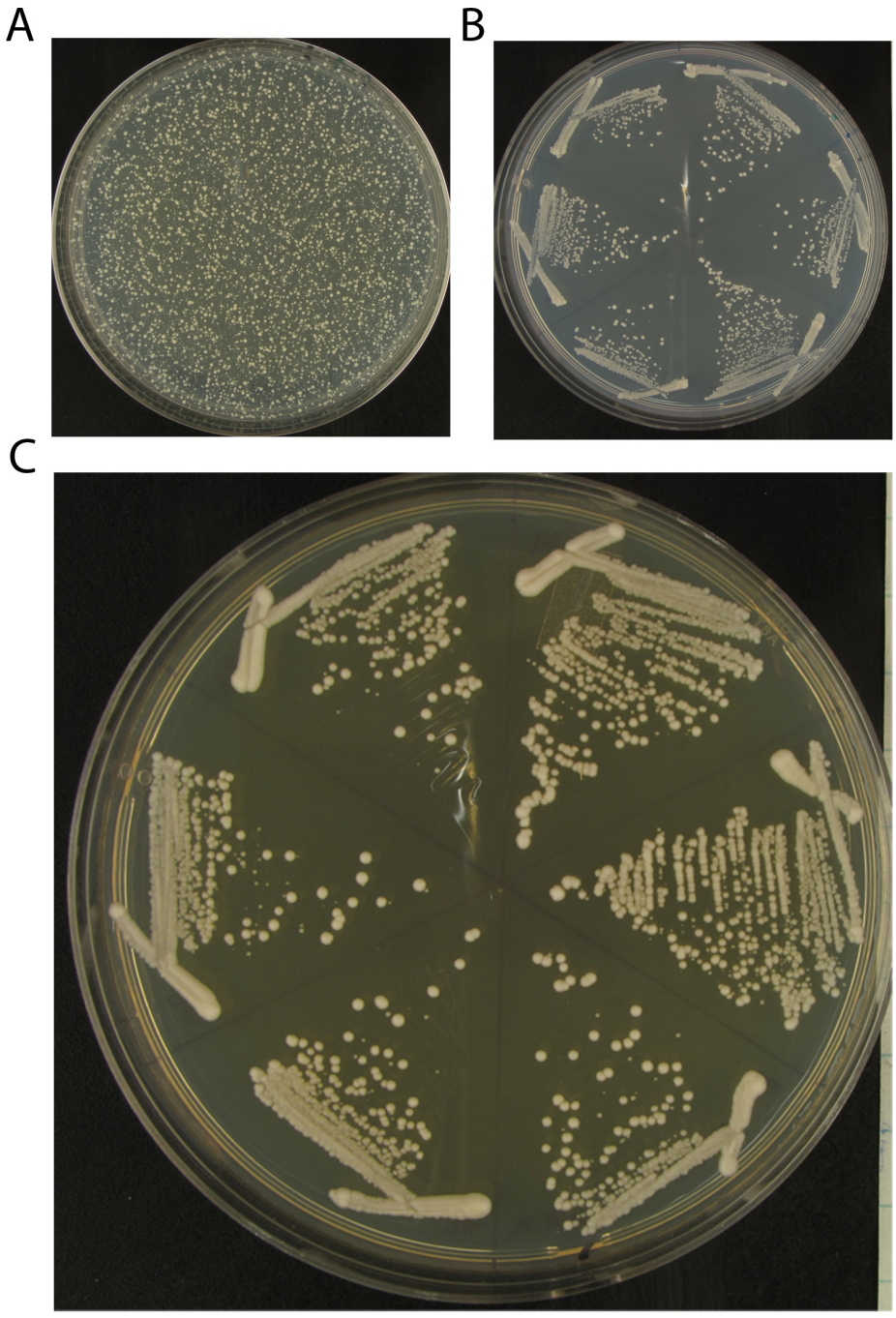


FIGURE 4-9 - SELECTION OF DISOMIC STRAINS. A) REPRESENTATIVE IMAGE OF INITIAL SELECTION PLATES (SC –HIS –URA). AS COLONIES WERE TOO DENSE, THEY WERE STREAKED FOR SINGLE COLONIES B) REPRESENTATIVE SINGLE COLONY STREAKS OF COLONIES (ON REPEATED SELECTION PLATES (SC –HIS –URA)) FROM INITIAL SELECTION PLATE. C) REPRESENTATIVE SINGLE COLONY STREAKS OF SINGLE COLONIES (ON YPD PLATES) FROM REPEATED SELECTION PLATES, SHOWING THE SMOOTH COLONY MORPHOLOGY OF THE COLONIES. ALL IMAGES ARE TAKEN ON DAY 2 OF GROWTH.

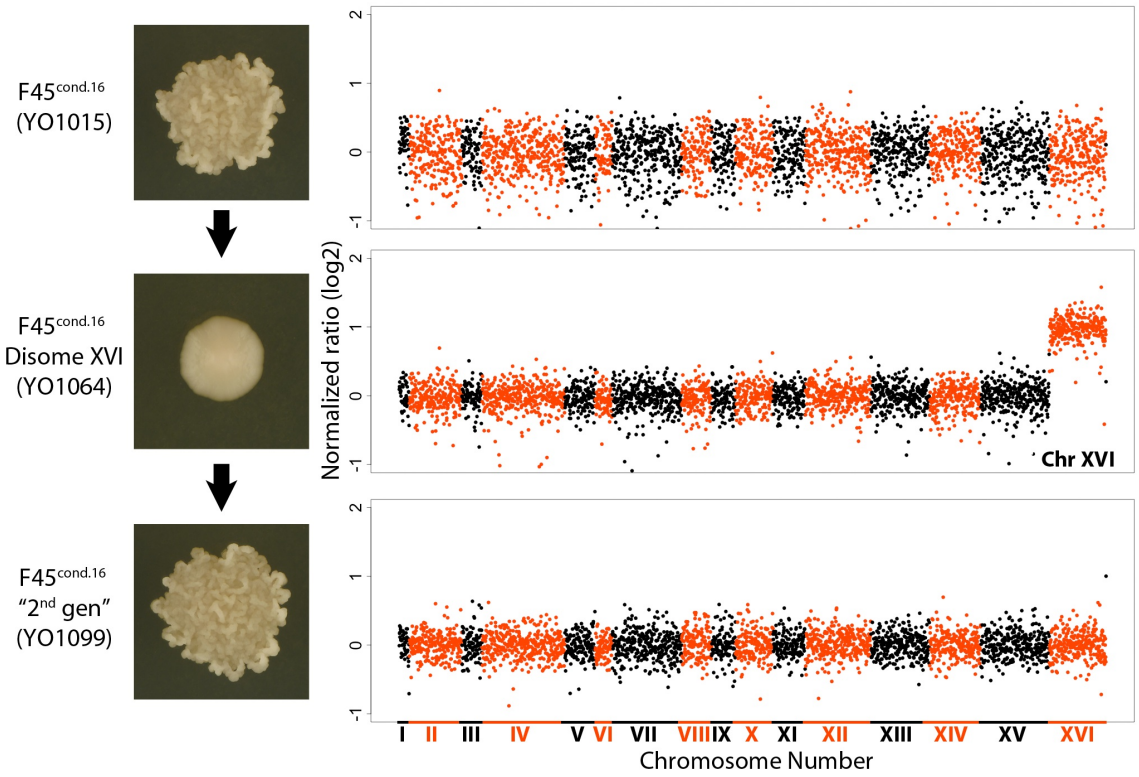


FIGURE 4-10 - CHANGE IN MORPHOLOGY ASSOCIATED WITH THE GAIN AND LOSS OF CHROMOSOME XVI (CONTAINING THE CONDITIONAL CENTROMERE). IMAGES ARE TYPICAL OF OBSERVED MORPHOLOGY, AND ARE TAKEN AFTER 4 DAYS OF GROWTH OF A SINGLE CELL. ACCOMPANYING PLOTS SHOW THE GAIN AND LOSS OF CHROMOSOME XVI AS IDENTIFIED BY RAD-SEQ AND ARE REPRESENTATIVE OF BIOLOGICAL REPLICATES.

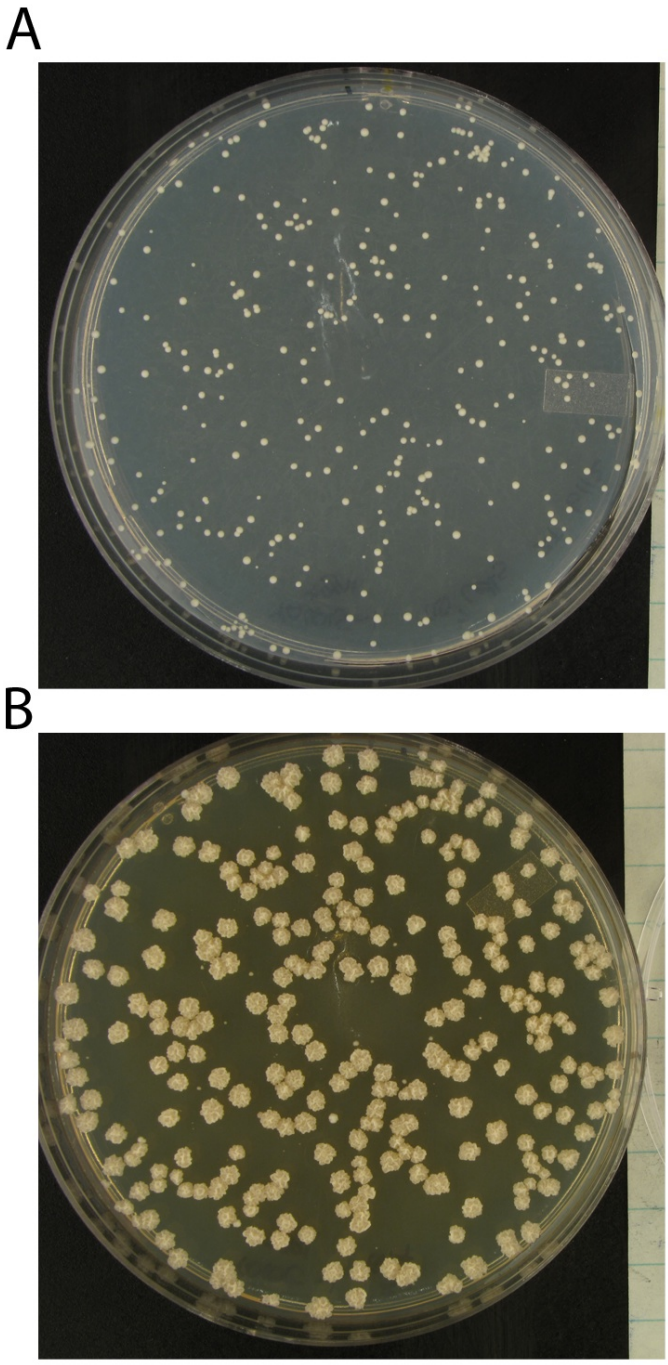


FIGURE 4-11 - SELECTION OF EUPLOID STRAINS. A) REPRESENTATIVE IMAGE OF INITIAL SELECTION PLATES (5-FOA). IN OUR HANDS, ~10^2 OUT OF 10^7 CELLS PRODUCED 5-FOA RESISTANT COLONIES. AS COLONY MORPHOLOGY IS DIFFICULT TO ASSESS ON 5-FOA PLATES, COLONIES WERE REPLICA-PRINTED ONTO YPD PLATES. B) REPLICA-PRINTED YPD PLATES, SHOWING THE FLUFFY COLONY MORPHOLOGY OF THE COLONIES. IMAGES WERE TAKEN ON DAY 2 OF GROWTH.

MULTIPLE ROUTES TO THE SAME PHENOTYPE

While analyzing additional fluffy to smooth isolates of F45, we were surprised to find that, in addition to chromosome XVI, several other aneuploidies altered the colony forming state of the strain. Disomies of chromosomes III, X, and XV displayed smooth colony morphologies, and chromosome V disomy displayed an intermediate phenotype (**Figure 4-12**). As with chromosome XVI, the subsequent reversion of all four disomic strains back to the original fluffy morphology was accompanied by restoration of the euploid karyotype (**Figure 4-13**). Thus, the ability of different disomies to trigger the bidirectional toggle, suggests that there are multiple, reversible routes to the same phenotypic state. We note that in the screen for smooth variants, we obtained 14 smooth isolates from a combination of the original F45 strain and some fluffy revertants (F45 "2nd gen") from F45 Smooth. 7 of the 14 were disomic strains, while the remaining strains show no discernible change in karyotype. These non-disomic smooth variants may have resulted from mutations, and were not followed up further. This suggests that additional mechanisms may exist.

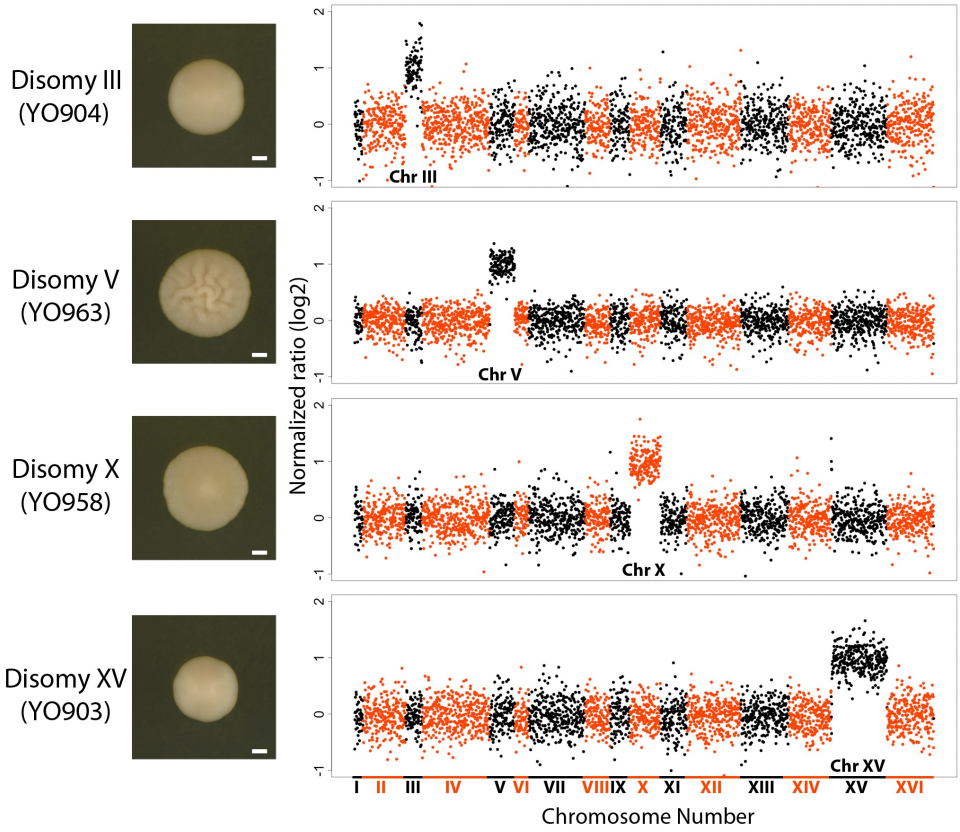


FIGURE 4-12 - MULTIPLE DISOMIES INDUCE A SIMILAR PHENOTYPE. COLONY MORPHOLOGY OF THE DIFFERENT DISOMIC YEAST STRAINS OBTAINED BY PHENOTYPIC SCREENING. PLOTS SHOW THE PRESENCE OF THE EXTRA CHROMOSOMES IDENTIFIED BY RAD-SEQ. SCALE BAR IS 1 MM. IMAGES WERE TAKEN ON DAY 4 OF COLONY GROWTH.

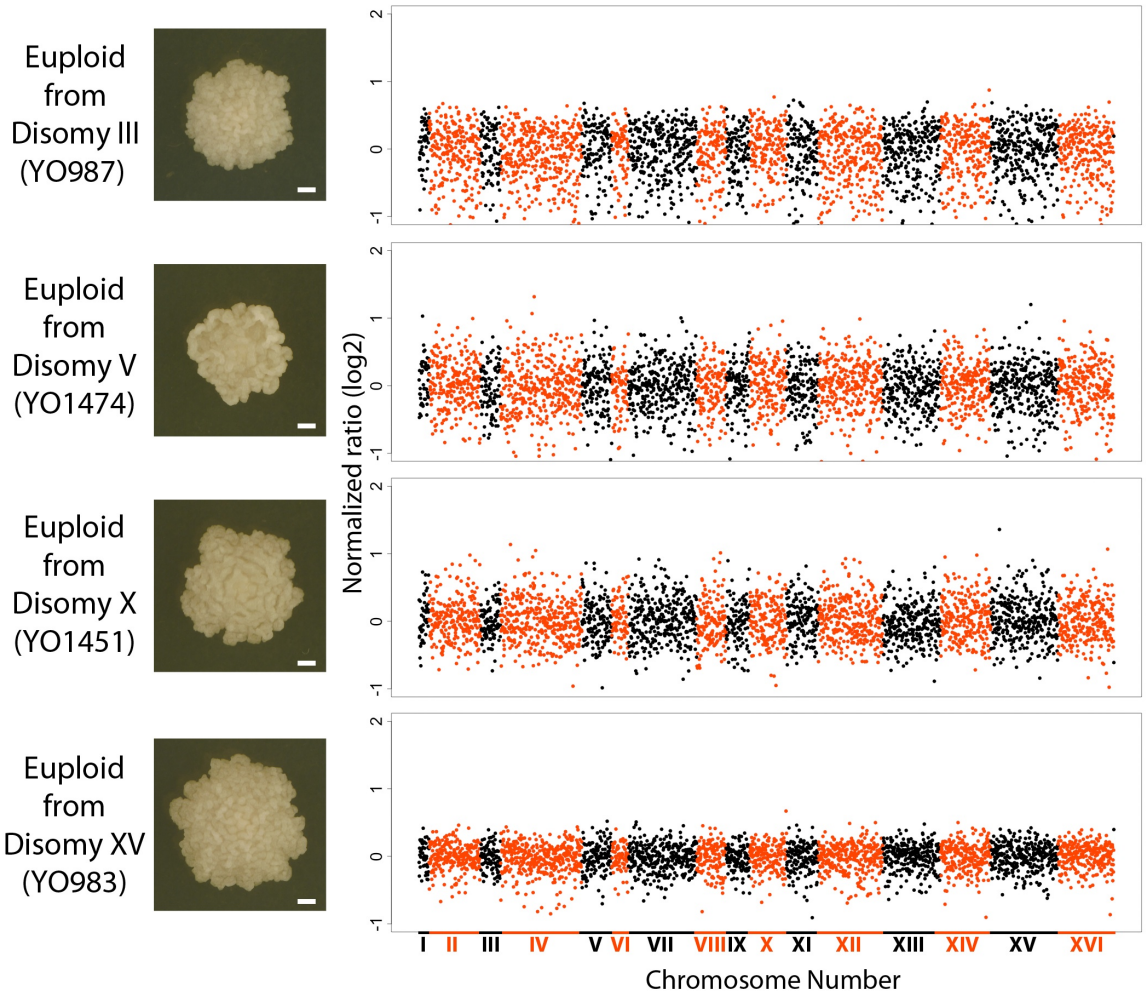


FIGURE 4-13 - RESTORATION OF EUPLOID KARYOTYPE RETURNS COLONIES TO ORIGINAL MORPHOLOGY. IMAGES SHOW REPRESENTATIVE COLONY MORPHOLOGY, AND ARE TAKEN AFTER 3 DAYS OF GROWTH OF A SINGLE CELL. PLOTS SHOW RESTORATION OF EUPLOID KARYOTYPE AS IDENTIFIED BY RAD-SEQ. SCALE BAR IS 1MM.

SWITCH MODULATED BY GENE-SPECIFIC COPY NUMBER VARIATION

Our results are consistent with either of two models. First, because numerous biological processes are required to form a fluffy colony[26-28], these five chromosomes could each contain dosage sensitive genes necessary for the function or regulation of key processes. Alternatively, the fluffy state could be sensitive to some more general aspect of aneuploidy, such as the DNA content of the cell. To test whether any chromosomal aneuploidy could confer the smooth

morphology, we introduced conditional centromere constructs onto the eleven remaining chromosomes (I, II, IV, VI, VII, VIII, IX, XI, XII, XIII and XIV) and selected for each disomy (**Figure 4-14** and **Table 4-2**). Of the eleven disomies selected, four were either too sick or unstable to assay (chromosome II, VI, XI, and XIII) and two could only be stably recovered in conjunction with additional aneuploidies (disome IV paired with VII and disome XII paired with XV). This last set of strains containing multiple disomies suggested that genetic interactions might play a role in the formation of colony morphology. For example, while strains disomic for chromosome III were smooth, and strains disomic for chromosome IX were fluffy, strains disomic for both III and IX were smooth (**Figure 4-15**). However, of the five chromosomes that could be stably recovered as individual disomies, only one (chromosome VII) yielded smooth colonies through day 4 of growth, while the remaining four disomies (chromosome I, VIII, IX and XIV) were all fluffy at the same time point (**Figure 4-15**). The fact that several disomies, spanning the continuum of chromosome sizes, failed to trigger the smooth transition strongly supports the hypothesis that the switch is due to specific gene effects rather than general DNA content changes.

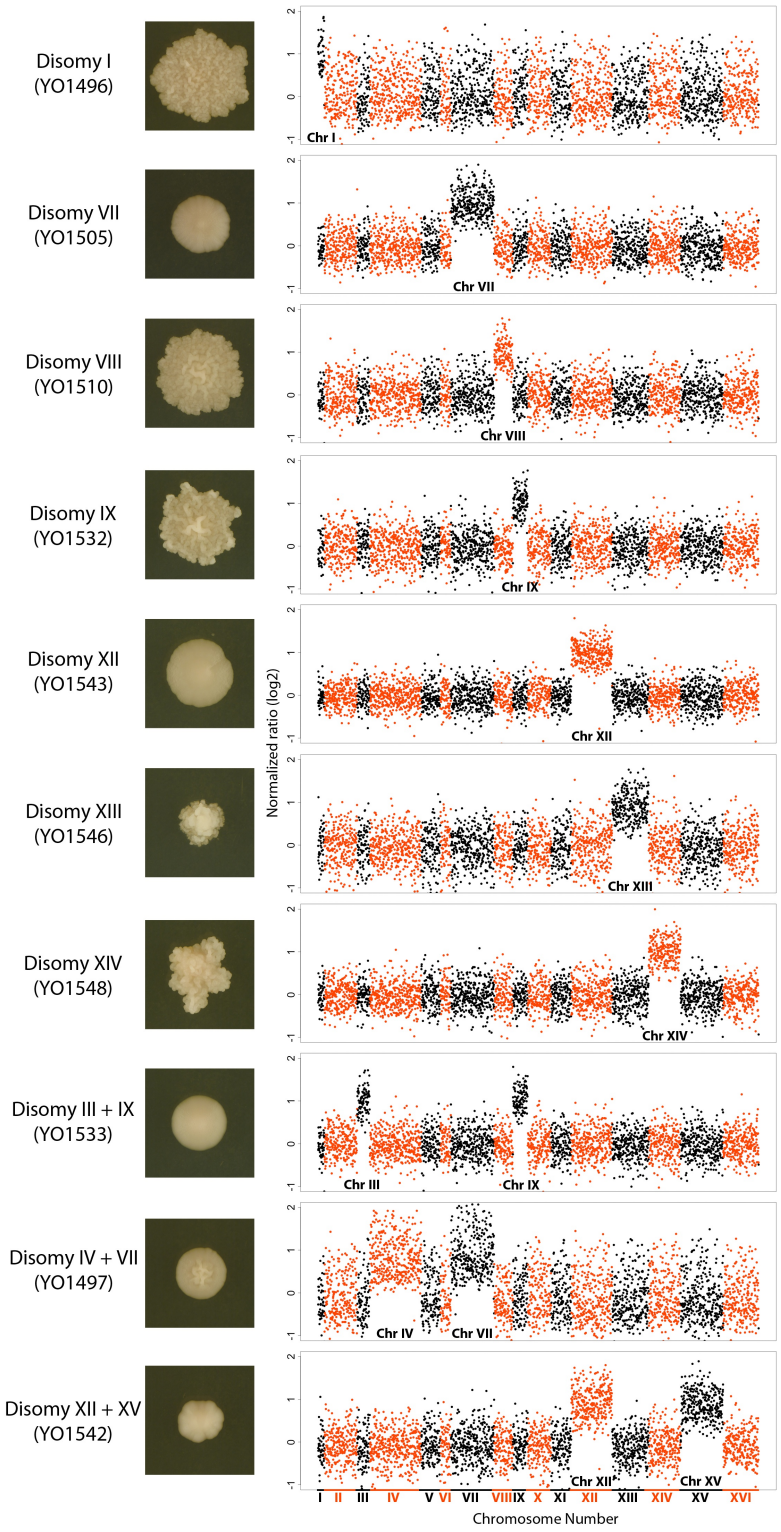


FIGURE 4-14 - KARYOTYPES OF ENGINEERED ANEUPLOID STRAINS. IMAGES SHOW REPRESENTATIVE COLONY MORPHOLOGY, PLOTS SHOW THE ADDITIONAL CHROMOSOMES AS IDENTIFIED BY RAD-SEQ. IMAGES WERE TAKEN ON DAY 4 OF COLONY GROWTH.

TABLE 4-2 - INITIAL KARYOTYPE AND MORPHOLOGY FOR STRAINS GENERATED UTILIZING THE CONDITIONAL CENTROMERE CONSTRUCT AND SELECTED FOR INITIAL RAD-SEQ ANALYSIS. ISOLATES FOR WHICH THE KARYOTYPE COULD NOT BE CALLED WERE OMITTED FROM THIS TABLE.

Chromosome harboring the conditional centromere	Karyotype following induction (disomy unless otherwise specified)	Proportion of strains with the karyotype	Morphology post-induction
I	I	14/18	Fluffy
	I trisomy	4/18	Fluffy
IV	IV+I	1/9	Fluffy
	IV+VII	8/9	Intermediate
VII	VII	12/12	Smooth
VIII	VIII	10/13	Fluffy
	VIII + XV	2/13	Fluffy
IX	IX	18/26	Fluffy
	IX + III	8/26	Smooth
XII	XII +XV	11/11	Smooth
XIII	XIII	5/11*	Fluffy
XIV	XIV	12/12	Fluffy

*Because the post-induction isolates exhibited a range of morphologies on the selective media, several isolates of each class, fluffy and smooth, were selected for karyotyping. All fluffy isolates were confirmed chromosome XIII disomes. The smooth isolates were suspected self-diploids which is consistent with what others have seen using this conditional centromere method[38].

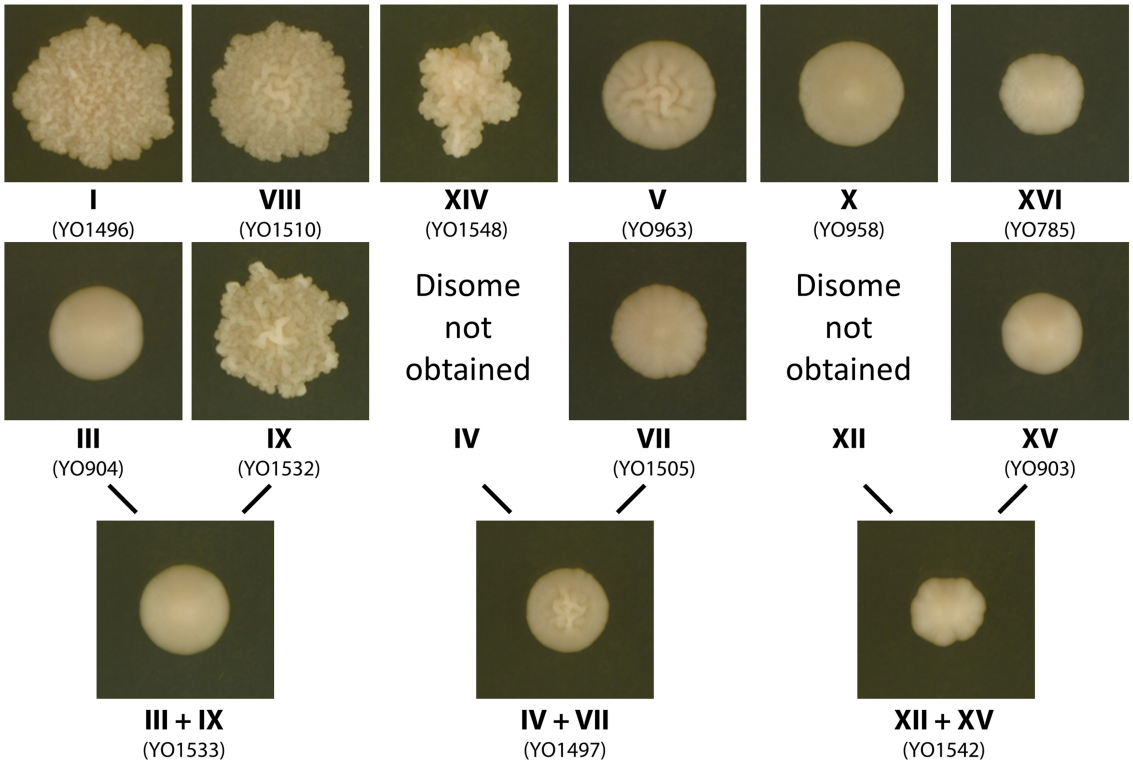


FIGURE 4-15 - COLONY MORPHOLOGY FOR DISOMIC STRAINS. MORPHOLOGIES OF REPRESENTATIVE EXAMPLES OF EACH KARYOTYPE ARE SHOWN ON YPD AGAR. ALL STRAINS IN THIS FIGURE WERE RE-KARYOTYPED FOLLOWING MICROMANIPULATION. ONLY STABLE, REPRODUCIBLE KARYOTYPES ARE DEPICTED. SEE SI OF REFERENCED PAPER FOR DETAILS ON SICK OR UNSTABLE DISOMIES (II, VI, XI AND XIII). IMAGES WERE TAKEN ON DAY 4 OF COLONY GROWTH.

To test whether this phenotypic toggle is in fact modulated by the copy number variation of specific genes, we transformed the original F45 (fluffy) strain with a set of low copy plasmids containing portions of chromosome XVI[45]. This screen identified a plasmid containing seven full-length genes that was able to confer the smooth state (**Figure 4-16A**). We subsequently determined that this effect was due to the increased copy number of a single gene *DIG1* (**Figure 4-16B**). *DIG1* encodes a known repressor of the MAPK pathway[46], and our results are consistent with a model in which overexpression of *DIG1* leads to the down-regulation of the fluffy trait[26]. However, the intermediate phenotype obtained by restoring *DIG1* copy number in the context of

the XVI disomy (**Figure 4-16B**) suggests that additional genes on chromosome XVI may contribute to the phenotypic switch.

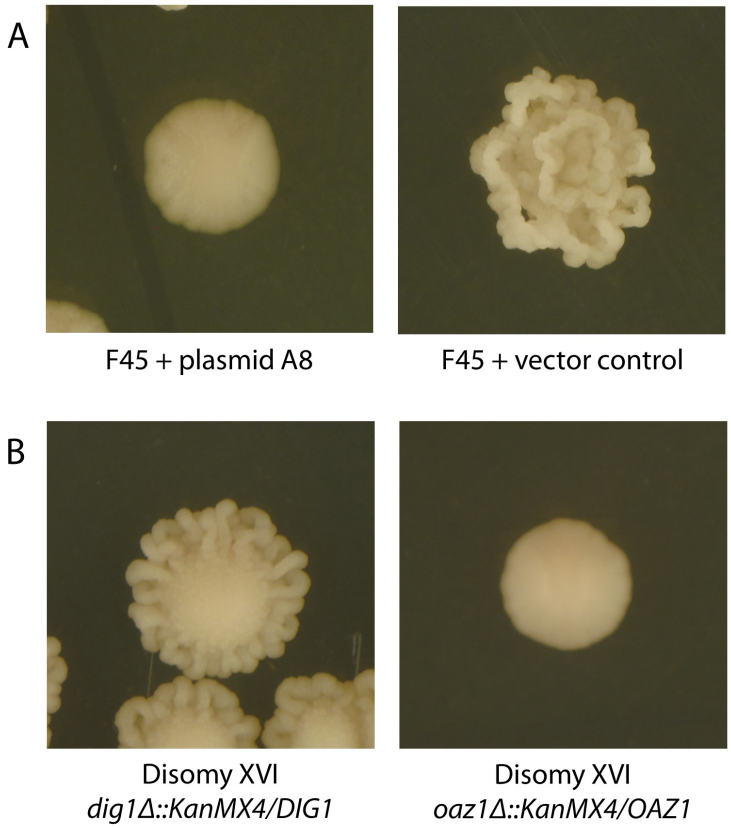


FIGURE 4-16 - *DIG1* COPY NUMBER INFLUENCES COLONY MORPHOLOGY. A) PLASMID PGP648 CONTAINING A 10.8 KB REGION OF CHROMOSOME XVI DNA (PLASMID A8) INDUCES A SWITCH TO THE SMOOTH PHENOTYPE IN F45. PLASMID A8 CONTAINS 7 COMPLETE ORFS (*OAZ1, ARL3, MNN9, DIG1, CAM1, SGF11, ELC1*) AND 2 PARTIAL ORFS (KTR6, VPS16). B) PARTIAL RESTORATION OF THE FLUFFY PHENOTYPE IS SEEN ONLY IN THE *DIG1Δ::KANMX4/DIG1* DISOMIC XVI STRAIN. ALL OTHER KNOCKOUTS REMAIN SMOOTH, INDICATING THAT *DIG1* IS RESPONSIBLE FOR THE CHANGE IN MORPHOLOGY INDUCED BY PLASMID A8.

GROWTH ADVANTAGES OF THE TWO STATES

Phenotypic switching suggests that the states may exhibit distinct fitness landscapes under different environmental conditions. Because the organization of cells into the fluffy colony

structure is a solid media phenomenon and because previous work has shown that fluffy colonies contain more cells than smooth colonies[47], we chose to compare the growth characteristics of smooth and fluffy colony forming strains in both liquid and solid media. We first compared the original fluffy euploid (F45) and the smooth XVI disome (F45 Smooth). On solid medium, the fluffy strain had a substantial growth advantage, as measured by the maximal growth rate achieved and number of cells produced by day 4 of growth (**Figure 4-17C**). While these results are consistent with previous comparisons of cell number in fluffy and smooth colonies[47], they could also merely reflect decreased growth rates resulting from aneuploidy[48]. Interestingly, in liquid medium the smooth, aneuploid strain showed only a modest decrease in growth rate and even a statistically significant increase in cell yield at saturation (**Figure 4-17E**). These results suggest that the smooth state has advantages for growth in liquid media even in the presence of the presumed growth disadvantage conferred by aneuploidy.

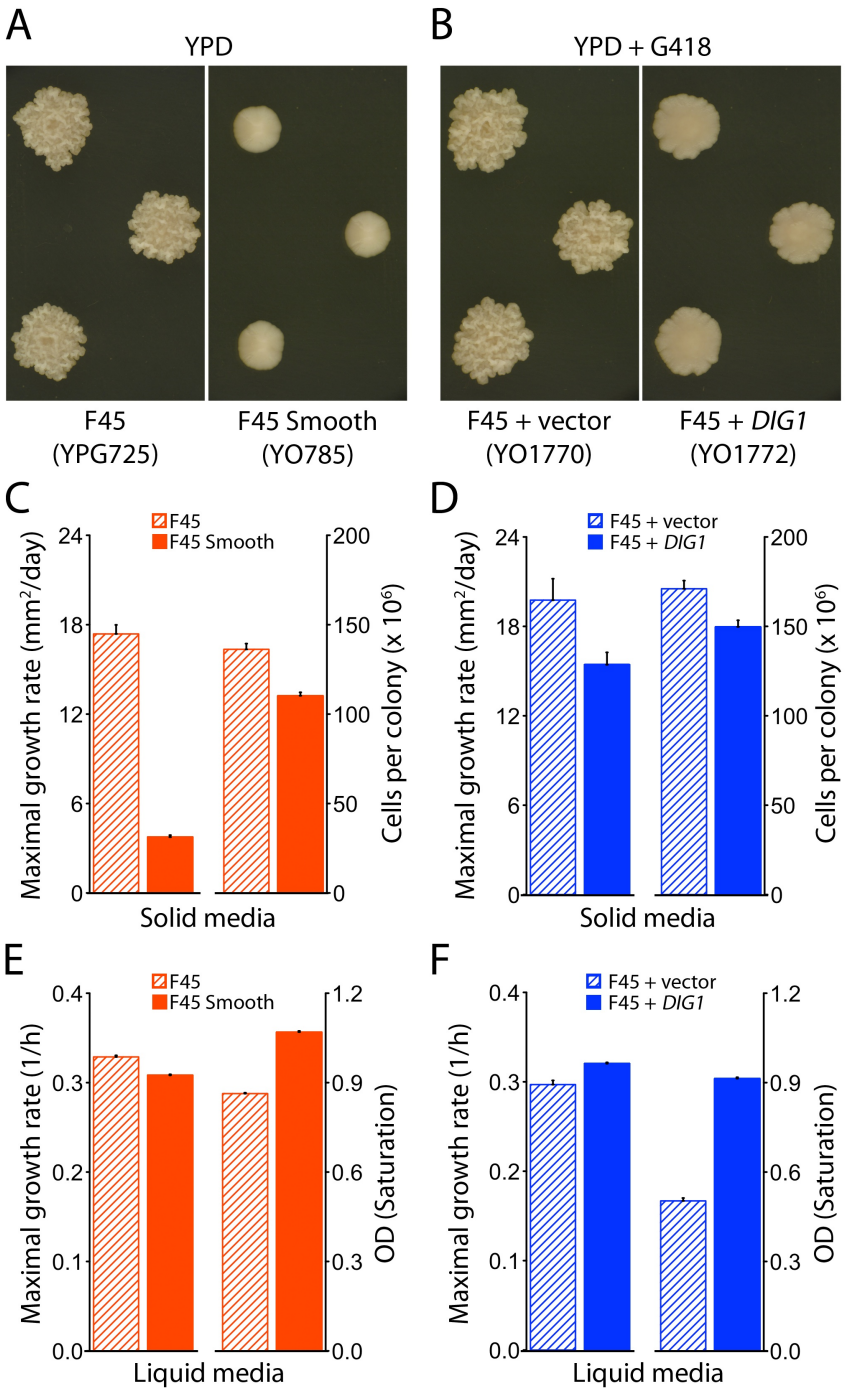


FIGURE 4-17 - GROWTH CHARACTERISTICS OF FLUFFY AND SMOOTH STRAINS. REPRESENTATIVE COLONIES OF A) F45 AND F45 SMOOTH GROWING ON YPD AGAR, B) F45 + VECTOR AND F45 + DIG1 GROWING ON YPD + G418 AGAR AS PLATED IN A "CHECKERBOARD" PATTERN. IMAGES WERE TAKEN ON DAY 4 OF COLONY GROWTH. C) MAXIMAL GROWTH RATE ($P < 10^{-4}$; N= 12 OR 14 COLONIES; WILCOXON RANK SUM TEST) AND CELLS PER COLONY ($P < 10^{-7}$; N= 20 COLONIES; STUDENT'S T-TEST) OF F45 AND F45

SMOOTH GROWN ON YPD AGAR. ERROR BARS ARE SEM. D) MAXIMAL GROWTH RATE (P = 0.025; N=7 OR 10 COLONIES; WILCOXON RANK SUM TEST) AND CELLS PER COLONY ($P < 10^{-3}$; N= 20 COLONIES; STUDENT'S T-TEST) OF F45 + VECTOR AND F45 + DIG1 GROWN ON YPD + G418 AGAR. ERROR BARS ARE SEM. E) MAXIMAL GROWTH RATES ($P < 10^{-5}$; N= 28 OR 24 WELLS; STUDENT'S T-TEST) AND SATURATION OD ($P < 10^{-5}$; N= 28 OR 24 WELLS; STUDENT'S T-TEST) OF F45 AND F45 SMOOTH GROWN IN LIQUID YPD. ERROR BARS ARE SEM AND ARE <1% OF THE VALUES OBTAINED. F) MAXIMAL GROWTH RATES ($P < 10^{-5}$; N= 14 OR 27 WELLS; STUDENT'S T-TEST) AND SATURATION OD ($P < 10^{-5}$; N= 14 OR 27 WELLS; STUDENT'S T-TEST) OF F45 + VECTOR AND F45 + DIG1 GROWN IN LIQUID YPD + G418. ERROR BARS ARE SEM AND ARE <5% OF THE VALUES OBTAINED.

Because the strain overexpressing *DIG1* exhibits a smooth colony morphology (**Figure 4-17B**) without the burden of an additional chromosome, the smooth state is genetically separable from the aneuploid state. We compared the growth characteristics of F45 transformed with either a low copy plasmid containing *DIG1* or an empty vector. In these experiments the fluffy (vector alone) strain also exhibited a growth advantage over the smooth (*DIG1*) strain on solid media (**Figure 4-17D**), while in liquid medium the *DIG1*-mediated smooth strain exhibited both a higher maximal growth rate and cell yield (**Figure 4-17F**). Thus, when the cost of aneuploidy is removed, the smooth state shows clear growth advantages in liquid.

4.5 CONCLUSIONS

The phenotypic changes brought about by aneuploidy can arise from a general increase in DNA content, dosage changes in specific gene products, or a combination of both factors. In *S. cerevisiae* some traits, such as decreased growth rates and G1 cell cycle delays, appear to be general effects of aneuploidy[48], while others, such as altered drug resistance and protein abundance, have been linked to the duplication of specific chromosomes or genes[49]. Our results add a complex, multicellular phenotype to the set of traits known to be regulated by aneuploidy. Cells in fluffy colonies organize to form intricate systems of hollow channels that increase the distances between cells, allowing for more rapid colonization of surfaces and greater acquisition

of resources[21]. At the same time, they perform metabolically intensive processes, such as the production of extracellular matrix, that while aiding the protection of the colony, require significant nutrient and energy expenditure[21]. Switching to the smooth state when such protection is not needed has been proposed as an energy conservation strategy[50]. Our results suggest that the copy number imbalance of specific chromosomes is one means by which cells can affect this state change, presumably by altering the dosage of a subset of genes required for the regulation or execution of these biological processes.

The surprisingly large number of chromosomes able to modulate the phenotypic toggle between fluffy and smooth may be proportional to the relatively large number of pathways required for the fluffy morphology[26-28] or may indicate that the trait is particularly sensitive to changes in gene dosage. In either case, the fact that aneuploidy provides cells multiple routes to the same phenotypic endpoint has fundamental implications for understanding how organisms respond to environmental perturbations. Because aneuploidy is strongly correlated with drug resistance in pathogenic microorganisms[51-54], rampant in cancer[55], and associated with high levels of cancer relapse after drug treatment[56], aneuploidy itself could be an important target for the development of novel antifungal and cancer therapeutics[55]. The complexity of the problem is compounded by the fact that the genome sequence (genetic background) of an individual has the potential to affect not only the traits displayed by aneuploid cells, but also how well specific aneuploidies are tolerated. Differences in the ability to recover specific disomes between our strain and one used in a previous study [48] may be one such example.

Phenotypic switching allows organisms to access multiple states with different growth or survival advantages. Here, we examine a system in which *S. cerevisiae* exists in two distinct multicellular states (fluffy and smooth) and is at times able to form heterogeneous populations containing both (**Figure 4-6**). Because the molecular mechanism underlying the fluffy-smooth switch is the mitotic gain and loss of intact chromosomes, it is likely the result of a stochastic process, although we have not directly tested this hypothesis. Furthermore, the fact that these two states exhibit different relative fitness profiles under different environmental conditions (solid and liquid media),

suggests that the organism may be leveraging the stochasticity of chromosome missegregation as a bet hedging strategy. Although we identified differences in growth, the two states could easily exhibit different advantages with respect to other traits, such as longevity, resistance to desiccation, and ease of dispersal, which may prove equally or even more beneficial to the organism in the natural environment. The importance of maintaining two states and the ability of the cells to switch between them is further supported by the recent identification a prion-based switch that can modulate colony morphology changes in other strains of *S. cerevisiae*[29].

The random gain and loss of complete chromosomes has many favorable properties as a mechanism for phenotypic switching, both in the context of a preemptive bet-hedging strategy as well as a rapid means for coping with unfamiliar environmental stresses[57]. While gene expression patterns can modulate phenotypic states[11], aneuploidy provides a more direct mode of inheritance. Additionally, because regulatory networks involve complex sets of molecular interactions that must evolve over time, aneuploidy might be a useful bet-hedging strategy for responding to environmental perturbations that are not frequently encountered. As aneuploidy perturbs hundreds of genes simultaneously, it has the potential to affect larger numbers of genes and processes per event relative to a spontaneous mutation. Importantly, unlike a spontaneous mutation, aneuploidy and any associated growth defects can also revert at a high frequency. This reversibility makes aneuploidy a particularly useful mechanism in conditions that are encountered infrequently and are likely to change. The inherent properties of aneuploidy thus first allow organisms to rapidly and extensively survey the fitness landscape imposed by new environmental conditions. Aneuploidy then allows organisms the flexibility of reverting to their previous state if conditions return, or stability as they fine-tune their strategy to over-express only the subset of genes responsible for the growth advantage if conditions remain stable[57]. Thus, aneuploidy is surprisingly well suited to provide a relatively quick, heritable, and reversible mechanism through which cells and organisms can toggle between phenotypes while searching for optimal fitness solutions.

4.6 FUTURE DIRECTIONS

The most obvious follow-up study would be to extend the search for dosage-sensitive genes that affect fluffy to the entire genome. The efforts made in this direction will be elaborated on in Chapter 5. However, the finding that we were unable to obtain many disomes (II, VI and IX) hints that there are dosage-sensitive genes that affect the *viability* of the strains. Finding out what these genes are and whether they are part of macromolecular complexes (which has been the only proven theory thus far[38]) would be interesting. The chromosomes that also seem to always appear together with another chromosome (e.g. chromosome XII, which is always seen with XV) also hint at this dosage-sensitive viability, with the complementing gene probably lying on the chromosome it has to be duplicated with. Finding such gene pairs would be very illuminating and exciting.

4.7 REFERENCES

1. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-2078 (2005).
2. Jain, N., Hasan, F. & Fries, B.C. Phenotypic Switching in Fungi. *Curr Fungal Infect Rep* **2**, 180-188 (2008).
3. Jain, N. & Fries, B.C. Antigenic and phenotypic variations in fungi. *Cell Microbiol* **11**, 1716-1723 (2009).
4. Levy, S.F., Ziv, N. & Siegal, M.L. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol* **10**, e1001325 (2012).
5. Dawson, C.C., Intapa, C. & Jabra-Rizk, M.A. "Persisters": survival at the cellular level. *PLoS Pathog* **7**, e1002121 (2011).
6. Lohse, M.B. & Johnson, A.D. Differential phagocytosis of white versus opaque Candida albicans by Drosophila and mouse phagocytes. *PLoS One* **3**, e1473 (2008).
7. Huang, G., Srikantha, T., Sahni, N., Yi, S. & Soll, D.R. CO(2) regulates white-to-opaque switching in Candida albicans. *Curr Biol* **19**, 330-334 (2009).
8. Veening, J.W., Smits, W.K. & Kuipers, O.P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62**, 193-210 (2008).
9. Beaumont, H.J., Gallie, J., Kost, C., Ferguson, G.C. & Rainey, P.B. Experimental evolution of bet hedging. *Nature* **462**, 90-93 (2009).
10. Slutsky, B. et al. "White-opaque transition": a second high-frequency switching system in Candida albicans. *J Bacteriol* **169**, 189-197 (1987).
11. Zordan, R.E., Miller, M.G., Galgoczy, D.J., Tuch, B.B. & Johnson, A.D. Interlocking transcriptional feedback loops control white-opaque switching in Candida albicans. *PLoS Biol* **5**, e256 (2007).
12. Kvaal, C. et al. Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of Candida albicans confers increased virulence in a mouse model of cutaneous infection. *Infect Immun* **67**, 6652-6662 (1999).
13. Lachke, S.A., Lockhart, S.R., Daniels, K.J. & Soll, D.R. Skin facilitates Candida albicans mating. *Infect Immun* **71**, 4970-4976 (2003).
14. Lockhart, S.R. et al. In Candida albicans, white-opaque switchers are homozygous for mating type. *Genetics* **162**, 737-745 (2002).
15. LaFleur, M.D., Kumamoto, C.A. & Lewis, K. Candida albicans biofilms produce antifungal-tolerant persister cells. *Antimicrob Agents Chemother* **50**, 3839-3846 (2006).
16. Lachke, S.A., Joly, S., Daniels, K. & Soll, D.R. Phenotypic switching and filamentation in Candida glabrata. *Microbiology* **148**, 2661-2674 (2002).
17. Srikantha, T. et al. Dark brown is the more virulent of the switch phenotypes of Candida glabrata. *Microbiology* **154**, 3309-3318 (2008).
18. Fries, B.C., Goldman, D.L., Cherniak, R., Ju, R. & Casadevall, A. Phenotypic switching in Cryptococcus neoformans results in changes in cellular morphology and glucuronoxylomannan structure. *Infect Immun* **67**, 6076-6083 (1999).
19. Fries, B.C., Taborda, C.P., Serfass, E. & Casadevall, A. Phenotypic switching of Cryptococcus neoformans occurs in vivo and influences the outcome of infection. *J Clin Invest* **108**, 1639-1648 (2001).
20. Cavalieri, D., Townsend, J.P. & Hartl, D.L. Manifold anomalies in gene expression in a vineyard isolate of Saccharomyces cerevisiae revealed by DNA microarray analysis. *Proc Natl Acad Sci U S A* **97**, 12369-12374 (2000).
21. Kuthan, M. et al. Domestication of wild Saccharomyces cerevisiae is accompanied by changes in gene expression and colony morphology. *Mol Microbiol* **47**, 745-754 (2003).
22. Karunanithi, S. et al. Shedding of the mucin-like flocculin Flo11p reveals a new aspect of fungal adhesion regulation. *Curr Biol* **20**, 1389-1395 (2010).
23. Vachova, L. et al. Flo11p, drug efflux pumps, and the extracellular matrix cooperate to form biofilm yeast colonies. *J Cell Biol* **194**, 679-687 (2011).

24. Reynolds, T.B. & Fink, G.R. Bakers' yeast, a model for fungal biofilm formation. *Science* **291**, 878-881 (2001).
25. Vopalenska, I., St'ovicek, V., Janderova, B., Vachova, L. & Palkova, Z. Role of distinct dimorphic transitions in territory colonizing and formation of yeast colony architecture. *Environ Microbiol* **12**, 264-277 (2010).
26. Granek, J.A. & Magwene, P.M. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823 (2010).
27. Ryan, O. et al. Global gene deletion analysis exploring yeast filamentous growth. *Science* **337**, 1353-1356 (2012).
28. Voordeckers, K. et al. Identification of a complex genetic network underlying Saccharomyces cerevisiae colony morphology. *Mol Microbiol* **86**, 225-239 (2012).
29. Holmes, D.L., Lancaster, A.K., Lindquist, S. & Halfmann, R. Heritable remodeling of yeast multicellularity by an environmentally responsive prion. *Cell* **153**, 153-165 (2013).
30. Rose, M.D., Winston, F.M., Hieter, P. & Cold Spring Harbor Laboratory. Methods in yeast genetics : a laboratory course manual. (Cold Spring Harbor Laboratory Press, [Cold Spring Harbor, N.Y.]; 1990).
31. Fay, J.C. & Benavides, J.A. Evidence for domesticated and wild populations of Saccharomyces cerevisiae. *PLoS Genet* **1**, 66-71 (2005).
32. Liti, G., Barton, D.B. & Louis, E.J. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics* **174**, 839-850 (2006).
33. Ezov, T.K. et al. Molecular-genetic biodiversity in a natural population of the yeast Saccharomyces cerevisiae from "Evolution Canyon": microsatellite polymorphism, ploidy and controversial sexual status. *Genetics* **174**, 1455-1468 (2006).
34. Sniegowski, P.D., Dombrowski, P.G. & Fingerman, E. Saccharomyces cerevisiae and Saccharomyces paradoxus coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Res* **1**, 299-306 (2002).
35. Baird, N.A. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376 (2008).
36. Lorenz, K. & Cohen, B.A. Small- and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. *Genetics* **192**, 1123-1132 (2012).
37. Hill, A. & Bloom, K. Genetic manipulation of centromere function. *Mol Cell Biol* **7**, 2397-2405 (1987).
38. Anders, K.R. et al. A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. *BMC Genet* **10**, 36 (2009).
39. Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Futcher, A.B. The WHI1+ gene of Saccharomyces cerevisiae tethers cell division to cell size and is a cyclin homolog. *Embo J* **7**, 4335-4346 (1988).
40. Haase, S.B. & Reed, S.I. Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* **1**, 132-136 (2002).
41. Yvert, G. et al. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat Genet* **35**, 57-64 (2003).
42. Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10**, 1793-1808 (1994).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
44. Tuite, M.F., Mundy, C.R. & Cox, B.S. Agents that cause a high frequency of genetic change from [psi+] to [psi-] in Saccharomyces cerevisiae. *Genetics* **98**, 691-711 (1981).
45. Hvorecny, K.L. & Prelich, G. A systematic CEN library of the Saccharomyces cerevisiae genome. *Yeast* **27**, 861-865 (2010).
46. Cook, J.G., Bardwell, L., Kron, S.J. & Thorner, J. Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast Saccharomyces cerevisiae. *Genes Dev* **10**, 2831-2848 (1996).

47. St'ovicek, V., Vachova, L., Kuthan, M. & Palkova, Z. General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol* **47**, 1012-1022 (2010).
48. Torres, E.M. et al. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916-924 (2007).
49. Pavelka, N. et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**, 321-325 (2010).
50. Palkova, Z. & Vachova, L. Life within a community: benefit to yeast long-term survival. *FEMS Microbiol Rev* **30**, 806-824 (2006).
51. Polakova, S. et al. Formation of new chromosomes as a virulence mechanism in yeast Candida glabrata. *Proc Natl Acad Sci U S A* **106**, 2688-2693 (2009).
52. Selmecki, A.M., Dulmage, K., Cowen, L.E., Anderson, J.B. & Berman, J. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet* **5**, e1000705 (2009).
53. Selmecki, A., Forche, A. & Berman, J. Aneuploidy and isochromosome formation in drug-resistant Candida albicans. *Science* **313**, 367-370 (2006).
54. Sionov, E., Lee, H., Chang, Y.C. & Kwon-Chung, K.J. Cryptococcus neoformans overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. *PLoS Pathog* **6**, e1000848 (2010).
55. Gordon, D.J., Resio, B. & Pellman, D. Causes and consequences of aneuploidy in cancer. *Nat Rev Genet* **13**, 189-203 (2012).
56. Sotillo, R., Schvartzman, J.M., Socci, N.D. & Benezra, R. Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. *Nature* **464**, 436-440 (2010).
57. Yona, A.H. et al. Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci U S A* (2012).

Chapter 5 - BROAD AND DISTINCT MODULATION OF EXTRACELLULAR GENES GUIDE THE DEVELOPMENT OF FLUFFY

This chapter is based on a manuscript in progress:

Zhihao Tan[1,2], Michelle Hays[2], Cecilia Garmendia-Torres[3], Gareth A. Cromie[1], Amy Sirr[1], Eric W. Jeffery[1] and Dudley AM[1,2]. Regulation of biofilm development through the simultaneous activation and repression of functionally distinct extracellular proteins.

1. Pacific Northwest Diabetes Research Institute, Seattle, WA
2. Molecular and Cellular Biology Program, University of Washington, Seattle, WA
3. Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

ZT, MH, CGT and AMD designed experiments; ZT, MH, CGT, AS and EWJ performed experiments; ZT, GAC and AMD performed analysis of the data; ZT and AMD wrote the manuscript.

Acknowledgements go to the Dudley lab, Patrick May, Barak Cohen, Michael White, Maitreya Dunham, Nikita Sakhanenko, David Galas for many helpful discussions. This work is funded by a strategic partnership between the Institute for Systems Biology and the University of Luxembourg. Zhihao Tan is funded by the Agency for Science, Technology, and Research, Singapore.

5.1 ABSTRACT

Biofilm formation by microorganisms is a major cause of recurring infections and their removal 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 and better infection outcomes. In this chapter I outline how we use complex colony morphology of *S.cerevisiae* as a model of biofilm development to uncover novel biology. Starting with an overexpression screen, we first uncover *HEK2*, *SAN1* and *TOS8* as novel regulators of biofilm formation. Subsequent RNA-seq analysis of biofilm and non biofilm-forming strains reveal that cells undertake an intuitive approach to the modification of the cell surface. While genes coding for extracellular products involved in cellular adhesion and structural constituents of the cell-wall are upregulated, genes encoding enzymatic cell wall proteins (e.g. hydrolases, glucosidases) are simultaneously repressed. Our analysis also suggested the involvement of *PHD1* and *SFG1* as potential regulators of these two classes of genes. The effects of deleting members of these two classes of extracellular genes, as well as *PHD1* and *SFG1*, confirmed their roles in biofilm formation. We propose a model in which the coordinated control of similarly localized but functionally distinct proteins allows *S. cerevisiae* to transition from a unicellular to a multicellular growth state.

5.2 INTRODUCTION

Many opportunistic human pathogens form biofilms - highly structured, multicellular communities that are a key factor in persistent infection and antimicrobial drug resistance[1-3]. Biofilms are a major cause of medical-device associated infections[1, 2, 4], chronic infections of the oral[5], respiratory[6] and urinary tract surfaces[7], and their formation often leads to chronic non-healing of wounds[5, 8]. Their natural ability to adhere to many organic and inorganic surfaces and their inherent drug resistance make them exceptionally difficult to remove and are therefore a significant and pressing clinical problem[1, 2, 9]. The transition from a planktonic, unicellular lifestyle

to a sessile, multicellular lifestyle requires the coordinated activation and repression of numerous genetic and biological programs[3, 10]. While some of these programs and pathways have been elucidated, many still remain poorly characterized. A greater understanding of the molecular mechanisms that lead to biofilm formation will help to guide the development of drug therapies that specifically target biofilms, an area that is still greatly underdeveloped[9, 11].

The yeast *Saccharomyces cerevisiae* has been established as a model to study biofilm formation[12]. While most *S.cerevisiae* laboratory strains form smooth, unstructured colonies when grown on solid media, some strains are able to form highly structured "fluffy" colonies that not only resemble bacterial and fungal biofilms, but also harbor properties that are characteristic of biofilms, e.g. presence of an extracellular matrix[13-15], localized expression of drug efflux pumps[15], use of cell-cell communication[15], and increased adherence to inorganic surfaces[16]. While many careful and detailed genetic screens have been performed to try and elucidate the genes involved in colony morphology, they have been all performed in the same strain background[17-19] (Σ1278b). Leveraging the use of yeast strains that are different from any previously used, we seek to get a better understanding of the genes and processes that are involved in the formation of these complex colonies.

5.3 METHODS

IMAGING

Unless stated otherwise, colonies imaged in this study were grown up from single cells micro-manipulated to a distance of 10mm to ensure consistency in growth. All images were taken with a Canon PowerShot SX10IS, F8.0, 1/8-s shutter speed, ISO80 under consistent lighting conditions, imaging distances and zoom. Images were cropped and scaled with Adobe Photoshop CS5.1 with no further modifications to the images.

YEAST STRAINS, MEDIA

Unless noted, standard media and methods were used for growth and genetic manipulation of yeast[20]. Strains used in this study are listed in **Table 5-1**. Strains were grown on YEP + 2% agar plates, with glucose levels varying between 1-2% depending on strain and experiment.

TABLE 5-1 – *S.CEREVISIAE* STRAINS USED IN THIS STUDY

Strain	Genotype	Background	Reference
YO1853	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *pRS41K-KanMX4* (AB47) plasmid	F45 (YPG725)	[21]
YO1773	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *DIG1-pFA6a-KanMX4* (AB340) plasmid	F45	[21]
YO1829	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *SAN1-p5472-KanMX4* plasmid	F45	This study
YO1833	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *TOS8-p5472-KanMX4* plasmid	F45	This study
YO1836	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *YHR177W-p5472-KanMX4* plasmid	F45	This study
YO1845	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine auxotroph, transformed with *SFL1-p5472-KanMX4* plasmid	F45	This study
YO1850	MATa *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*, serine	F45	This study

	auxotroph, transformed with *HEK2-p5472-KanMX4* plasmid		
YO486	MATa *SPS2:EGFP:kanMX4, hoΔ::HphMX6*	Derived from UC5 (Sake)	[22]
YO502	MATα *SPS2:EGFP:NatMX4*, *hoΔ::HphMX6*	Derived from DBVPG1853 (White tecc)	[23]
YO780 (F13)	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4*	Haploid progeny of YO486 and YO502	This study
YO1737	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, dig1Δ::KanMX4*	F13	This study
YO1779	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, tec1Δ::KanMX4*	F13	This study
YO1781	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, ste12Δ::KanMX4*	F13	This study
YO1786	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, dig2Δ::KanMX4*	F13	This study
YO1898	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, san1Δ::KanMX4*	F13	This study
YO1902	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, tos8Δ::KanMX4*	F13	This study
YO1908	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, yhr177wΔ::KanMX4*	F13	This study
YO1910	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, hek2Δ::KanMX4*	F13	This study
YO1914	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4,*	F13	This study

	sfl1Δ::KanMX4		
YO1942	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, cis3Δ::KanMX4*	F13	This study
YO1962	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, msb2Δ::KanMX4*	F13	This study
YO1970	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, flo11Δ::KanMX4*	F13	This study
YO1979	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, bud8Δ::KanMX4*	F13	This study
YO1985	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, pry2Δ::KanMX4*	F13	This study
YO2005	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, sfg1Δ::KanMX4*	F13	This study
YO2008	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, phd1Δ::KanMX4*	F13	This study
YO2015	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, egt2Δ::KanMX4*	F13	This study
YO2022	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, hpf1Δ::KanMX4*	F13	This study
YO2027	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, bud9Δ::KanMX4*	F13	This study
YO2037	MATα *hoΔ::HphMX6*, *SPS2:EGFP:NatMX4, cts1Δ::KanMX4*	F13	This study

MOLECULAR-ORF BARCODED YEAST SCREEN

MoBY library pools were created by pinning the master library[24] (Thermo Scientific) to LB Lennox +Kan+Chlor+Tet Omnitrays and scraping 10 96-well plates into a pool pellet, giving a total of 6 library pools. Selection was performed on YPD + G418 (0.2mg/ml) plates. With an assumed 1000 plasmids per pool, screening 4600 transformants should give a >99% chance of seeing all plasmids at least once. We decided to screen at least 5000 colonies per pool, thus with 6 pools, >30,000 total CFU were screened. ~1000-1500 CFU were plated per bioassay tray (Teknova) for the actual screen. Following transformation, completely smooth as well as intermediate colony phenotypes were picked. Genomic DNA was extracted from candidates and used to create amplicons to Sanger sequence (Beckman Coulter) both the uptag and downtag barcodes in the MoBY vectors. We decided to individually verify only gene candidates that appeared more than once in our screen by transformation of F45 with each plasmid and screening for colony morphology again. Plasmids that induced a phenotypic change again were verified to ensure the documented ORF was present. From an initial 375 candidates at the picking stage, our pipeline left us with 5 verified candidates (*SAN1*, *HEK2*, *TOS8*, *SFL1* and *YHR177W*) at the end.

RNA-SEQ

To ensure consistent development of colonies, colonies harvested for RNA-seq were generated by depositing single cells 12.7 mm apart in a checkerboard pattern with a FACSAria II cell sorter (BD Biosciences, Franklin Lakes, NJ). After 2 days, entire colonies were harvested by scraping them from the surface of the agar plate with a pipette tip. To obtain sufficient amounts of RNA, 5 colonies were pooled together for each sample. Due to logistical reasons, samples had to be collected and processed in 2 batches. The first batch, referred to from here on as Set A, consisted of the samples wild-type – F13, *tec1Δ, dig1Δ* and *sfl1Δ*. The second batch, referred to from here on as Set B, consisted of the samples WT, *tos8Δ* and *yhr177wΔ*. 3 replicates (of 5 colonies each) of Set A and 4 replicates (of 5 colonies each) of Set B were processed further. Total RNA was prepared as described previously[25].

Library preparation: Following extraction, total RNA from the pooled colonies was checked for integrity and quantified on a Bioanalyzer (Agilent). 5ug of total RNA for each sample was then processed into sequencing libraries using the Tru-Seq stranded mRNA kit (Illumina) following manufacturer recommendations for dual-indexing (9 or 12-plex).

Pooling: The individual libraries were then multiplexed into the final pools for sequencing. Each multiplex pool consisted of either 12 (one set of biological replicates for Set A, with only 4 of the 12 are being reported in this study) or 9 individual libraries (one set of biological replicates for Set B, with only 3 of the 9 are being reported in this study) with one replicate for each condition at 10nM each in 50ul. Sequencing was performed on an Illumina Hiseq2000 with 1 pool per lane for paired-end, 50 nucleotide sequencing with index read (Covance, Seattle). Therefore, each Hiseq lane represents one set of biological replicates. A total of 7 Hiseq lanes were used (as Set A had 3 biological replicates, and Set B had 4 biological replicates). Sequencing yielded an average of ~175 million paired-end reads (~350 million total reads) per lane for a total number of paired-end reads of 1,227,909,850. The total number of paired-end reads used in this study is 483,383,969.

DATA PROCESSING

Sequence quality was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which revealed good quality calls across the entire read (lower quartile phred scores ranging from 31-40 across all bases). Therefore, reads were not trimmed and were further processed per se. Alignment was done using Bowtie2[26] (version 2.1.0). As the full genomic sequence and annotation of F13 was not available at the time of analysis, sequences were aligned to the S288C reference genome supplied by SGD (downloaded 4/28/2014) allowing 1 mismatch per read. The GFF file was similarly downloaded from SGD at the same time and was appended with data (also downloaded from SGD on 4/28/2014) to include CUTs[27], SUTs[27], MUTs[28] and XUTs[29]. Read pairs were aligned to the S288C genome using BWA (version 0.7.5a) via the mem command with default settings. Read pairs where neither member of the pair was aligned to the reference were identified and were used to assemble de novo contigs using Velvet

(version 1.2.10) via the commands velveth (kmer length=31) and velvetg (-cov_cutoff 20 -exp_cov 300 -ins_length 250). Nucleotide BLAST was used to identify potential ORFs in the novel contigs, and these ORFs were appended to both the reference sequence and the annotation file. Counting was done using featureCounts (version 1.4.0) in the Subread package[30], with overlapping reads not counted.

DIFFERENTIAL EXPRESSION ANALYSIS

Analysis of differential expression was done using edgeR[31]. The table of raw counts produced by featureCounts was used as the input for edgeR. Data was normalized for library size and library distribution as per the Classic edgeR pipeline. Each subset of the data always had a wild-type control for which comparisons are made to. To look for differentially expressed genes between samples, we conducted pairwise testing, choosing a p-value of 0.01 for significance. The biological coefficient of variation (as determined by the plotBCV function in edgeR) was 0.0843 for Set A and 0.1308 for Set B. This is probably due to sample preparation differences and might account for the smaller number of differentially expressed genes detected for Set B.

LISTS OF DOCUMENTED REGULATED GENES

We turned to Saccharomyces Genome Database[32] (SGD) and Yeastract[33], 2 well-known yeast resources for evidence of regulation. The list of target genes provided by each website collates and combines information from multiple references on both documented binding data (ChIP-chip) and gene expression changes (e.g. mutant vs. wild-type). To obtain as comprehensive a list of known target genes as we could, we decided to combine both lists (union) into 1 and use that as our reference list. Gene lists were obtained from the resource Yeastract and SGD on 2/14/14. Lists were separated into documented binding targets (mostly ChIP-chip data) or differential expression evidence (mostly microarray, transcription factor mutant data). The number of

references would be too many to list here, and readers are advised to obtain the references from the respective databases if interested.

FUNCTIONAL ENRICHMENT OF GENE LISTS

We looked for functional enrichment of gene lists using YeastMine provided on SGD, with analysis performed on 5/5/2014. Enrichment p-values were Holm-Bonferroni corrected, and a max p-value of 0.05 was accepted.

TRANSCRIPTIONAL REGULATION OF GENE LISTS

We looked for evidence of transcriptional regulation using Yeastract. We used the "Rank Genes by TF" function, filtered for documented regulations (DNA binding plus expression evidence, TF acting as both activator or inhibitor) and checked for all TFs using the TF Rank algorithm[34] using a Heat diffusion coefficient of 0.25.

5.4 RESULTS

SCREEN FOR MODULATORS OF FLUFFY REVEALS NEW SUPPRESSORS

Our previous study indicated that changes in copy number of certain chromosomes, and even a single gene (*DIG1*), was sufficient to reduce the fluffy morphology of colonies[21]. To follow up our study, we decided to search for more modulators of this trait by performing a screen for genes whose increase in copy number would cause a reduction in fluffy morphology. To conduct our screen, we transformed F45, the strain characterized in our previous study[21], with a low copy-number plasmid (CEN) library containing 4981 ORFs controlled by their native promoter[24] in a pooled fashion. While the screen was unlikely to be saturated, the screen still revealed 5 genes whose mild increase in copy number led to a reduction in complex colony morphology: *HEK2,*

SAN1, SFL1, TOS8 and *YHR177W* (**Figure 5-1A**). *SFL1* is closely linked to colony morphology as it has been previously identified as a transcriptional suppressor of the flocculation genes and specifically *FLO11*, a gene that is crucial for colony morphology[16, 35, 36]. *Candida albicans SFL1* has also been shown to be a negative regulator of flocculation and filamentation[37]. *YHR177W* is described as a paralog of *MIT1*, a transcriptional regulator of pseudohyphal growth[38], and an ortholog of *WOR1*, a master regulator of the white-opaque switch in *C.albicans*[39]. Interestingly, while *YHR177W* has been shown to have DNA-binding properties, its deletion had little effect in one study[38], and seemed to act as an activator of complex colony morphology in another[40]. To the best of our knowledge, the remaining 3 genes are newly identified regulators of the trait. *SAN1* encodes a ubiquitin ligase[41]; *HEK2* encodes a RNA-binding protein[42]; *TOS8* encodes a homeodomain-containing transcription factor that has targets enriched for bud growth[43].

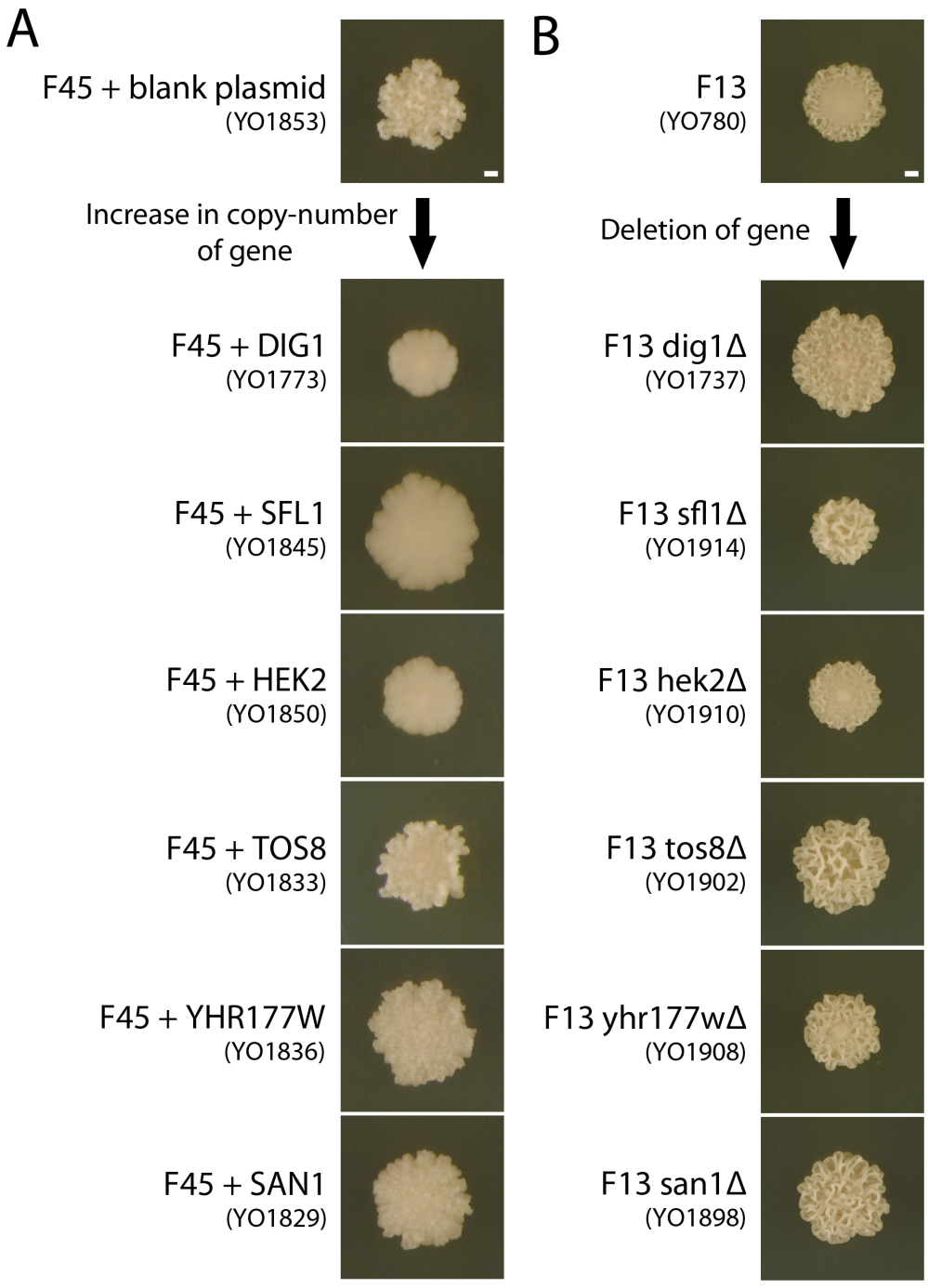


FIGURE 5-1 - EFFECTS OF GENES UNCOVERED IN SCREEN FOR MODULATORS OF FLUFFY. A) INCREASE IN COPY NUMBER OF *DIG1, SFL1, HEK2, TOS8, YHR177W* AND *SAN1* LEAD TO A REDUCTION IN FLUFFY MORPHOLOGY IN STRAIN F45. IMAGES ARE TAKEN ON DAY 4 OF COLONY GROWTH, STRAINS ARE GROWN ON YEP + 2% GLUCOSE AGAR PLATES. B) DELETION OF THE SAME GENES LEAD TO AN INCREASE IN FLUFFY MORPHOLOGY IN STRAIN F13. IMAGES ARE TAKEN ON DAY 5 OF COLONY GROWTH, STRAINS ARE GROWN ON YEP + 1% GLUCOSE AGAR PLATES. SCALE BAR IS 1MM.

We wanted to independently verify the role of these genes in the modulation of the trait, and also attempt to tease out their genetic role. If these genes essentially act as suppressors of the trait, deletion of these genes should increase the fluffiness of colonies. If their stoichiometric ratios in macromolecular complexes are crucial, deleting the genes should lead to a similar effect as increasing their copy number, i.e. a loss of fluffiness. Unable to test this in the strain we carried out our initial screen in (due to its extreme fluffiness), we decided to test this, and all subsequent experiments, in a strain that shares the same background but has a more moderate phenotype (F13, also a haploid progeny of a Japanese sake strain and an Ethiopian white tecc strain, (**Table 5-1**). This strain follows an interesting developmental path where colonies initially develop as completely smooth colonies. As they develop, only the leading growth border of the growing colonies adopt a fluffy morphology, giving rise to smooth center, fluffy border colonies (**Figure 5-1B**). This pattern might vary depending on the distance between the colonies and nutrient availability, but is highly reproducible when these factors are controlled for. More importantly, this strain allows us to test if a perturbation causes either an increase or decrease of the phenotype. We thus tested the deletion of the genes that turned up in our screen, and also the deletion of *DIG1* as a form of positive control. Strikingly, all 6 of our mutant strains showed an increase in the fluffiness of the colonies (**Figure 5-1B**), confirming their role in the modulation of the trait. Our results also strongly indicate that at least in our strain background, these genes most likely are acting as suppressors of the trait.

MODULATION OF TRAIT BY THE FILAMENTATION MAPK PATHWAY

We noticed that *DIG1* is a known repressor of the mitogen activated protein kinase (MAPK) transcription factor Ste12[44], and wanted to delve into this interaction a little further. Ste12 is a transcription factor that regulates multiple pathways, and differentiates between them by changing its binding partners. Ste12 can either activate genes involved in the mating pathway by forming homodimers to bind to pheromone response elements[45] (PREs), or activate genes involved in the pseudohyphal/invasive growth pathway by forming a heterodimer with Tec1 to

bind to filamentation response elements[46] (FREs). If removal of the repressor Dig1 causes an increase in fluffiness, then removal of the activators Ste12 and/or Tec1 should cause a decrease in fluffiness. We created mutants lacking either *STE12* or *TEC1*, and loss of either of these transcription factors had a very strong effect. Colonies formed by both these mutants entirely lost the ability to form fluffy colonies (**Figure 5-2**).

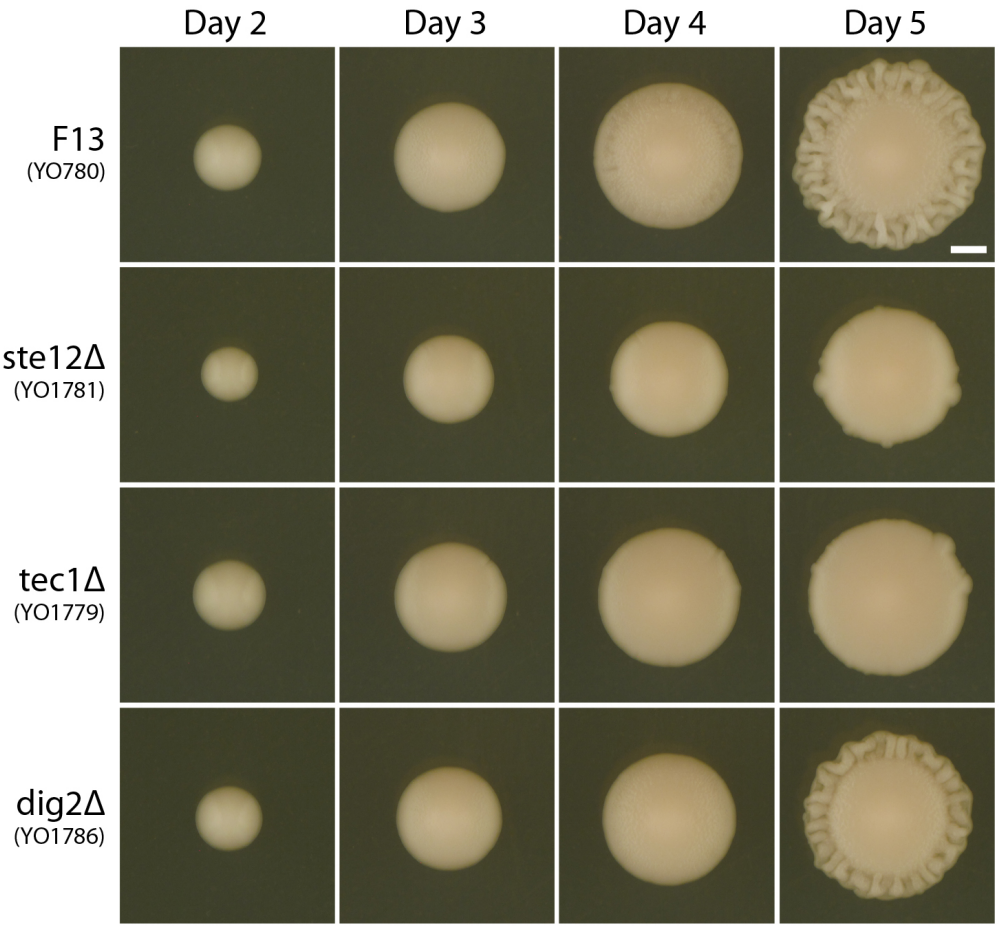


FIGURE 5-2 - EFFECTS OF DELETION OF GENES IN THE MAPK PATHWAY. STRAINS ARE GROWN ON YEP + 1% GLUCOSE AGAR PLATES. SCALE BAR IS 1MM.

With these results, we wanted to see if we could differentiate the relative importance of both pathways. *DIG1* has a paralog, *DIG2*, with which it is able to form repressor complexes[44]. Dig1 is thus able to mediate its effects through 2 repressor complexes: a Ste12/Tec1/Dig1 complex and a

Ste12/Dig1/Dig2 complex[47, 48]. We tested to see if the loss of *DIG2* would cause as strong an effect as a *dig1Δ* mutant. Interestingly, deletion of *DIG2* had very little effect in our strain (**Figure 5-2**). Thus, our results seem to indicate that the Ste12/Tec1/Dig1 repressor complex, and thus perturbation of the filamentation pathway, probably plays a larger role in modulation of the trait.

MANIFOLD GENE EXPRESSION CHANGES FOLLOWING DELETIONS OF KNOWN AND PUTATIVE TRANSCRIPTIONAL FACTORS AND REPRESSORS

Our results thus far implicated many transcriptional activators and repressors, some known and some novel, as modulators of this trait. We decided to focus our attention on them, and use the transcriptional changes that arise in colonies formed by their mutants as a handle to understand the phenotypic changes observed. Picking one of the smooth mutants (*tec1Δ*), the fluffy mutants of the known and putative transcriptional repressors that we picked up in our screen (*sfl1Δ*, *tos8Δ* and *yhr177wΔ*), and the fluffy *dig1Δ* mutant, we performed RNA-seq on cells harvested from their colonies (**Figure 5-3**). We reasoned that analyzing the transcriptome when the colonies were growing the most actively, as opposed to well developed, would provide insight into the processes that are important in the development of a colony. Thus, we extracted RNA from colonies that were harvested in the early stage of colony development (day 2). Due to logistical reasons, we had to separate the samples into 2 experiments, and are thus showing them separately with their respective wild-type controls.

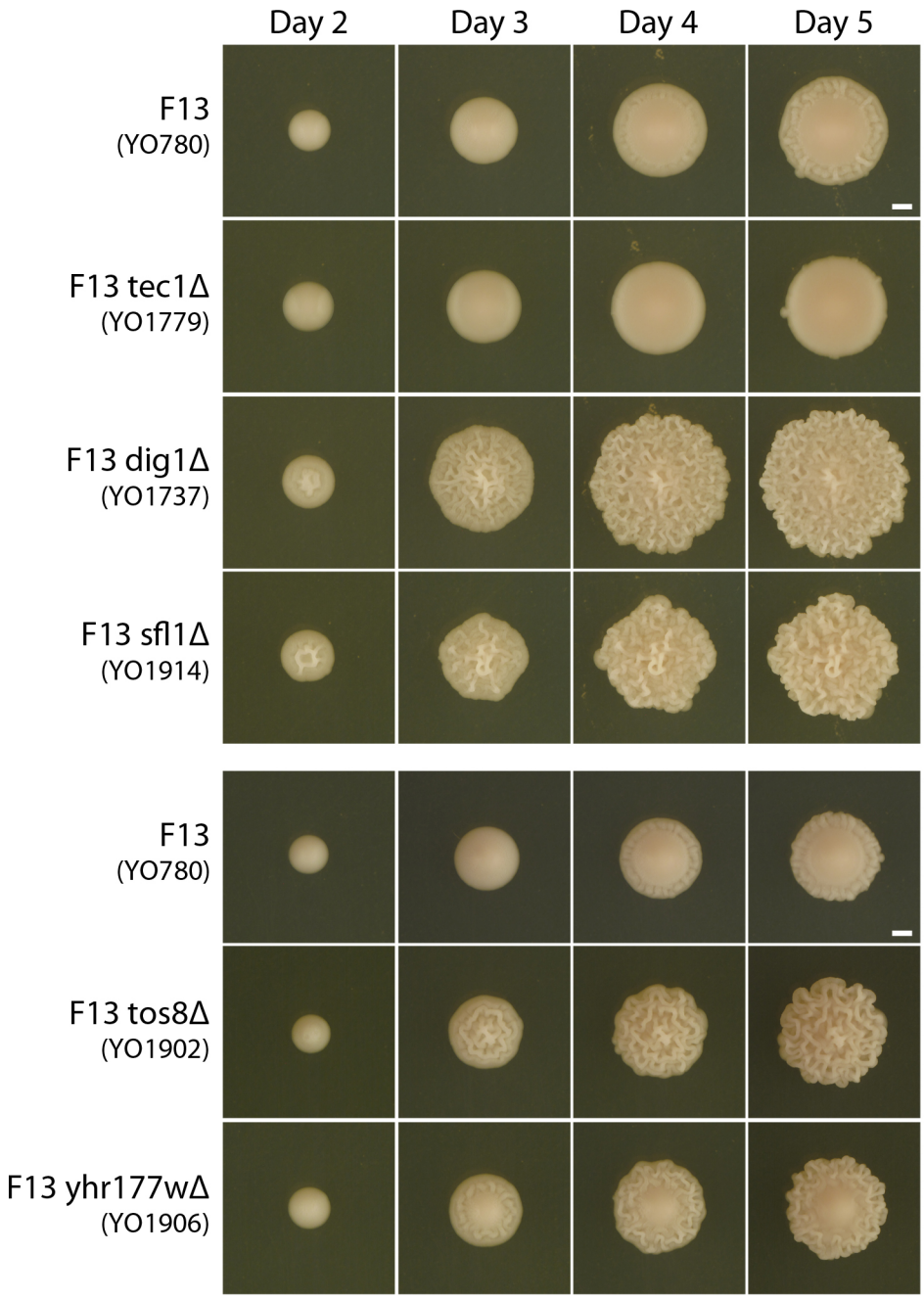


FIGURE 5-3 - COLONY DEVELOPMENT OF STRAINS USED IN RNA-SEQ ANALYSIS. IMAGES SHOW THE DEVELOPMENT OF COLONIES OF WILD-TYPE (F13), *TEC1Δ*, *DIG1Δ*, *SFL1Δ*, *TOS8Δ* AND *YHR177WΔ* NULL MUTANTS. 2 WILD-TYPE CONTROLS ARE SHOWN AS EXPERIMENTS HAD TO BE SEPARATED FOR LOGISTICAL REASONS. STRAINS ARE GROWN ON YEP + 1.5% GLUCOSE AGAR PLATES. SCALE BAR IS 1MM.

As non-coding RNAs (ncRNAs) have been shown to play important roles in the regulation of transcription at specific loci[49, 50], we decided to look for differential expression for both open reading frames (ORFs) and ncRNAs (as annotated in SGD, and adding CUTs[27], SUTs[27], MUTs[28] and XUTs[29]). In addition, we detected genomic sequences in our strain that are not present in the reference S288C genome and thus added predicted genes from those regions to our differential expression analysis pipeline. Differential expression analysis of the samples was performed using edgeR[31], and showed extensive changes in the transcriptome for most mutants compared to our wild-type strain (**Figure 5-4**). To get a broad sense of our data, we first compared just the ORFs that we detected as differentially expressed with previously documented regulation of the known transcriptional regulators (*TEC1*, *DIG1*, *SFL1* and *TOS8*). Also, as the regulation of ORFs by transcription factors can be highly dependent on the experimental conditions (e.g. liquid vs. solid media), it might be enlightening to compare our dataset with others, most of which were collected in liquid media. For each regulator, we obtained a combined list of documented target genes from Saccharomyces Genome Database[32] (SGD) and Yeastract[33], well known resources that document and analyze regulatory associations in yeast. We separated the list into evidence of regulation either by DNA-binding studies (**Figure 5-4,** Binding) or differential expression studies (**Figure 5-4**, Expression), and compared the list with the ORFs that were differentially expressed in our experiment (**Figure 5-4**, Experimental). Overlap between the sets varied, e.g. our *tec1Δ* mutant showed high overlap between experimental and documented ORFs (57.1% previously documented), but our *tos8Δ* mutant showed almost no overlap (3.33% previously documented). Particularly interesting was the group of ORFs that had documented binding evidence but did not previously show any differential expression evidence (lower-left intersection, overlap of Binding (red) and Experimental (green)). With the exception of *TOS8*, we relatively consistently detected about a quarter of them as being differentially expressed under our experimental conditions (*tec1Δ* = 56/225 (24.9%), *dig1Δ* = 211/829 (25.5%), and *sfl1Δ* = 8/26 (30.8%)). We performed functional enrichment analysis on these “newly identified” overlapping ORFs and could detect functional enrichment for our *dig1Δ* mutant, with categories “cytosolic ribosome” (GO:0022626, p-

value=1.79e-3) and “site of polarized growth” (GO:0030427, p-value=2.70e-3) and being the most significant.

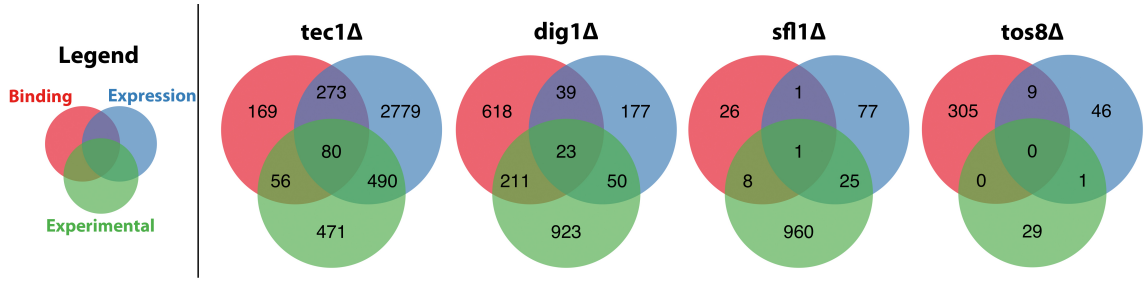


FIGURE 5-4 - OVERLAP OF REGULATED ORFS AS REPORTED PREVIOUSLY WITH DIFFERENTIALLY EXPRESSED ORFS DETECTED IN THIS STUDY. THE 3 CLASSES SHOWN ARE ORFS THAT HAVE DNA-BINDING DATA (BINDING), ORFS THAT ARE DETECTED AS DIFFERENTIALLY EXPRESSED PREVIOUSLY (EXPRESSION), OR ORFS THAT ARE DETECTED IN THIS STUDY AS DIFFERENTIALLY EXPRESSED IN *TEC1Δ*, *DIG1Δ*, *SFL1Δ* AND *TOS8Δ* MUTANTS AS COMPARED TO WILD-TYPE (EXPERIMENTAL).

Our *yhr177wΔ* mutant showed the weakest phenotypic effect (**Figure 5-3**), and consistent with this only showed 11 ORFs (*AQY2, FLO5, FLO9, FLO11, PHO89, RGS2, SPL2, YFL051C, YLL053C, YOR338W* and *YHR177W* itself) and 5 ncRNAs (*PWR1, SUT032, SUT362, MUT533.1* and *SUT1869.2*) as differentially expressed. We note that the biological variation was higher in our second experimental set and might have resulted in a decrease in ability to detect differentially expressed genes. The small number of genes detected as differentially expressed belie their importance though, as many of these genes play important roles in modulating cellular adhesion (e.g. the flocculation genes *FLO5*, *FLO9*, *FLO11* (reviewed in [51]), and the ncRNA *PWR1* that regulates *FLO11* levels[52]).

EXPRESSION DIFFERENCES OF FLUFFY AND SMOOTH STRAINS

We then turned to analyzing the data with the aim of understanding more about the processes involved in the development of these complex structures. We reasoned that comparing the samples that have the largest difference in phenotype would give us the most power to detect transcriptional differences. Thus, we looked for the set of genes that were commonly differentially expressed between the *tec1Δ* mutant (smooth at day 2, develops into an entirely smooth colony) and the *dig1Δ* or *sfl1Δ* mutant (both strongly fluffy at day 2, continue developing as fluffy colonies) (**Figure 5-3**). Our analysis revealed that a total of 261 ORFs and 95 ncRNAs were differentially expressed at a p-value of 0.01 (**Figure 5-5A, Figure 5-5B**). Unsupervised hierarchical clustering of the expression levels of these genes revealed large and distinct modules of both ORFs and ncRNAs that clearly differentiated the "fluffier" colonies (*dig1Δ* and *sfl1Δ* mutants) from the "smoother" colonies (WT and *tec1Δ* mutant) (**Figure 5-5C, Figure 5-5D,** indicated by black bar). Consistent with the strength of the phenotype, the majority of these genes also reflected expression levels in the WT samples that were intermediate between the *tec1Δ* mutant and the *dig1Δ* or *sfl1Δ* mutant. There were sets of genes that were specific to a mutant, but generally these were few in number.

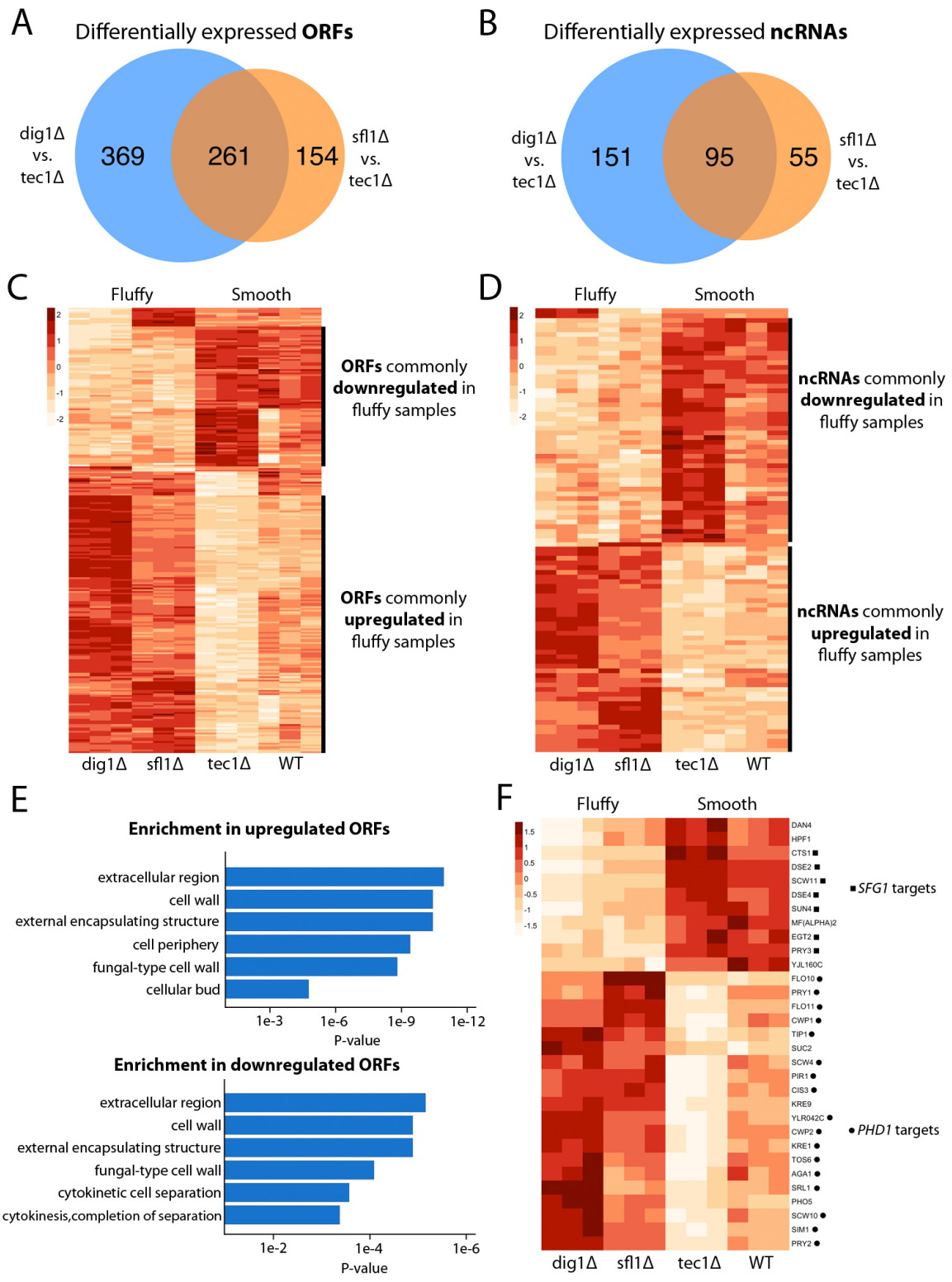


FIGURE 5-5 - ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES WITH RESPECT TO PHENOTYPE. A) OVERLAP OF ORFS DETECTED AS DIFFERENTIALLY EXPRESSED IN A *DIG1Δ* OR *SFL1Δ* MUTANT AS

COMPARED TO A *TEC1Δ* MUTANT. B) OVERLAP OF NCRNAS DETECTED AS DIFFERENTIALLY EXPRESSED IN A *DIG1Δ* OR *SFL1Δ* MUTANT AS COMPARED TO A *TEC1Δ* MUTANT. C) HEATMAP SHOWING UNSUPERVISED HIERARCHICAL CLUSTERING OF THE 261 ORFS PRESENT IN THE INTERSECT ABOVE. SCALE REFLECTS ROW-NORMALIZED LOG COUNTS PER MILLION VALUES. D) HEATMAP SHOWING UNSUPERVISED HIERARCHICAL CLUSTERING OF THE 95 NCRNAS PRESENT IN THE INTERSECT ABOVE. SCALE REFLECTS ROW-NORMALIZED LOG COUNTS PER MILLION VALUES. E) THE 6 FUNCTIONAL CATEGORIES THAT HAVE THE LOWEST P-VALUES FOLLOWING FUNCTIONAL ENRICHMENT ANALYSIS ARE SHOWN FOR BOTH THE ORFS COMMONLY UPREGULATED OR DOWNREGULATED IN THE FLUFFY SAMPLES. F) HEATMAP SHOWING JUST THE GENES ANNOTATED AS PRESENT IN THE "EXTRACELLULAR REGION". DOCUMENTED *SFG1* AND *PHD1* TARGETS ARE MARKED. SCALE REFLECTS ROW-NORMALIZED LOG COUNTS PER MILLION VALUES.

Looking at the 151 ORFs commonly upregulated in the fluffy samples, many genes tightly linked to the trait are seen: both the flocculation genes *FLO10* and *FLO11* are upregulated, the former not normally expressed in lab strains but when expressed induces complex colony morphology[53]; *AQY1* is a gene that, together with *FLO11*, was seen to be always upregulated in a set of natural variants that displayed complex colony morphology[36]; *PHD1* is a transcriptional activator that regulates the levels of *FLO11* among many other genes[54], and is a key target hub protein in pseudohyphal differentiation[55]; *MSB2* and *SHO1* function at the head of the filamentous growth signaling pathway[56], with Msb2 also having been seen to be cleaved and shed from the cell wall into the surrounding region[57]. Other notable genes upregulated in the fluffy samples include *FBP1*, a key gluconeogenesis pathway enzyme, and the bud-site selection genes *BUD8*, *RAX1* and *RAX2*. Interesting gene sets that were identified in the same functional screen include the enolase related region genes *ERR1*, *ERR2* and *ERR3*, the latter two being able to complement the loss of *ENO1* and *ENO2*[58]; targets of SBF complex *TOS2* and *TOS6*; and the cell-wall integrity signaling genes *WSC2* and *WSC3*. Interestingly, most of the Y' helicase genes that are located on the telomeric regions of chromosomes were also detected as upregulated (*YRF1-1, YRF1-3, YRF1-5, YRF1-6, YRF1-7* and *YRF1-8*), fitting well with previous studies showing telomeric genes being often upregulated in fluffy colonies[13, 59].

For the 82 ORFs commonly downregulated in the fluffy samples, there were also a few interesting gene sets. This included the daughter-specific genes *DSE1, DSE2, DSE4*, and many of the heat-shock genes: *HSP30, HSP42, HSP78, HSP82, HSP104* and *SSA3*. Other notable genes include the bud-site selection gene *BUD9*. The entire family of Pathogen-Related Yeast (PRY) genes also came up as differentially expressed, but with *PRY1* and *PRY2* being commonly upregulated and *PRY3* being commonly downregulated in the fluffy samples.

GENES ENCODING FOR EXTRACELLULAR PRODUCTS ARE REGULATED DISTINCTLY ACCORDING TO FUNCTION

Seeking to understand which biological processes could be more important in the formation of these complex colonies, we performed functional enrichment analysis on the differentially expressed ORFs, splitting them up into ORFs that are commonly up or downregulated in the fluffier strains (**Figure 5-5C**). For the ORFs commonly upregulated in the fluffier samples, the most significant categories are “extracellular region” (GO:0005576, p-value=1.13e-11), and “cell wall” (GO:0005618, p-value=3.58e-11) (**Figure 5-5E**, **Materials and Methods**). However, functional enrichment analysis of the ORFs commonly downregulated in the fluffier samples also revealed that the most highly significant categories are once again “extracellular region” (GO:0005576, p-value=7.00e-6), and “cell-wall” (GO:0005618, p-value=1.29e-5) (**Figure 5-5E**).

The results of the functional enrichment analysis surprised us. Expecting categories that would display phenotype specific enrichment, the most significant category of “extracellular region” (including cell wall proteins) was enriched in both sets of ORFs, encompassing a total of 31 out of the 233 genes that we detected as commonly differentially expressed (p-value=6.08e-20, **Figure 5-5F**). However, upon closer inspection, an interesting pattern emerged. Many of the genes encoding for extracellular proteins that are upregulated in the fluffy samples are cell wall genes likely contributing to cellular adhesion due to their serine/threonine rich regions that lead to their highly glycosylated nature (*AGA1, CIS3, CWP1, CWP2, KRE1, KRE9, PIR1, PRY1, PRY2, SRL1*

and *TIP1*), with some also annotated as contributing to the structural integrity of the cell wall (*TIP1, CIS3, CWP1, CWP2, PIR1* and *KRE1*). On the other hand, many of the genes encoding for extracellular proteins that are downregulated in the fluffy samples were also cell wall genes, but concurrently displayed an enzymatic function. For example, 7 of the 11 extracellular genes commonly downregulated were cell wall genes annotated to have hydrolase activity (*SCW11, DSE2, DSE4, HPF1, SUN4, CTS1* and *EGT2*), with the former 5 (out of 7 annotated in the yeast genome) having glucosidase activity.

To test the functional importance of the genes we detected, we decided to perform deletions of some of the genes that we were interested in. All the selected genes came from the sets of ORFs commonly up (*BUD8, CIS3, CWP1, FLO11, PHO5, MSB2, PRY2, SIM1, SRL1, SUC2* and *YLR042C*) or downregulated (*BUD9, CTS1, DSE2, DSE4, EGT2, HPF1, PRY3, SCW11, SUN4* and *YJL160C*) in the fluffy samples, with most of the ORFs annotated as being extracellular. Most of the deletions did not show an effect, or had effects too small to reliably detect. Of those that had an effect, the *flo11Δ*, *cis3Δ* and *msb2Δ* mutants showed the largest effect, with their colonies forming entirely smooth colonies (**Figure 5-6**). The colonies formed by the *bud8Δ*, *bud9Δ*, *pry2Δ* and *hpf1Δ* mutants showed a reduction but not complete loss of ability to form complex morphology, with the *bud9Δ* and *hpf1Δ* mutants results rather surprising as they seemed to go in the opposite direction of what was expected. Only the *cts1Δ* and *egt2Δ* mutants showed a slight increase in complex morphology, seeming to increase the complexity of the fluffy region. Interestingly, all the mutants that show a modulation and not complete loss of the phenotype (*bud8Δ, bud9Δ, pry2Δ, hpf1Δ, cts1Δ* and *egt2Δ*) still maintain the “center smooth outside fluffy” phenotype, only modulating the complexity of the fluffy region. This might be consistent with these genes not affecting the actual developmental path of the colony but interfering with the downstream mechanistic effects.

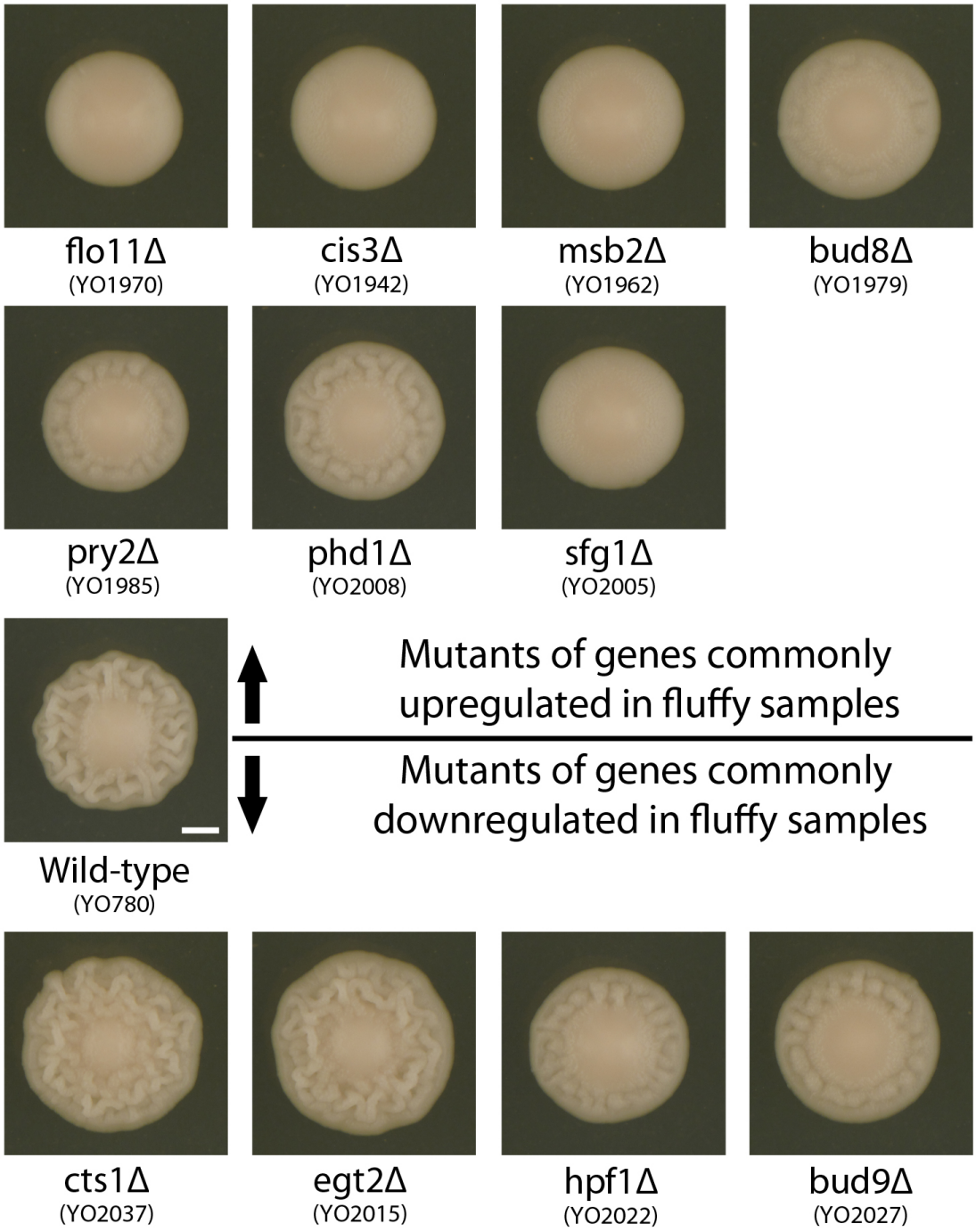


FIGURE 5-6 - EFFECTS OF GENE DELETIONS ON FLUFFY MORPHOLOGY. MUTANTS OF GENES THAT ARE COMMONLY UPREGULATED IN FLUFFY SAMPLES ARE SHOWN ABOVE AND MUTANTS OF GENES COMMONLY DOWNREGULATED IN FLUFFY SAMPLES ARE SHOWN BELOW. IMAGES ARE TAKEN ON DAY 4 OF COLONY GROWTH, STRAINS ARE GROWN ON YEP + 1.5% GLUCOSE AGAR PLATES. SCALE BAR IS 1MM.

POTENTIAL TRANSCRIPTIONAL REGULATION OF MODULES

As a last step in our analysis, since genes that perform the same processes are often coregulated, we wanted to determine if we could detect any transcriptional control of these modules. We used TF Rank (from Yeastract) to perform transcription factor enrichment analysis[34], first using the genes commonly upregulated in the fluffy samples. *PHD1* was a clear top hit in the analysis, with 57 of the 151 commonly upregulated and 18 of the 21 extracellular ORFs being documented to be regulated by *PHD1* (**Figure 5-5F**). Since it also came up in our analysis above as being upregulated in the fluffy samples, we decided to test the loss of *PHD1* in our strain background. The *phd1Δ* mutant displayed an attenuation of the trait, maintaining a smooth center but reducing the complexity of the fluffy region (**Figure 5-6**).

We then turned to the genes commonly downregulated in the fluffy colonies, but there were no clear candidate transcription factors that came up in the enrichment analysis. However, we identified *SFG1*, a very interesting gene that is also commonly upregulated in the fluffy samples. Sfg1 is a putative transcription factor that has been seen to modulate superficial pseudohyphal growth, i.e. growth of pseudohyphae on the surface of the agar[60]. Interestingly, it was also suggested to be a transcriptional repressor for 17 genes in a previous study[61], 7 of which are extracellular genes commonly downregulated in our fluffy samples, and of which 6 have hydrolase activity (**Figure 5-5F**). Since *SFG1* is commonly upregulated in the fluffy samples, it could thus be mediating the repression of the enzymatic cell wall proteins observed. *SFG1* also shows some sequence similarity to *PHD1* and *C.albicans EFG1*, which encodes a transcription factor required in biofilm formation[60, 62, 63]. The finding that opposing transcriptional modules could be regulated by transcription factors that show structural similarity was intriguing, and we thus decided to test the loss of *SFG1*. Rather surprisingly, the loss of *SFG1* had a very strong effect, with colonies forming completely smooth colonies (**Figure 5-6**).

5.5 CONCLUSIONS

NEW AND ALTERED REGULATION

Our initial copy-number screen identified *HEK2*, *SAN1*, and *TOS8* as novel regulators of the trait, with our data arguing that they act as suppressors of the trait. While we chose to focus our attention on the putative transcriptional factors in our screen, *HEK2* and *SAN1* might play interesting roles in the modulation of the trait. Hek2 is a RNA-binding protein that binds the transcripts of many genes, so deciphering the specific mechanism through which it affects the trait might be difficult. However, Hek2 has been shown to directly bind the RNA transcripts of 54 genes that play a role in cell wall function, as well as binding specifically to the repetitive sequences present in the transcript of FLO11[64]. Rather strikingly, its targets include 24 of the 31 extracellular region genes that we detected as differentially expressed, and could thus be affecting the trait by interfering with the translation of these transcripts. San1 is a ubiquitin ligase and the presence of altered levels of many heat-shock proteins in our analysis suggests an interesting link between protein folding and alteration of colony morphology. San1 is nuclear localized[65], and might thus be affecting the trait through modulation of transcription factor protein levels. The recent finding of prion-forming transcription factors that modulate the levels of FLO11 makes the above hypothesis particularly interesting[66].

Our results also indicate that *YHR177W* seems to be acting as a suppressor in our strain background, albeit a relatively weak one with minimal effects on the transcriptome. The strain background does differ between our studies and others, and we did detect single-nucleotide polymorphisms in *YHR177W* and other genes that could possibly affect their function and could explain the differences seen. Since we did not perform any DNA-binding studies, we are unable to predict if the transcriptional regulation by *YHR177W* in our strain is direct or indirect. But the fact that there was good overlap (3 genes) between the genes we detected as differentially expressed (13 total) and Yhr122W's predicted binding sites[38] (where only 24 sites were detected) do however suggest that there is some extent of conservation between our proteins. All 3 such genes (*FLO9*, *FLO11* and *YFL051C*) are upregulated in our mutant. If *YHR177W* is indeed acting

as a transcriptional repressor in our strain, the possible direct derepression of *FLO9* and *FLO11* in our mutant could be enough to explain the phenotypic changes seen.

Apart from our *yhr177wΔ* mutant, we see substantial changes in the transcriptional profiles of the other transcription factor mutants as compared to our wild-type strain. By adding the ncRNAs and non-reference ORFs into our analysis, we also obtained a more comprehensive picture of the transcriptional regulation for this trait. While it is difficult to predict if the differential expression of genes is through direct regulation or the result of transcriptional cascades, we feel it is worth noting from these profiles how their intersection with previously annotated regulation differ, with our study picking up numerous differentially expressed genes that have previous binding data but no previous differential expression data. It is well established that there is often poor overlap between genes whose promoters are bound by a transcription factor and genes whose transcription are perturbed following deletion of the same transcription factor[67, 68]. While this could hint at the complexity of regulation dynamics and co-regulation of genes, another explanation could be that regulation by transcription factors change substantially depending on environmental factors, hinting at the need to assay regulation in different experimental conditions[68]. In our study, we assayed differential expression during growth on solid media, a condition that is arguably less well-studied. With *S.cerevisiae* being commonly found in nature on solid surfaces, e.g. grape skins and tree bark[69], one can imagine that *S.cerevisiae* would have well developed transcriptional responses to growth in this condition. By seeing sharp phenotypic effects from the removal of the transcription factors, we are sure that the transcription factors we tested are indeed having an effect on the developmental processes of the colony, hinting at the value of assaying the regulation of these transcription factors on solid surfaces.

SIMULTANEOUS MODULATION OF COMPLEMENTARY SETS OF EXTRACELLULAR PROTEINS

The formation of these complex multicellular structures is sure to have a physiological effect on the colony, and it has been well established that cellular metabolism in smooth and structured colonies differ[13, 59]. However, we believe that by analyzing the transcriptomes of colonies very early in their development but still displaying strong phenotypic differences, we are shifting the focus away from the metabolic differences present in the more developed colonies used in those studies to the genes that are important in the active development of these complex structures, with our results clearly placing the focus on the modulation of extracellular genes as being crucial.

Our confirmation of *FLO11* as a crucial gene in this trait comes as no surprise as many others have documented its importance previously[16, 35, 59]. However, we also see a large cluster of extracellular genes being upregulated together with *FLO11*. The proteins present on and around the cell surface mediate a cell's interaction with its environment and thus determine the adhesive properties of a cell with its neighbors and other surfaces. The common upregulation of numerous extracellular proteins, many of them heavily glycosylated, thus increases the adhesive properties of the colony and starts to account for the mechanical rigidity and tensile strength needed in the formation of multicellular structures. To our knowledge, this is the first time that *CIS3* and *PRY2* have been implicated in biofilm formation, and to find that *CIS3* has as strong an effect as *FLO11*, i.e. a complete loss of ability to form biofilms, was rather surprising. Cis3 is a protein that has similarities to a family of proteins with internal repeats[70] and Hsp150[71]. While overexpression of *CIS3* has been seen to suppress the temperature-sensitive growth of a *cik1Δ* strain[71], it is probably Cis3's O-glycosylation and localization to the cell wall that is responsible for its contribution to the trait[70, 72]. Pry2 is another O-glycosylated protein that together with Pry1, is involved in the export of sterols from the cell[73]. Their combined upregulation in fluffy samples could indicate an increased secretion of such compounds in fluffy colonies, and that these colonies would be more efficient at the detoxification of small hydrophobic compounds[73].

What might be the most unexpected result we obtained though was the relatively large set of enzymatic cell wall genes that are commonly downregulated in the fluffy samples. Having glucosidases on the cell surface that degrade all the newly formed glycosylated cell-wall proteins would be counter-productive, and thus their downregulation would serve as a good energy conservation strategy in fluffy colonies. There have also been reports of downregulation of glucanases and glycosidases in mature *C.albicans* biofilms, indicating that this might be a conserved response[74]. Many of these downregulated genes have also been annotated as functioning in cytokinesis and daughter cell separation, with the endoglucanase *EGT2* and the endochitinase *CTS1* playing particularly important roles[75, 76]. With our finding that these 2 genes, and furthermore the bud site localization genes *BUD8* and *BUD9* all playing a role in our trait, proper localization of proteins to separating cells thus seems an important factor in the modulation of this trait. The formation of a chain of interconnected cells can be thought of once again as providing additional structural strength for the formation of more complex structures[54]. Alternatively, the formation of branched structures might supply the initial spatial bifurcation that leads to the creation of more complex structures. The formation of linked cells has close links to pseudohyphal differentiation and haploid invasion and might explain why there is often overlap between the genes found[18]. Further studies will be needed to help tease out the similarities and differences between these traits.

While most of these genes present in our clusters are annotated as cell wall proteins, many of them are secreted beyond the cell wall. *FLO11, MSB2, SIM1, PRY1, PRY2* and *SUN4* have all been previously shown to be detected in the medium surrounding the cells[14, 57, 73, 77]. With the static nature of colonies grown on solid medium, these secreted proteins could very well contribute to the extracellular matrix that has been seen in many of these colonies[13]. In that study, investigators were unable to identify the highly glycosylated proteins present in the extracellular matrix despite multiple extraction and treatment techniques, suggesting that not only might there be multiple proteins involved, but also multiple post-translational modifications present, making the extracellular matrix a highly complex and resistant structure. In the same vein, since many of the differentially expressed extracellular hydrolases and glycosidases are at the cell surface, one

can hypothesize that these enzymes could possibly be affecting the trait through the degradation of the extracellular matrix. This dynamic secretion and degradation of extracellular matrix material could potentially provide the cells with protection or nutrients as and when they are needed[78].

FURTHER TRANSCRIPTIONAL CONTROL

Our transcriptomic and regulation analysis provided further insight, picking up *PHD1* and *SFG1* as additional regulators of biofilm formation. *PHD1* has been previously shown to regulate pseudohyphal development, though it has not been assayed or linked to biofilm formation on solid media in previous large-scale screens[17-19]. In our strain, the loss of *PHD1* has a weak effect compared to the other transcription factors found, only displaying a slight attenuation of the trait following its loss. On the other hand, the loss of *SFG1* has a very strong effect in our strain. *SFG1* is a relatively uncharacterized gene whose similarity to *PHD1* and previous role in the modulation of superficial pseudohyphal growth highlighted it as an interesting candidate. The surprisingly strong effect following the loss of *SFG1* potentially suggests an alternate path for the drug targeting of biofilms. While genes that enable an organism to have increased adherence or give it the ability to form an extracellular matrix should be targeted, our findings indicate that interrupting the transcriptional repression of enzymatic cell surface proteins can have as strong an effect in the disruption of the formation of higher order structures. In fact, the finding that many surface adhesins can substitute for each other's function if expressed and localized similarly[79] could mean that targeting a single adhesin might fail to achieve a strong effect, while targeting a gene that serves the same purpose as *SFG1* would have a better outcome.

Our results thus indicate that the formation of complex morphology is highly dependent on an intuitive two-pronged approach to the modification of the cell surface. Though the exact role of *SFG1* has yet to be determined in our strain, we tentatively propose a model of regulation that accounts for our findings thus far (**Figure 5-7**). At the heart of our model lies the bimodular upregulation of structural cell wall proteins that is coordinated with the simultaneous downregulation of enzymatic cell wall proteins. In our strain, *YHR177W* seems to only act through

the repression of the structural proteins, while *TOS8* and *SFL1* act through both the repression of the structural proteins and the repression of *SFG1* that in turn might lead to the activation of the enzymatic proteins. *TEC1* activates both the structural enzymes and *SFG1* (leading to the repression of the enzymatic proteins), while *DIG1* mediates its effects through the repression of *TEC1*. While our model is arguably overly simplistic and does not take into account much of the transcriptional regulation elucidated by many other groups[18, 19, 55], we believe it provides a foundation to start mechanistically understanding this complex phenotype. One of the final steps in the development of biofilms is the dispersal of planktonic cells from the mature biofilm, and requires the reversal of the genetic programs that led to the increased adherence and non-separation of cells needed initially for biofilm formation (reviewed in [10]). Being able to rapidly modulate the levels of a core set of cell adhesion proteins or enzymatic proteins in a coordinated fashion will allow an organism to rapidly switch between these two growth states. The remarkable coordinated control of similarly localized but functionally distinct proteins as observed here suggests an approach that *S.cerevisiae* takes to transition from a unicellular to a multicellular growth state, and should serve as further insight for the generation of drug targets for biofilm formation.

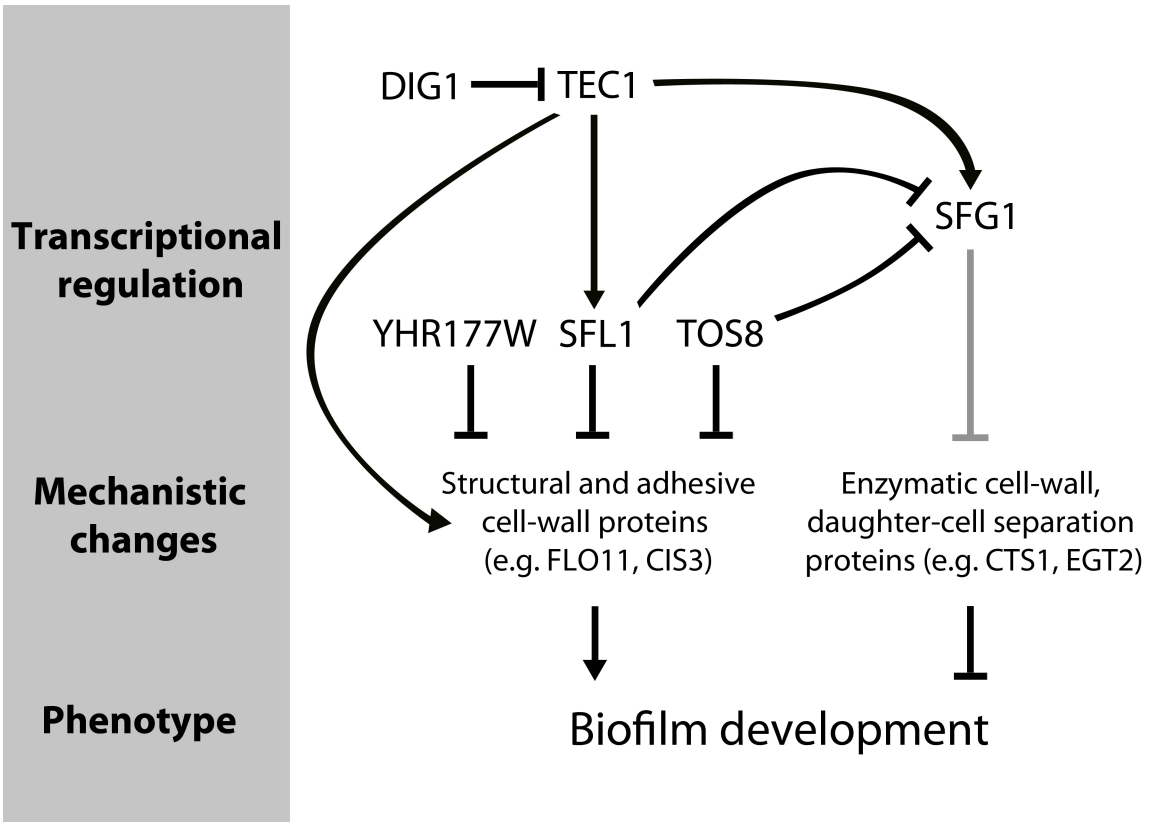


FIGURE 5-7 - TENTATIVE RELATIONSHIP BETWEEN THE TRANSCRIPTION FACTORS AND THEIR EFFECTS ON BIOFILM FORMATION. ARROWS SHOW RELATIONSHIPS AS DETECTED BY THE GENE EXPRESSION CHANGES SEEN IN THIS STUDY. THE RELATIONSHIP BETWEEN *SFG1* AND ENZYMATIC PROTEINS WAS NOT SHOWN IN THIS STUDY AND IS THUS GREYED OUT.

5.6 FUTURE DIRECTIONS

Our screen of genes on chromosome XVI which when increased in copy number will lead to a reduction in fluffy morphology stopped at *DIG1*, though it is clear that there are possibly more causal genes left on the chromosome. This study was an attempt to find the other genes not only on the chromosome, but also in the genome. The lack of saturation in the study even with the large scale of the screen (as judged by the low number of biological repeats in the initial screen), and the lack of completeness of the Moby library (which seems to contain a rather alarming number of false constructs) restricts the conclusions that we can make from this study, and

questions the feasibility of getting a more comprehensive list of genes. There is a little correlation between the location of the genes that we picked up in the screen and the chromosomes which when increased in copy number cause a switch to smooth, though some genes are on chromosomes which disomies we were unable to attain (*HEK2* is on chromosome II, *SAN1* is on chromosome IV). *SFL1* is on chromosome 15, and we do see a switch to smooth with the disome XV strain. The genes that induce the switch on some chromosomes are still unaccounted for though (chromosome III, V and X).

The seemingly complete loss of fluffiness in some of the mutants is particularly interesting and might be worth following up on, especially so for *SFG1* and *CIS3*. Sfg1 is a transcription factor that is not well characterized. It should be worth performing RNA-seq on *sfg1Δ* colonies and probably ChIP-seq. Confirming whether *SFG1* really transmits its effects through the repression of extracellular enzymatic proteins would be satisfying. If that result is positive, an attempt could be made to uncover the exact nature of the mechanism, i.e. is the overexpression of a single enzyme or a combination of them responsible for the attenuation of the trait? While the results of the deletions of the extracellular proteins might hint that more than a single protein contributes to the change in phenotype, an actual overexpression study of the enzymatic proteins could aid in this determination. On the other hand, *CIS3* should be affecting the trait through a completely different mechanism. As a cell wall protein, its effects should be downstream of most pathways, and might be affecting the phenotype in a way that is similar to *FLO11*, i.e. through the gain of intercellular adhesion. Taking a look at the expression profile of a *cis3Δ* mutant should help to confirm this hypothesis. Uncovering the mechanistic role that *CIS3* plays might also help to shed light on the exact role that *FLO11* plays in modulating the trait. Be it through the formation of an extracellular matrix[13, 14], the formation of “velcro-like” fibers[15], or any other mechanism, figuring out if Cis3 shares the same properties as Flo11 and is able to functionally replace Flo11 would be illuminating.

It would also be satisfying to understand the role of *YHR177W* a little more, though previous investigators have had difficulties in uncovering its function and thus indicate that different

approaches might be needed[38]. For example, the binding and expression profiles could be investigated in different environmental conditions and in the presence and absence of its paralog, *MIT1*. Also, if *YHR177W* indeed only regulates a very small set of genes, confirming its role in the direct regulation of *FLO11* could set the stage for further studies, e.g. understanding how *YHR177W* has evolved alongside *MIT1* and why it has a smaller but just as important subset of targets.

Though the variability seen in the second set of experiments should limit its use (**Figure 5-8**), the comparison of all the mutants together should reveal interesting sets of genes to interrogate. For example, looking at the sets of genes detected as differentially expressed between the wild-type and a mutant, the intersection of all the sets reveals (fortunately, and unfortunately) that *FLO11* is central to this trait (**Figure 5-9**). One overlap in this figure that might be interesting to chase down is the "all the mutants minus *YHR177W*" overlap, which only contains 4 genes but contains *SFG1*. The other genes include *BTN2* (v-SNARE binding protein), *DAN4* (a cell wall mannoprotein expressed under anaerobic conditions) and *YLR042C*.

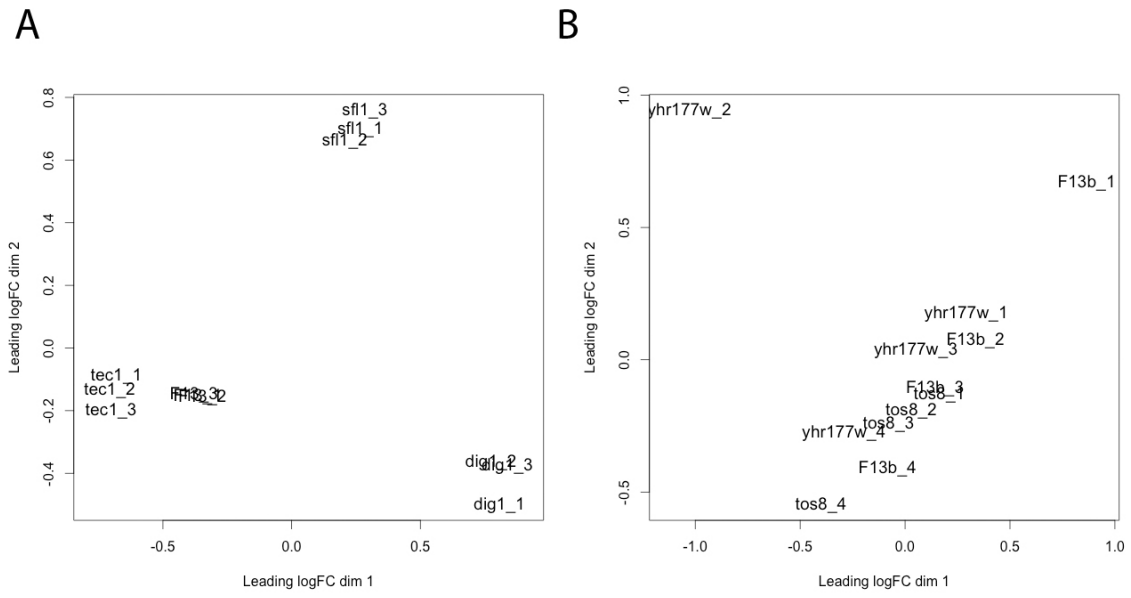


FIGURE 5-8 - MULTIDIMENSIONAL PLOT SHOWING HOW THE REPLICATES FROM EACH STRAIN CLUSTERS. A) CLUSTERING OF SAMPLES ACCORDING TO GENOTYPE IS GOOD FOR THIS EXPERIMENT. THE WT AND *TEC1*

MUTANT (SMOOTH) ARE ALSO CLUSTERED CLOSER TOGETHER THAN THE *DIG1* AND *SFL1* MUTANT (FLUFFY). B) CLUSTERING OF SAMPLES ACCORDING TO GENOTYPE IS POOR FOR THIS EXPERIMENT. REPLICATES ARE NOT CLUSTERING BY GENOTYPE OR PHENOTYPE.

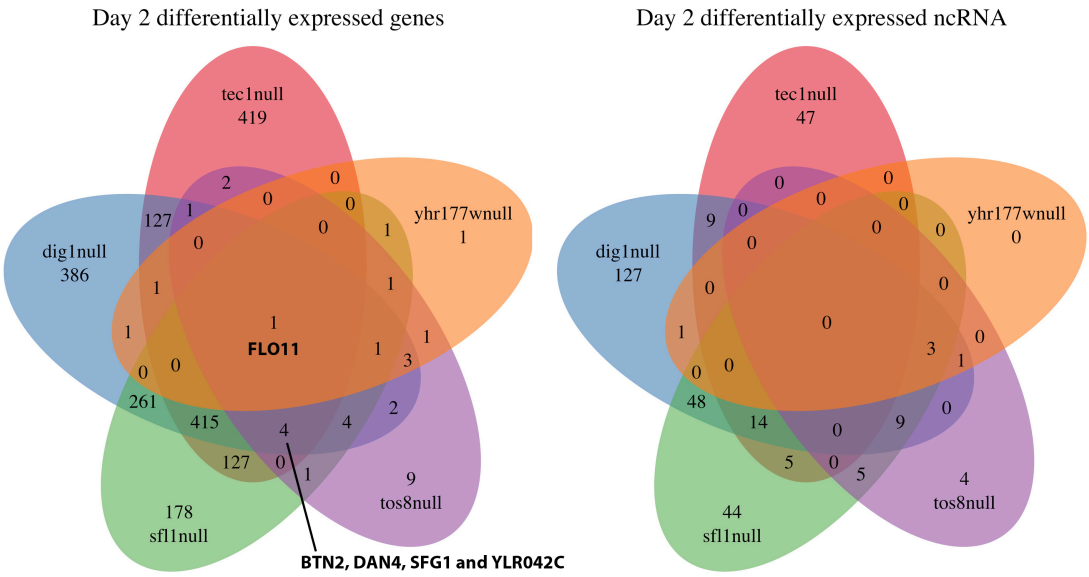


FIGURE 5-9 - OVERLAP OF GENES OR NCRNA THAT ARE DIFFERENTIALLY EXPRESSED IN THE MUTANTS. NOTE THAT THE DIRECTIONS (UP OR DOWNREGULATED) OF DIFFERENTIAL EXPRESSION MIGHT BE DIFFERENT BETWEEN MUTANTS

5.7 REFERENCES

1. Costerton, J.W., Stewart, P.S. & Greenberg, E.P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318-1322 (1999).
2. Donlan, R.M. & Costerton, J.W. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev* **15**, 167-193 (2002).
3. Fanning, S. & Mitchell, A.P. Fungal biofilms. *PLoS Pathog* **8**, e1002585 (2012).
4. Darouiche, R.O. Treatment of infections associated with surgical implants. *N Engl J Med* **350**, 1422-1429 (2004).
5. Mancl, K.A., Kirsner, R.S. & Ajdic, D. Wound biofilms: lessons learned from oral biofilms. *Wound Repair Regen* **21**, 352-362 (2013).
6. Kobayashi, H. Airway biofilms: implications for pathogenesis and therapy of respiratory tract infections. *Treat Respir Med* **4**, 241-253 (2005).
7. Tenke, P. et al. Update on biofilm infections in the urinary tract. *World J Urol* **30**, 51-57 (2012).
8. Scali, C. & Kunimoto, B. An update on chronic wounds and the role of biofilms. *J Cutan Med Surg* **17**, 371-376 (2013).
9. Ramage, G., Culshaw, S., Jones, B. & Williams, C. Are we any closer to beating the biofilm: novel methods of biofilm control. *Curr Opin Infect Dis* **23**, 560-566 (2010).
10. Finkel, J.S. & Mitchell, A.P. Genetic control of Candida albicans biofilm development. *Nat Rev Microbiol* **9**, 109-118 (2011).
11. Pierce, C.G. & Lopez-Ribot, J.L. Candidiasis drug discovery and development: new approaches targeting virulence for discovering and identifying new drugs. *Expert Opin Drug Discov* **8**, 1117-1126 (2013).
12. Reynolds, L.E. et al. Tumour angiogenesis is reduced in the Tc1 mouse model of Down's syndrome. *Nature* **465**, 813-817 (2010).
13. Kuthan, M. et al. Domestication of wild Saccharomyces cerevisiae is accompanied by changes in gene expression and colony morphology. *Mol Microbiol* **47**, 745-754 (2003).
14. Karunanithi, S. et al. Shedding of the mucin-like flocculin Flo11p reveals a new aspect of fungal adhesion regulation. *Curr Biol* **20**, 1389-1395 (2010).
15. Vachova, L. et al. Flo11p, drug efflux pumps, and the extracellular matrix cooperate to form biofilm yeast colonies. *J Cell Biol* **194**, 679-687 (2011).
16. Reynolds, T.B. & Fink, G.R. Bakers' yeast, a model for fungal biofilm formation. *Science* **291**, 878-881 (2001).
17. Granek, J.A. & Magwene, P.M. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823 (2010).
18. Ryan, O. et al. Global gene deletion analysis exploring yeast filamentous growth. *Science* **337**, 1353-1356 (2012).
19. Voordeckers, K. et al. Identification of a complex genetic network underlying Saccharomyces cerevisiae colony morphology. *Mol Microbiol* **86**, 225-239 (2012).
20. Rose, M.D., Winston, F.M., Hieter, P. & Cold Spring Harbor Laboratory. Methods in yeast genetics : a laboratory course manual. (Cold Spring Harbor Laboratory Press, [Cold Spring Harbor, N.Y.]; 1990).
21. Tan, Z. et al. Aneuploidy underlies a multicellular phenotypic switch. *Proc Natl Acad Sci U S A* **110**, 12367-12372 (2013).
22. Fay, J.C. & Benavides, J.A. Evidence for domesticated and wild populations of Saccharomyces cerevisiae. *PLoS Genet* **1**, 66-71 (2005).
23. Liti, G., Barton, D.B. & Louis, E.J. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics* **174**, 839-850 (2006).
24. Ho, C.H. et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol* **27**, 369-377 (2009).

25. Swanson, M.S., Malone, E.A. & Winston, F. SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol Cell Biol* **11**, 4286 (1991).
26. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
27. Xu, Z. et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033-1037 (2009).
28. Lardenois, A. et al. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc Natl Acad Sci U S A* **108**, 1058-1063 (2011).
29. van Dijk, E.L. et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114-117 (2011).
30. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
31. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
32. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-705 (2012).
33. Teixeira, M.C. et al. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic Acids Res* **34**, D446-451 (2006).
34. Goncalves, J.P. et al. TFRank: network-based prioritization of regulatory associations underlying transcriptional responses. *Bioinformatics* **27**, 3149-3157 (2011).
35. Robertson, L.S. & Fink, G.R. The three yeast A kinases have specific signaling functions in pseudohyphal growth. *Proc Natl Acad Sci U S A* **95**, 13783-13787 (1998).
36. St'ovicek, V., Vachova, L., Kuthan, M. & Palkova, Z. General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol* **47**, 1012-1022 (2010).
37. Bauer, J. & Wendland, J. Candida albicans Sfl1 suppresses flocculation and filamentation. *Eukaryot Cell* **6**, 1736-1744 (2007).
38. Cain, C.W., Lohse, M.B., Homann, O.R., Sil, A. & Johnson, A.D. A conserved transcriptional regulator governs fungal morphology in widely diverged species. *Genetics* **190**, 511-521 (2012).
39. Zordan, R.E., Galgoczy, D.J. & Johnson, A.D. Epigenetic properties of white-opaque switching in Candida albicans are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* **103**, 12807-12812 (2006).
40. Furukawa, K., Furukawa, T. & Hohmann, S. Efficient construction of homozygous diploid strains identifies genes required for the hyper-filamentous phenotype in Saccharomyces cerevisiae. *PLoS One* **6**, e26584 (2011).
41. Schnell, R., D'Ari, L., Foss, M., Goodman, D. & Rine, J. Genetic and molecular characterization of suppressors of SIR4 mutations in Saccharomyces cerevisiae. *Genetics* **122**, 29-46 (1989).
42. Irie, K. et al. 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 (2002).
43. Horak, C.E. et al. Complex transcriptional circuitry at the G1/S transition in Saccharomyces cerevisiae. *Genes Dev* **16**, 3017-3033 (2002).
44. Cook, J.G., Bardwell, L., Kron, S.J. & Thorner, J. Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast Saccharomyces cerevisiae. *Genes Dev* **10**, 2831-2848 (1996).
45. Dolan, J.W., Kirkman, C. & Fields, S. The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc Natl Acad Sci U S A* **86**, 5703-5707 (1989).
46. Madhani, H.D. & Fink, G.R. Combinatorial control required for the specificity of yeast MAPK signaling. *Science* **275**, 1314-1317 (1997).
47. Olson, K.A. et al. Two regulators of Ste12p inhibit pheromone-responsive transcription by separate mechanisms. *Mol Cell Biol* **20**, 4199-4209 (2000).
48. Chou, S., Lane, S. & Liu, H. Regulation of mating and filamentation genes by two distinct Ste12 complexes in Saccharomyces cerevisiae. *Mol Cell Biol* **26**, 4794-4805 (2006).

49. Bumgarner, S.L. et al. Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Mol Cell* **45**, 470-482 (2012).
50. van Werven, F.J. et al. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**, 1170-1181 (2012).
51. Teunissen, A.W. & Steensma, H.Y. Review: the dominant flocculation genes of Saccharomyces cerevisiae constitute a new subtelomeric gene family. *Yeast* **11**, 1001-1013 (1995).
52. Bumgarner, S.L., Dowell, R.D., Grisafi, P., Gifford, D.K. & Fink, G.R. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc Natl Acad Sci U S A* **106**, 18321-18326 (2009).
53. Halme, A., Bumgarner, S., Styles, C. & Fink, G.R. Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell* **116**, 405-415 (2004).
54. Pan, X. & Heitman, J. Sok2 regulates yeast pseudohyphal differentiation via a transcription factor cascade that regulates cell-cell adhesion. *Mol Cell Biol* **20**, 8364-8372 (2000).
55. Borneman, A.R. et al. Target hub proteins serve as master regulators of development in yeast. *Genes Dev* **20**, 435-448 (2006).
56. Cullen, P.J. et al. A signaling mucin at the head of the Cdc42- and MAPK-dependent filamentous growth pathway in yeast. *Genes Dev* **18**, 1695-1708 (2004).
57. Vadaie, N. et al. Cleavage of the signaling mucin Msb2 by the aspartyl protease Yps1 is required for MAPK activation in yeast. *J Cell Biol* **181**, 1073-1081 (2008).
58. Kornblatt, M.J. et al. The Saccharomyces cerevisiae enolase-related regions encode proteins that are active enolases. *Yeast* **30**, 55-69 (2013).
59. Stovicek, V., Vachova, L., Begany, M., Wilkinson, D. & Palkova, Z. Global changes in gene expression associated with phenotypic switching of wild yeast. *BMC Genomics* **15**, 136 (2014).
60. Fujita, A., Hiroko, T., Hiroko, F. & Oka, C. Enhancement of superficial pseudohyphal growth by overexpression of the SFG1 gene in yeast Saccharomyces cerevisiae. *Gene* **363**, 97-104 (2005).
61. White, M.A., Riles, L. & Cohen, B.A. A systematic screen for transcriptional regulators of the yeast cell cycle. *Genetics* **181**, 435-446 (2009).
62. Stoldt, V.R., Sonneborn, A., Leuker, C.E. & Ernst, J.F. Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. *Embo J* **16**, 1982-1991 (1997).
63. Ramage, G., VandeWalle, K., Lopez-Ribot, J.L. & Wickes, B.L. The filamentation pathway controlled by the Efg1 regulator protein is required for normal biofilm formation and development in Candida albicans. *FEMS Microbiol Lett* **214**, 95-100 (2002).
64. Wolf, J.J. et al. Feed-forward regulation of a cell fate determinant by an RNA-binding protein generates asymmetry in yeast. *Genetics* **185**, 513-522 (2010).
65. Gardner, R.G., Nelson, Z.W. & Gottschling, D.E. Degradation-mediated protein quality control in the nucleus. *Cell* **120**, 803-815 (2005).
66. Holmes, D.L., Lancaster, A.K., Lindquist, S. & Halfmann, R. Heritable remodeling of yeast multicellularity by an environmentally responsive prion. *Cell* **153**, 153-165 (2013).
67. MacQuarrie, K.L., Fong, A.P., Morse, R.H. & Tapscott, S.J. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* **27**, 141-148 (2011).
68. Hughes, T.R. & de Boer, C.G. Mapping yeast transcriptional networks. *Genetics* **195**, 9-36 (2013).
69. Hyma, K.E. & Fay, J.C. Mixing of vineyard and oak-tree ecotypes of Saccharomyces cerevisiae in North American vineyards. *Mol Ecol* **22**, 2917-2930 (2013).
70. Moukadiri, I., Jaafar, L. & Zueco, J. Identification of two mannoproteins released from cell walls of a Saccharomyces cerevisiae mnn1 mnn9 double mutant by reducing agents. *J Bacteriol* **181**, 4741-4745 (1999).
71. Manning, B.D., Padmanabha, R. & Snyder, M. The Rho-GEF Rom2p localizes to sites of polarized cell growth and participates in cytoskeletal functions in Saccharomyces cerevisiae. *Mol Biol Cell* **8**, 1829-1844 (1997).
72. Mrsa, V., Seidl, T., Gentzsch, M. & Tanner, W. Specific labelling of cell wall proteins by biotinylation. Identification of four covalently linked O-mannosylated proteins of Saccharomyces cerevisiae. *Yeast* **13**, 1145-1154 (1997).

73. Choudhary, V. & Schneiter, R. Pathogen-Related Yeast (PRY) proteins and members of the CAP superfamily are secreted sterol-binding proteins. *Proc Natl Acad Sci U S A* **109**, 16882-16887 (2012).
74. Nett, J.E., Lepak, A.J., Marchillo, K. & Andes, D.R. Time course global gene expression analysis of an in vivo Candida biofilm. *J Infect Dis* **200**, 307-313 (2009).
75. Kovacech, B., Nasmyth, K. & Schuster, T. EGT2 gene transcription is induced predominantly by Swi5 in early G1. *Mol Cell Biol* **16**, 3264-3274 (1996).
76. Kuranda, M.J. & Robbins, P.W. Chitinase is required for cell separation during growth of Saccharomyces cerevisiae. *J Biol Chem* **266**, 19758-19767 (1991).
77. Kuznetsov, E., Kucerova, H., Vachova, L. & Palkova, Z. SUN family proteins Sun4p, Uth1p and Sim1p are secreted from Saccharomyces cerevisiae and produced dependently on oxygen level. *PLoS One* **8**, e73882 (2013).
78. Flemming, H.C. & Wingender, J. The biofilm matrix. *Nat Rev Microbiol* **8**, 623-633 (2010).
79. Guo, B., Styles, C.A., Feng, Q. & Fink, G.R. A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc Natl Acad Sci U S A* **97**, 12158-12163 (2000).

Chapter 6 - CLOSING COMMENTS

In this concluding chapter, I provide a brief summary of our work, and speculate on some of the future areas of advancement that I feel will help to broaden our understanding of this trait.

6.1 SUMMARY

What began initially as a perplexing observation by an undergraduate in the Dudley lab (the appearance of strangely shaped colonies) has now developed into an expansive undertaking with numerous routes of investigation in our lab. The decision was made to pursue the detailed study of colony morphology because it was strongly believed that such a pursuit would provide insight into many important fields of study. While not limited to these, understanding colony morphology will help us broaden our knowledge of fungal biofilms, hint at the regulatory pathways that guide multicellular shape development, and allow us to assess the variability in the genetic architecture of complex traits. While this dissertation has not covered all of these fields, it has unexpectedly covered other areas (e.g. aneuploidy), and much groundwork has been laid by the results obtained in the above studies.

At the time of the start of this thesis, very few systematic studies of the genes underlying this trait had been conducted, and the few studies that had been performed described the trait in a qualitative manner. Pilot studies in our lab quickly led to the recognition that these complex colonies were heavily influenced by the environmental conditions, and in order to obtain more rigorous data that would allow for reliable and reproducible comparisons and further quantitative analysis, we set out to determine suitable environmental conditions for our experiments. We took a major step forward in this area when we realized that micromanipulation of single cells onto the agar at fixed distances produced colonies whose shapes were highly reproducible. This set-up of placing colonies at fixed distances was then improved in a high-throughput manner with the

option of placing single cells using an automated cell sorter. Results from experiments detailing the changes in colony morphology could now be confidently compared and analyzed. The images of the now highly reproducible colony shapes provided the raw material that allowed our collaborators and us to formally create an image analysis pipeline (as described in Chapter 3) that will aid others and us in the quantitative analysis of these complex colonies.

With the framework for the image analysis solidified, I set about to create the hundreds of fluffy strains that would allow us to investigate the genetic determinants of both fluffy and colony shape. However, it was about this time that we became sidetracked with the interesting phenomenon of the reversible fluffy and smooth phenotypic switch in our strains. It was a rather fortunate sidetrack, as not only was the underlying mechanism dissected quickly, our result provided a solid addition to the literature on aneuploidy (as described in Chapter 4). Our discovery that it was the gain and loss of whole chromosomes that was responsible for the phenotypic switching has led not only to the appreciation of how the toggling of this trait can arise abruptly during a single cell cycle and drastically affect both the phenotype and fitness of an organism, but has also added to the growing literature on the effects, both beneficial and detrimental, of aneuploidy. Our work suggests that *S.cerevisiae* could use aneuploidy as a form of bet hedging to reversibly switch between phenotypic states, with this mechanism possibly particularly favorable in nature, where environmental conditions often change quickly and rapidly.

The importance and frequency of aneuploidy that we observe in the strains mentioned in this thesis and other natural variants in the lab then motivated a copy number screen that led to the identification of novel suppressors of this trait (*TOS8, HEK2, SAN1* and *ROF1/YHR177W*). Together with the mutants of the MAPK pathway, a panel of fluffy and smooth strains was thus created, allowing further comparisons to be made between phenotypically different strains from the same genetic background. Transcriptomic analysis of these strains uncovers the importance of genes encoding for extracellular proteins and how these proteins are distinctly regulated according to their function. Detailed analysis also points to a potentially exciting inhibitory regulation of a mechanistic pathway by the relatively unknown transcription factor *SFG1*. If *SFG1*

is indeed a single focal point through which the inhibition of cell-wall enzymes can be controlled, it would represent an exciting target for drug development against biofilm formation. Many genes have also been newly identified as modulators of the trait in this study, and it is exciting that these genes cover both genes that potentially regulate the trait through transcriptional regulation (*TOS8, ROF1/YHR177W, PHD1* and *SFG1*), and genes that regulate the trait through intermediate or downstream mechanisms (*CIS3, HPF1, EGT2, CTS1, BUD8, BUD9, HEK2* and *SAN1*). Subsequent dissection of the molecular mechanisms by which these genes modulate this trait will substantially help further our understanding of this complex trait.

The recent completion of many large-scale studies on colony morphology has started to give investigators a good sense of the pathways and genes involved in this trait. Our own studies have added to this knowledge base, and laid the ground for further investigations. The progress that the field has made in the last few years has been impressive, and it highlights that fact that investigators have recognized of the importance of this trait. Much of the foundation has now been laid, and in the next few years, we should be able to get a good handle on both the broad principles and molecular mechanisms that underlie this visually striking and fundamentally important trait.

6.2 PERSPECTIVES

There are a few broad themes of investigation that I believe can contribute to this field significantly. Firstly, all of the large-scale deletion and mutagenesis screens have been done in the same strain background (Σ1278b), and what we gain in depth by doing so, we miss in breadth. As with any phenotype, assessing the trait in strains from a different background will help to uncover which observations might be strain specific (e.g. every cross between the BY strain and a fluffy strain will lead to the identification of *FLO8* as a QTL), and which might be more general principles that we can move forward with and consolidate around. The field has reached the point where efforts should be put into understanding the diversity present in this trait. The

large strain collection in the Dudley lab that holds many fluffy strains from diverse backgrounds will no doubt help in this pursuit.

Secondly, while many of the transcriptional regulators and nutrient sensing pathways have been implicated in the modulation of the trait, there is still much left to uncover about the downstream effectors. Is it really just all about *FLO11*? It seems that there are proteins that have similar functions to Flo11, i.e. proteins secreted beyond the cell wall, e.g. Msb2, etc. Is Flo11 the most important of all these proteins because it is being expressed at the highest level? Is intercellular adhesion enough, or do you need the creation of an extracellular matrix? Our own transcriptomic study is a small but exciting hint into what some of the other important downstream effectors might be and how they are possibly co-regulated in interesting ways. More importantly, I feel our study is starting to help drive towards a mechanistic understanding of the trait. It seems to me the crux to understanding this trait might lie in uncovering the "why, when, what and how" of the creation of the extracellular matrix. The extracellular matrix has already been shown to be extremely important in bacterial biofilms, and has been characterized much better in bacterial biofilms compared to fungal biofilms. One can imagine though that similar environmental sensing pathways and downstream effectors can be found for both bacterial and fungal biofilms, and many broad principles can be transferred between these areas of study.

As to the reason for my belief of the importance of the extracellular matrix for this trait, it comes down basically to the need for mechanical strength in the creation of complex multicellular structures. In order for crevices, folds, and arches to be formed, there needs to be tensile strength that holds individual yeast cells together. I speculate that there are at least 3 physiological properties that yeast cells have that could aid in creating the mechanical strength needed. They are in increasing importance: the non-division of cells, intercellular surface adhesion and the extracellular matrix.

The non-division of cells is often seen with fungal species, with the formation of multinucleate hyphae a defining feature separating fungi from the unicellular yeasts such as *S.cerevisiae*. *S.cerevisiae* is able to form pseudohyphae though, elongated chains of uninucleate cells that are

connected through incomplete separation of their cell wall. The formation of a chain of cells could be thought of a beam that allows more complex shapes and structures to be formed, and our results implicating *EGT2* and *CTS1*, both involved in proper cell division, might support this hypothesis. Alternatively, the formation of branches of cells could provide the spatial bifurcations needed to initiate colony morphology changes. However, this property would only allow the formation of structures that are dozens of cells wide, and thus probably contributes to the fluffy trait in a minimal extent.

Intercellular adhesion of cells provides the possibility to form structures a lot more complex than those formed by the non-division of cells alone. Adhesion also more importantly provides tensile strength, and that allows cells to be moved and molded *en masse*, with the adhesive strength determining how large the structures can be. This intercellular adhesion has to be performed by the interactions between the cell walls for *S.cerevisiae*, and the presence of many glycosylated proteins on the surface of the cell wall provides for the ability to achieve this. The numerous genes that encode for cell wall proteins that we picked up in our study (e.g. *CIS3, TIP1, HPF1 CWP1, CWP2, PIR1*, etc.) probably contribute to the trait through this mechanism.

Though intercellular adhesion alone could possibly allow the formation of large multicellular structures, whether it alone can allow the formation of colony-sized structures is questionable. This is where I believe the extracellular matrix comes in. It basically acts as flexible “concrete” to fill in the spaces between cells and significantly strengthens the colony structure, allowing the formation of structures that consist of millions to billions of cells. If we are to pursue the investigation of the extracellular matrix, the complexity that the extracellular matrix has been seen to have (probably consisting of multiple proteins each with numerous and diverse post-translational modifications) makes this a formidable task, but the field is taking concrete steps forward. One such step forward would be the gathering of more detailed spatial and temporal information about the development of these colonies. Both others and we have identified many interesting genes that are involved in the modulation of the trait, both at the regulatory and effector levels. Tracking their expression both spatially and temporally in the development of the

colony will help to determine which events are upstream, downstream, and maybe even causal or consequential. The development of high-throughput imaging pipelines by our lab, and the colony section imaging that the Palkova lab performs will no doubt help to drive this area forward. Colony morphology is no doubt a complex trait, and piecing together all the elements will require the further work of many labs, each with a different expertise.

Note on appendices: In the course of my investigations, there have been many more interesting observations and data generated than I have had time to follow up on. The following appendices list some of the promising areas of study that I feel have the potential to be pursued. A brief introduction to the background and motivations that drove some of the hypotheses is given. Some of the pilot experiments conducted are described, and possible future directions are discussed. They are not meant to be comprehensive literature searches on the topics discussed, and results should be verified before proceeding on future work.

Appendix A - QUORUM SENSING

A.1 SMALL MOLECULES THAT INDUCE CHANGES IN THE BEHAVIOR OF MICROORGANISMS

Microorganisms have the ability to communicate with each other using small molecules that they are able to secrete[1, 2]. These molecules can both signal and induce the transition between distinct growth states (partially reviewed in [3]) and allow them to discover the presence and density of organisms that are around them, making it possible for them to modify their growth dynamics in a coordinated fashion. Microorganisms can also use these signals in an attempt to inhibit the growth of other organisms in order to obtain more nutrients and/or territory. Being able to dissect this network of molecules and pathways could allow researchers to create drugs that inhibit the growth and development of pathogenic microorganisms.

For *S.cerevisiae*, the aromatic alcohols phenylethyl alcohol and tryptophol from conditioned medium (medium in which cells were grown in) induced filamentation and induction of *FLO11,* while at the same time reducing filamentation in *C.albicans* [4]. I have conducted pilot experiments that show that phenylethyl alcohol does indeed induce a phenotypic change in our strains. Interestingly, phenylethyl alcohol, but not ethanol, reduces colony morphology in F45 (**Figure A-1B, D**), seemingly in opposition to what would be expected in an *S.cerevisiae* strain (as induction of *FLO11* should make a colony more fluffy). The reduction in fluffy morphology also seems to correspond with a slight loss of invasion into the agar (**Figure A-1C**). This effect seems to be dependent upon genotype though, as phenylethyl alcohol does not have an effect on F13 (**Figure A-1A**). If the response is indeed varying according to genotype, teasing out the genetic differences could point towards genes or pathways that play important roles in this modulation.

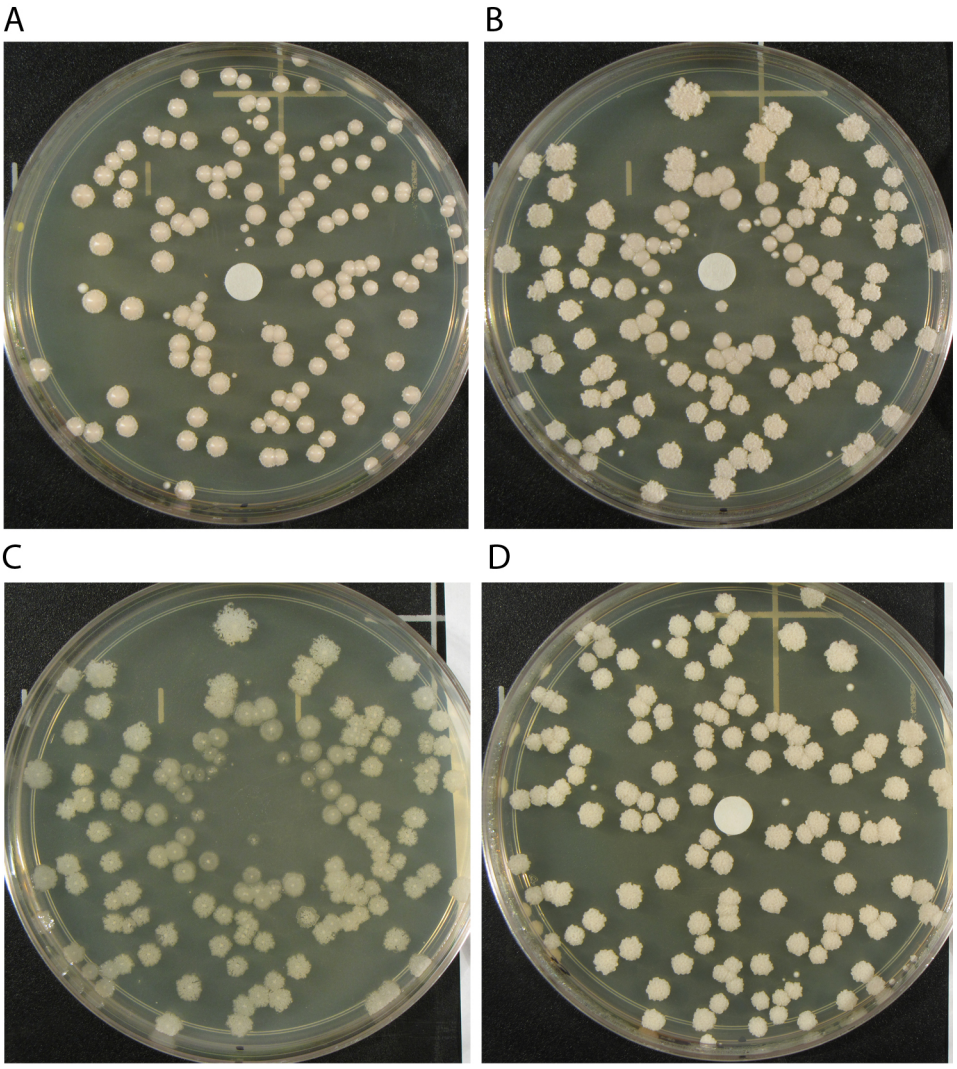


FIGURE A-1 – PHENYLETHYL ALCOHOL AFFECTS COLONY MORPHOLOGY OF F45 BUT NOT F13. A) F13 COLONIES GROWING UP AROUND A DISC CONTAINING 100% PHENYLETHYL ALCOHOL. B) F45 COLONIES GROWING UP AROUND A DISC CONATINING 100% PHENYLETHYL ALCOHOL. NOTE THE LOSS OF COLONY MORPHOLOGY AROUND THE DISC. C) PLATE FROM (B) AFTER WASHING, SHOWING A SLIGHT DIFFERENCE IN INVASION OF COLONIES TOO. D) A FILTER DISC CONATINING 100% ETHANOL DOES NOT AFFECT COLONY MORPHOLOGY OF F45.

An increase in filamentaion does not always equate to an increase in fluffy morphology though, and an observation that would be interesting to make is whether or not F45 shows increased filamentation despite its reduction in fluffy morphology. Beyond that, there might be the need to

show the levels at which phenylethyl alcohol is being produced, or to find other molecules that are being produced. To that end, an interesting experimental protocol could be to use centrifugal spin columns (**Figure A-2**). Cool molten agar can be poured into centrifugal columns and allowed to solidify. Colonies are then grown on the surface for a desired time duration and the column can be centrifuged at low speeds to collect the “secretome”, i.e. compounds that have been released from the colony into the agar, of the colony.

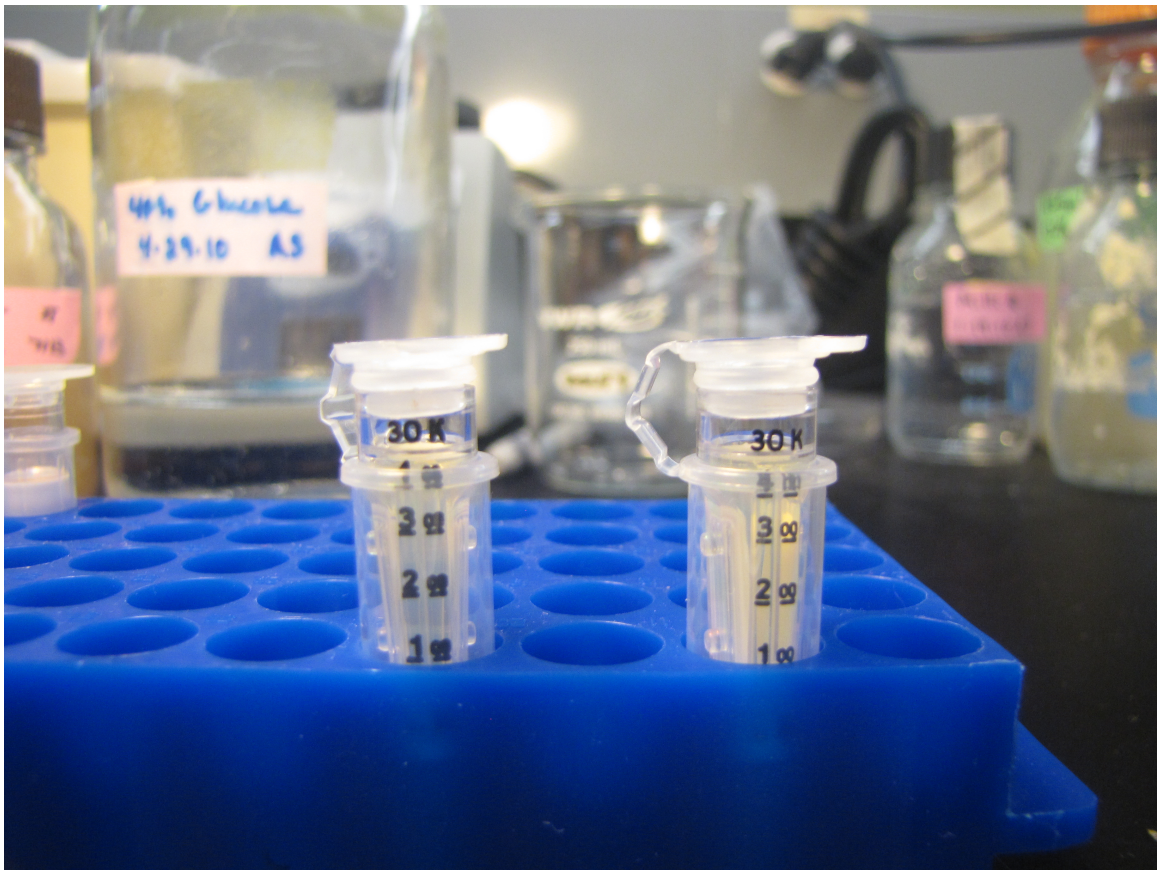


FIGURE A-2 - CENTRIFUGAL SPIN COLUMNS CAN BE USED TO GROW FLUFFY COLONIES

A.2 POSSIBILITY FOR AUTOINDUCTION OF FLUFFY MORPHOLOGY BY MATING PHEROMONES

Among some of the surprising and interesting results from Chapter 5 is the identification of the mating pheromone *MF(ALPHA)2* as a gene that is differentially expressed between fluffy and smooth strains (**Figure 5-5F**). Extra caution should be taken in the interpretation of this result as Tec1 and Dig1 have interactions with Ste12 and this result might be solely due to the perturbations made. Nonetheless, the hypothesis of mating pheromones being able to induce biofilm formation is not completely unfounded as it has been recently seen in *C.albicans*[5] and *C.tropicalis*[6]. Also, both others[7] and I have seen that a difference in mating type can lead to different colony morphologies (**Figure A-3**). The finding that mating pheromone can induce cell death also points to a possible role for it in the remodeling of colony morphology[8].

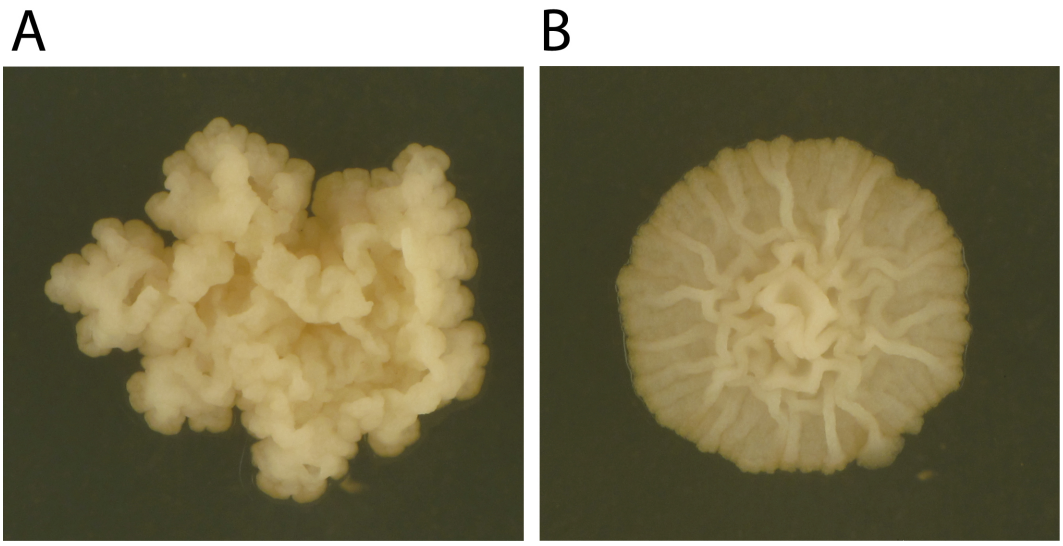


FIGURE A-3 - MATING TYPE CAN AFFECT COLONY MORPHOLOGY. A) COLONY MORPHOLOGY OF F45 MATING TYPE A (YO894). B) COLONY MORPHOLOGY OF F45 MATING TYPE ALPHA (YO970)

Also rather surprisingly, the same RNA-seq experiment also revealed that instead of being completely repressed, very low levels of STE2 (the receptor for the alpha factor) RNA are being

expressed in an alpha strain (which secretes alpha factor, **Figure A-4**). If this translates into low levels of Ste2 on the cell surface, the possibility exists that there is a low-level autoinduction of filamentation/biofilm pathways in the colony by the pheromones secreted by the cells in the colony itself. The recent finding that an artificially created circuit that allowed the secretion and sensing of the same molecule allowed yeast cells to achieve diverse social behaviors also supports this possibility[2].

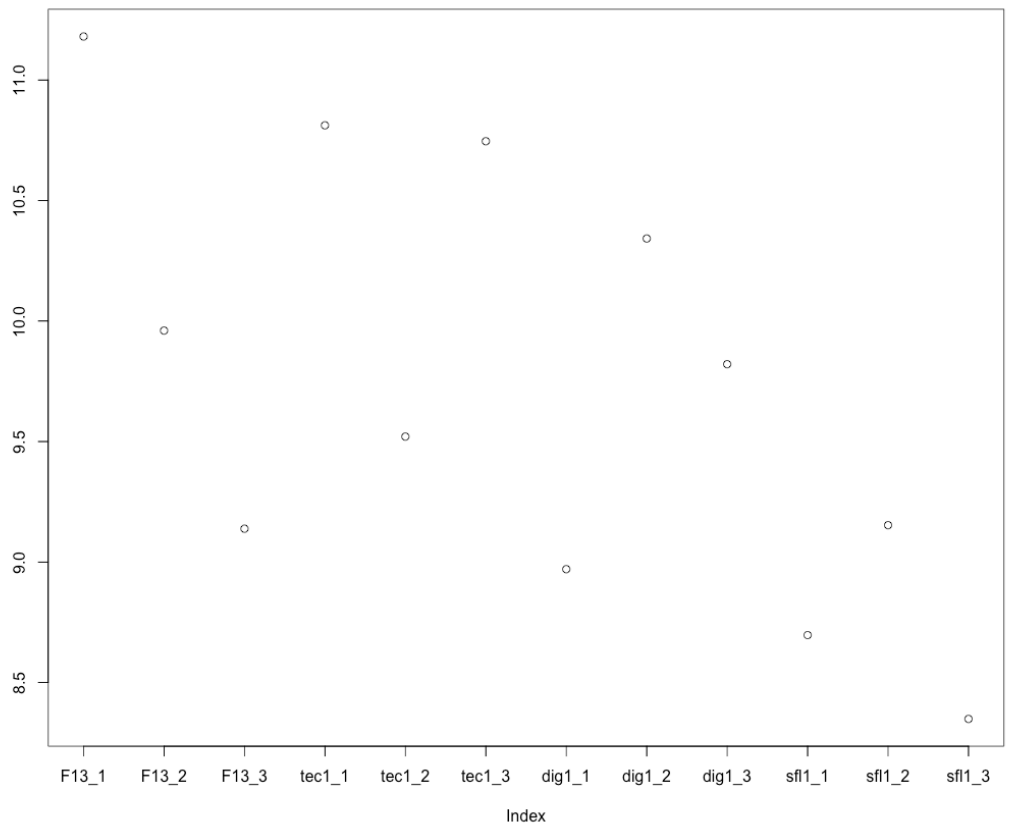


FIGURE A-4 – STE2 CPM OF DAY 2 SAMPLES OF WT, TEC1NULL, DIG1NULL AND SFL1NULL COLONIES

In a similar vein, *C.tropicalis* has been seen to achieve expansion of biofilms through the ability to switch mating types[6]. Thus, another interesting experiment that could be done is to reintroduce *HO* into fluffy strains. The *HO* endonuclease that allows mating type switching has been removed

in most, and also our, lab strains in order to not confound results. It would be interesting to analyze the development of fluffy colonies in strains in which a functional *HO* still exists. Additional experiments that would be interesting to attempt are the deletions of *MF(ALPHA)1* (the major gene locus that encodes mating factor alpha), *BAR1* (the extracellular protease that breaks down alpha factor) and *STE2* (the receptor for alpha factor).

A.3 REFERENCES

1. Fuqua, W.C., Winans, S.C. & Greenberg, E.P. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* **176**, 269-275 (1994).
2. Youk, H. & Lim, W.A. Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* **343**, 1242782 (2014).
3. Finkel, J.S. & Mitchell, A.P. Genetic control of Candida albicans biofilm development. *Nat Rev Microbiol* **9**, 109-118 (2011).
4. Chen, H. & Fink, G.R. Feedback control of morphogenesis in fungi by aromatic alcohols. *Genes Dev* **20**, 1150-1161 (2006).
5. Lin, C.H. et al. Genetic control of conventional and pheromone-stimulated biofilm formation in Candida albicans. *PLoS Pathog* **9**, e1003305 (2013).
6. Jones, S.K., Jr., Hirakawa, M.P. & Bennett, R.J. Sexual biofilm formation in Candida tropicalis opaque cells. *Mol Microbiol* **92**, 383-398 (2014).
7. Reynolds, T.B. & Fink, G.R. Bakers' yeast, a model for fungal biofilm formation. *Science* **291**, 878-881 (2001).
8. Severin, F.F. & Hyman, A.A. Pheromone induces programmed cell death in S. cerevisiae. *Curr Biol* **12**, R233-235 (2002).

Appendix B - MANIFOLD RNA CHANGES IN MUTANTS

B.1 TEC1 LOCUS CONTAINS A NCRNA UPSTREAM OF ORF

A rather shocking finding with the RNA-seq analysis presented in Chapter 5 was the presence of an unannotated sense transcript that is upstream of *TEC1*. This transcript is ~1kb long, seems to be distinct from the TEC1 transcript itself, and is expressed to strikingly high levels in the *dig1Δ* mutant (**Figure B-1, Figure B-2**).

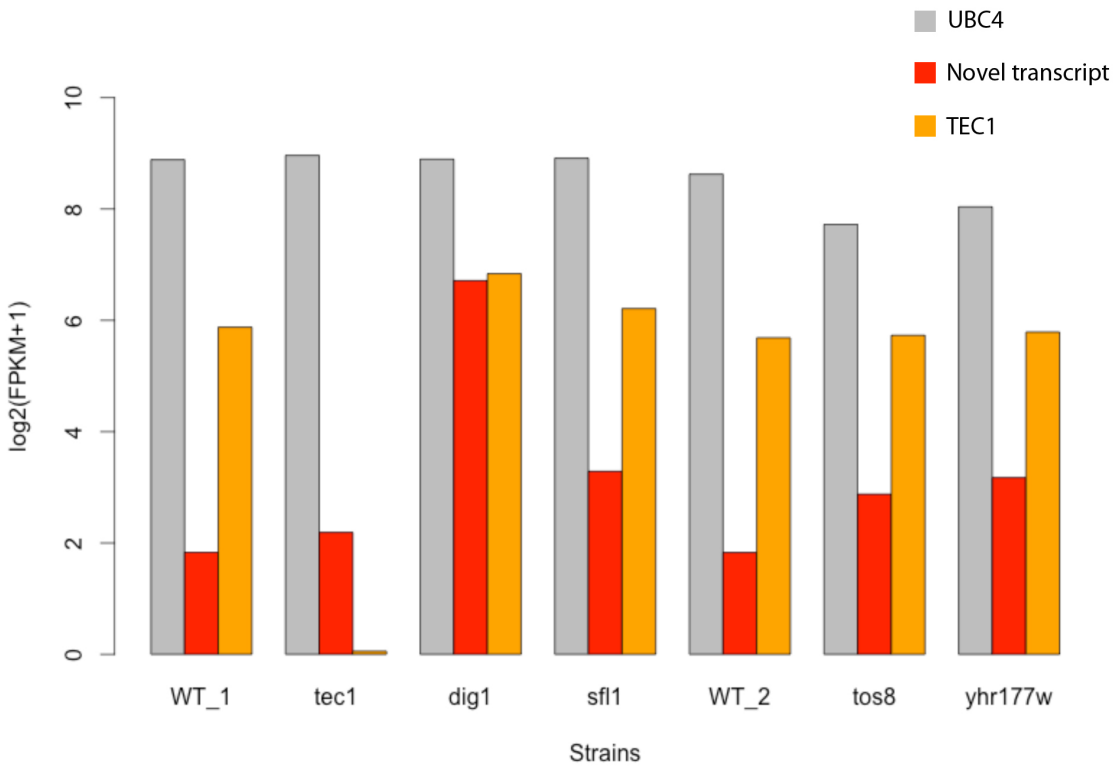


FIGURE B-1 - LOG2(FPKM+1) VALUES OF UBC4 (NEIGHBOR OF *TEC1*), TEC1 AND THE NOVEL TRANSCRIPT (GIVEN A LENGTH OF 1KB, COUNTING UPSTREAM FROM THE START CODON OF *TEC1*) ACROSS THE VARIOUS MUTANTS CREATED. THE LEVELS OF THE NOVEL TRANSCRIPT INCREASES ~8 FOLD FOLLOWING DELETION OF *DIG1*.

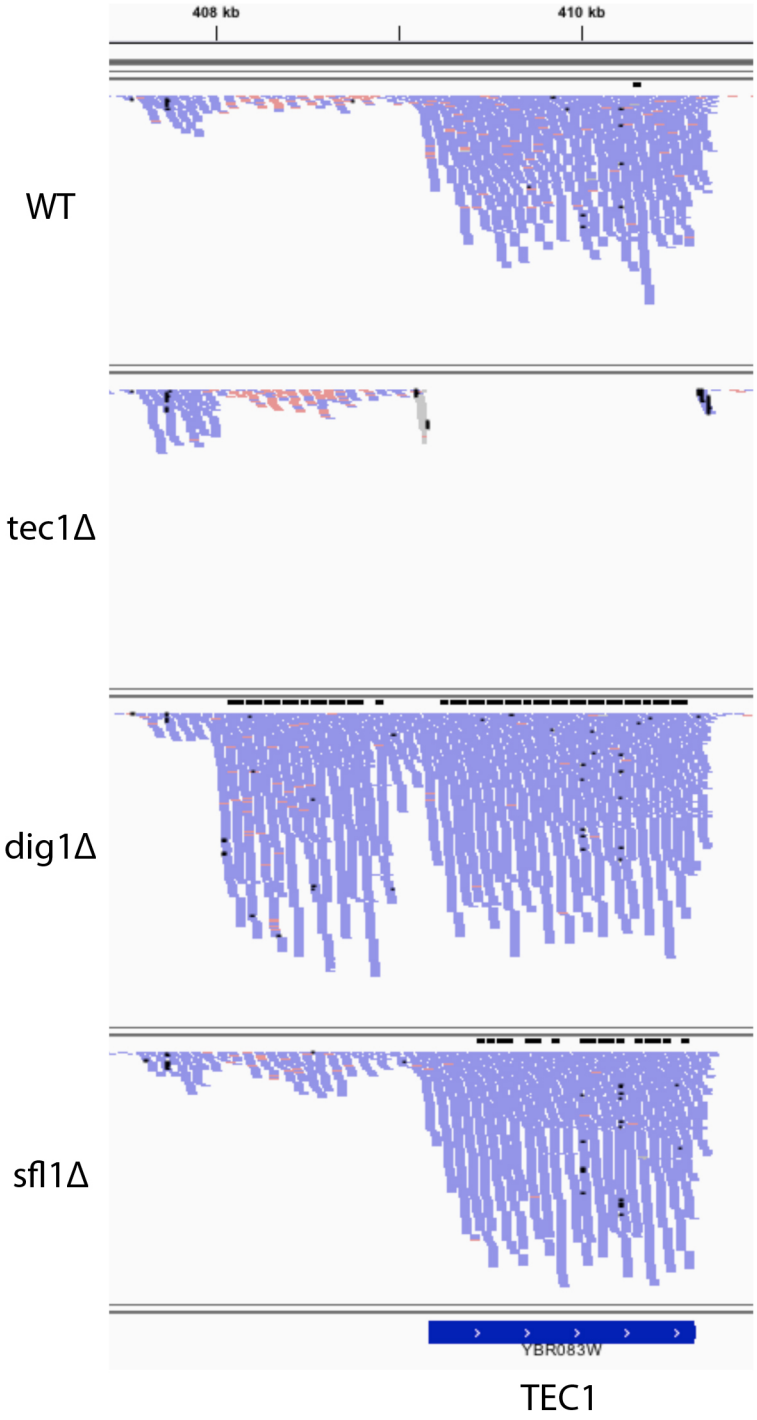


FIGURE B-2 - SCREENSHOT FROM IGV SHOWING THE *TEC1* LOCUS AND THE PRESENCE OF THE NOVEL TRANSCRIPT IN THE *DIG1* MUTANT.

No ORFs seemed to be present in any of the 3 translations of the transcript, and BLAST did not turn up any hits either. It thus seems likely to be a ncRNA that might be having a cis effect on the transcriptional levels of *TEC1* itself. A similar architecture of having sense transcription upstream of the 5' end of an ORF seems to generally attenuate transcription of the ORF itself, probably due to interference of transcription factor binding at the promoter of the ORF[1]. An interesting hypothesis would be that the massive upregulation of the novel transcript could be attenuating the TEC1 RNA levels in a *dig1Δ* mutant, and removal of the novel transcript would increase TEC1 RNA levels. However, Dig1 is a negative regulator of Tec1[2], and TEC1 RNA levels are already increased in a *dig1Δ* mutant (**Figure B-1**), albeit to a small degree. Thus, the presence of the novel transcript could be dialing back the upregulation of TEC1 RNA levels in a *dig1Δ* mutant. This seeming double negative regulation of *TEC1* by the presence and absence of *DIG1* would be rather startling, to say the least.

B.2 CUTs, SUTs, MUTs AND XUTs, OH MY!

In recent years, systematic attempts have been made to annotate the ncRNAs present in the yeast genome. From these works come the ncRNA categories of cryptic unstable transcripts (CUTs)[3], stable uncharacterized transcripts (SUTs)[3], meiotic unannotated transcripts (MUTs)[4] and XRN1-sensitive unstable transcripts (XUTs)[5]. While some seem to be general transcripts that are quickly degraded and can only be seen if the exonuclease *XRN1* is removed[5], others seem to be expressed at specific timings, e.g. meiosis[4]. Since their discovery though, little progress has been made in ascribing functions to these transcripts.

A finding from Chapter 5 that did not receive much attention is the large cluster of C/S/M/XUTs that are indeed differentially expressed between fluffy and smooth samples (**Figure 5-5D**). These ncRNAs show very similar profiles to the ORFs, i.e. clear clusters differentiating fluffy and smooth, with WT levels often intermediate between *tec1Δ* and *dig1Δ* or *sfl1Δ*. It should be worth systematically combing through this list to find and categorize interesting transcriptional profiles,

e.g. transcripts antisense or upstream of other ORFs, transcripts not close to other ORFs, etc. An enlarged version of the heatmap with names attached is shown below (**Figure B-3**).

FIGURE B-3 - NCRNAS THAT ARE DIFFERENTIALLY EXPRESSED IN FLUFFY AND SMOOTH STRAINS. FIGURE IS SIMILAR TO FIGURE 5-5D BUT WITH NAMES INCLUDED.

B.3 ORFs THAT HAVE INTERESTING TRANSCRIPT PROFILES

Using Integrated Genome Viewer (IGV) to manually look at some of the raw transcript profiles of ncRNAs and ORFs of interest revealed loci that were very interesting. Not having time to follow up on any of these, I will list some of the striking ones that I have found below. It should be noted the sequences were mapped to the S288C reference genome, and that some of these genes are located close to the telomeres and could have highly divergent sequences leading to erroneous mapping of transcripts.

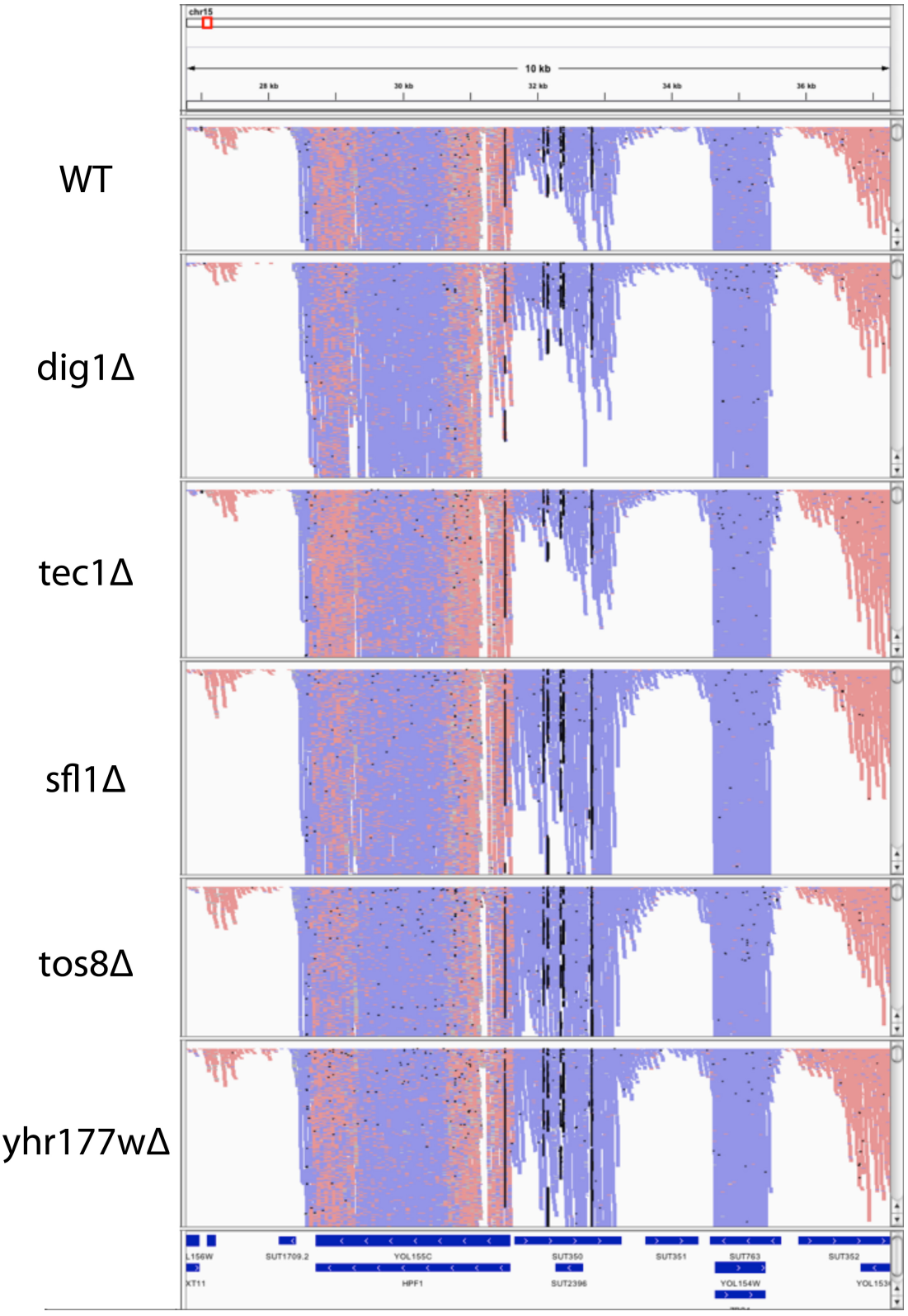


FIGURE B-4 - HPF1 (COLORED RED) SHOWS INTERESTING INTERNAL TRANSCRIPTIONAL PROFILE, AND MASSIVE ANTISENSE TRANSCRIPTION

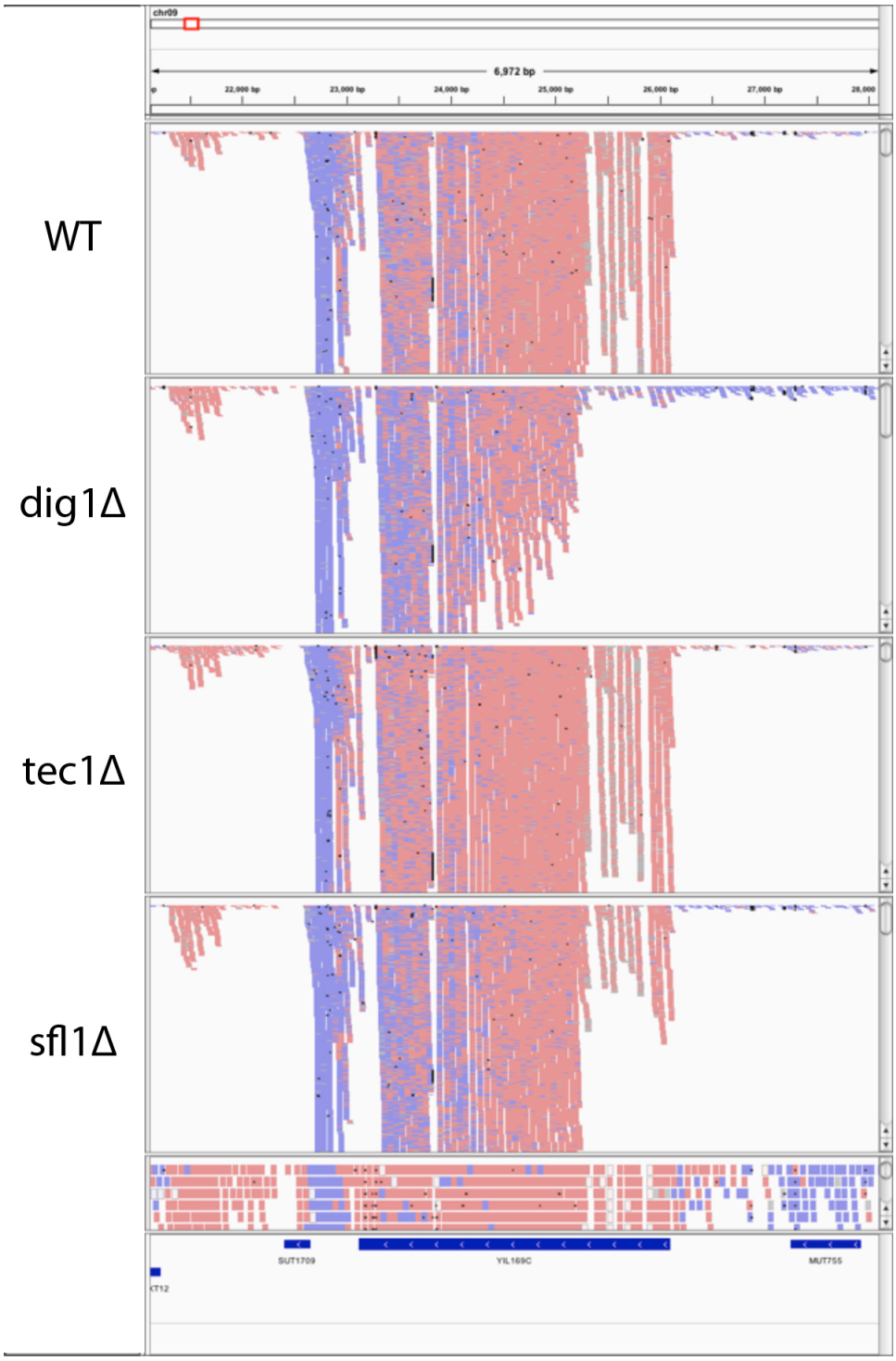


FIGURE B-5 - HPF1'/YIL169C (RED) ALSO SHOWS INTERESTING INTERNAL TRANSCRIPTION, WITH THE 5' END PARTICULAR INTERESTING FOR THE *DIG1* MUTANT.

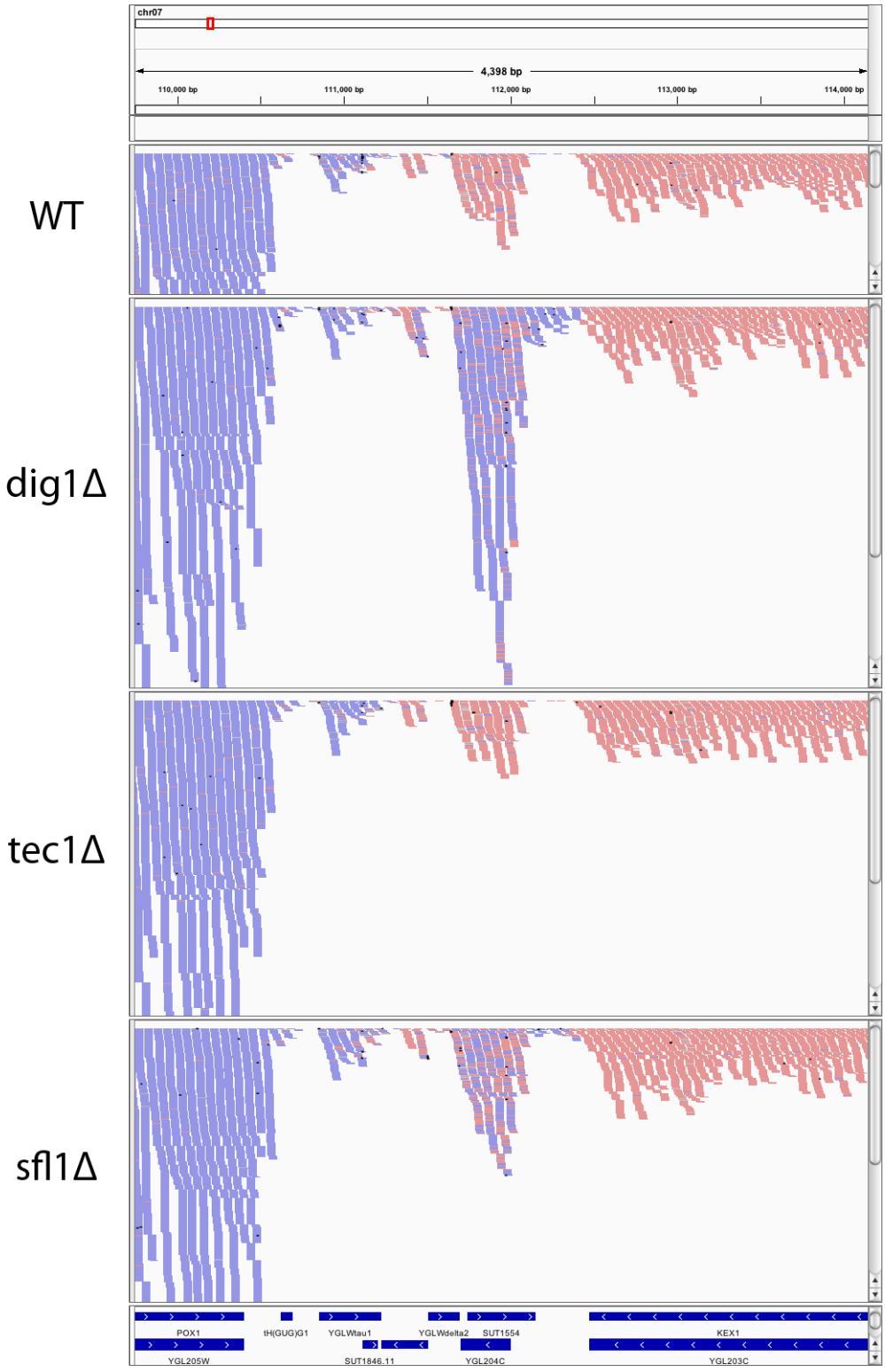


FIGURE B-6 - YGL204C (RED) AND SUT1554 (BLUE) SHOW INTERESTING PROFILES, ESPECIALLY FOR THE *DIG1* MUTANT

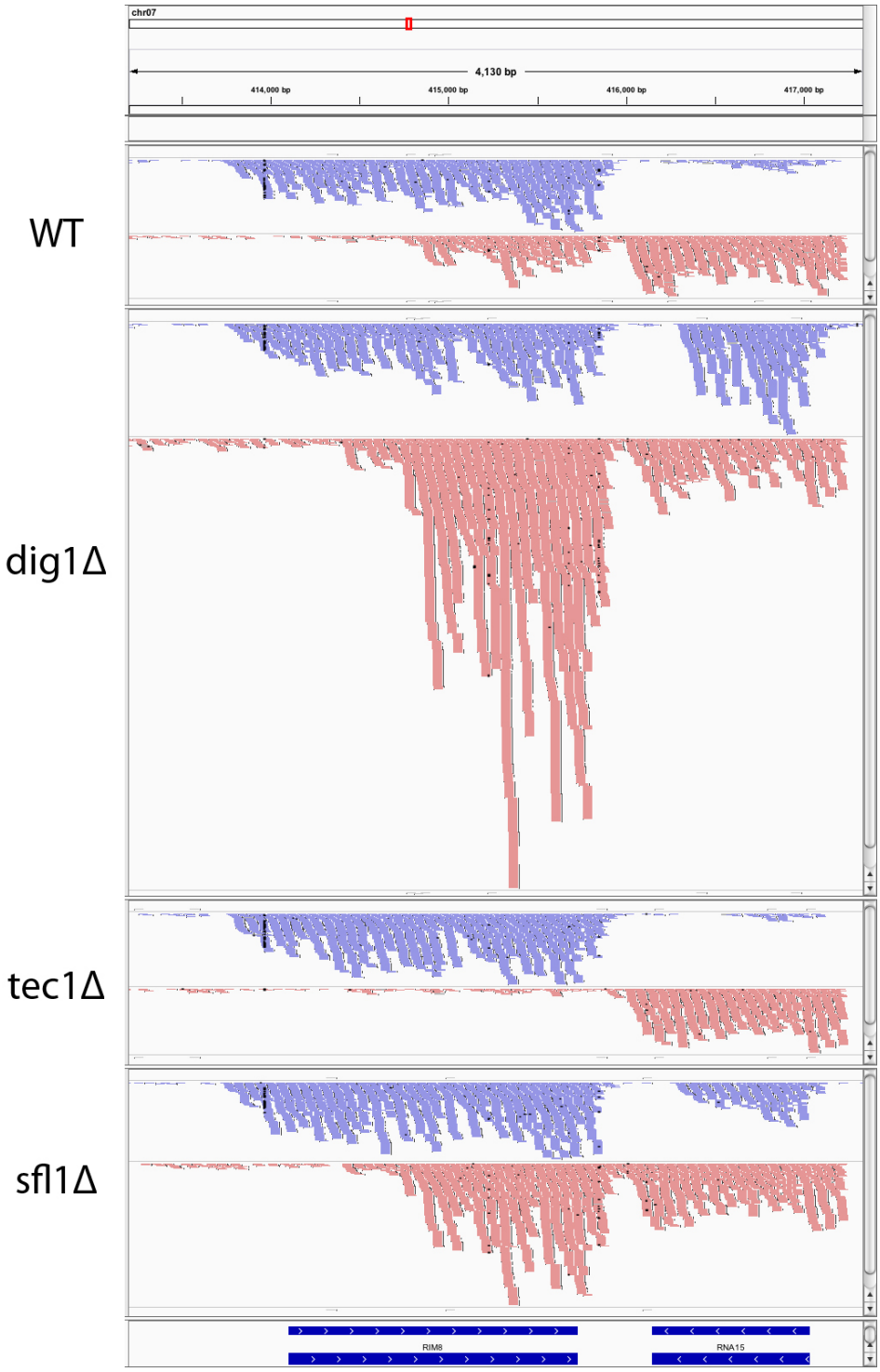


FIGURE B-7 – RIM8 (BLUE) AND RNA15 (RED) SHOW INTERESTING ANTISENSE PROFILES. TRANSCIPTS ARE SPLIT INTO WHICH STRAND THEY ORIGINATED FROM. PARTICULARLY STRIKING IS THE INDUCTION OF THE ANTISENSE TRANSCRIPTS IN THE *DIG1* MUTANT, AND THE ALMOST COMPLETE LOSS OF THEM IN THE *TEC1* MUTANT

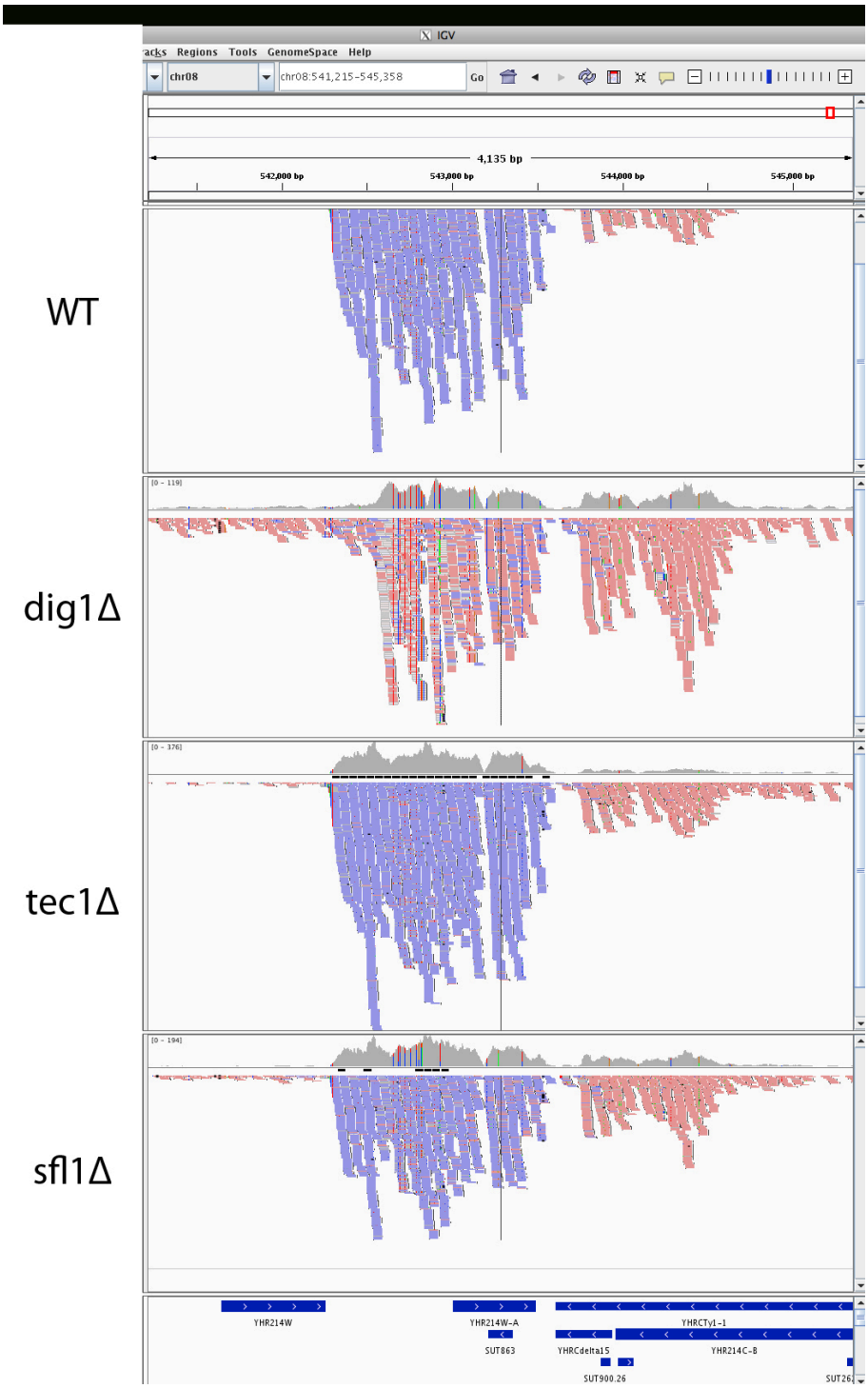


FIGURE B-8 - YHR214W-A (BLUE) AND SUT863 (RED) SHOW INTERESTING PROFILES. IT IS SURPRISING THAT THE LOSS OF THE TRANSCRIPTIONAL REPRESSOR *DIG1* LEADS TO LOSS OF EXPRESSION OF *YHR214W-A*. BUT NOTE THE HIGH LEVELS OF SNPS IN THE ANTISENSE TRANSCRIPT.

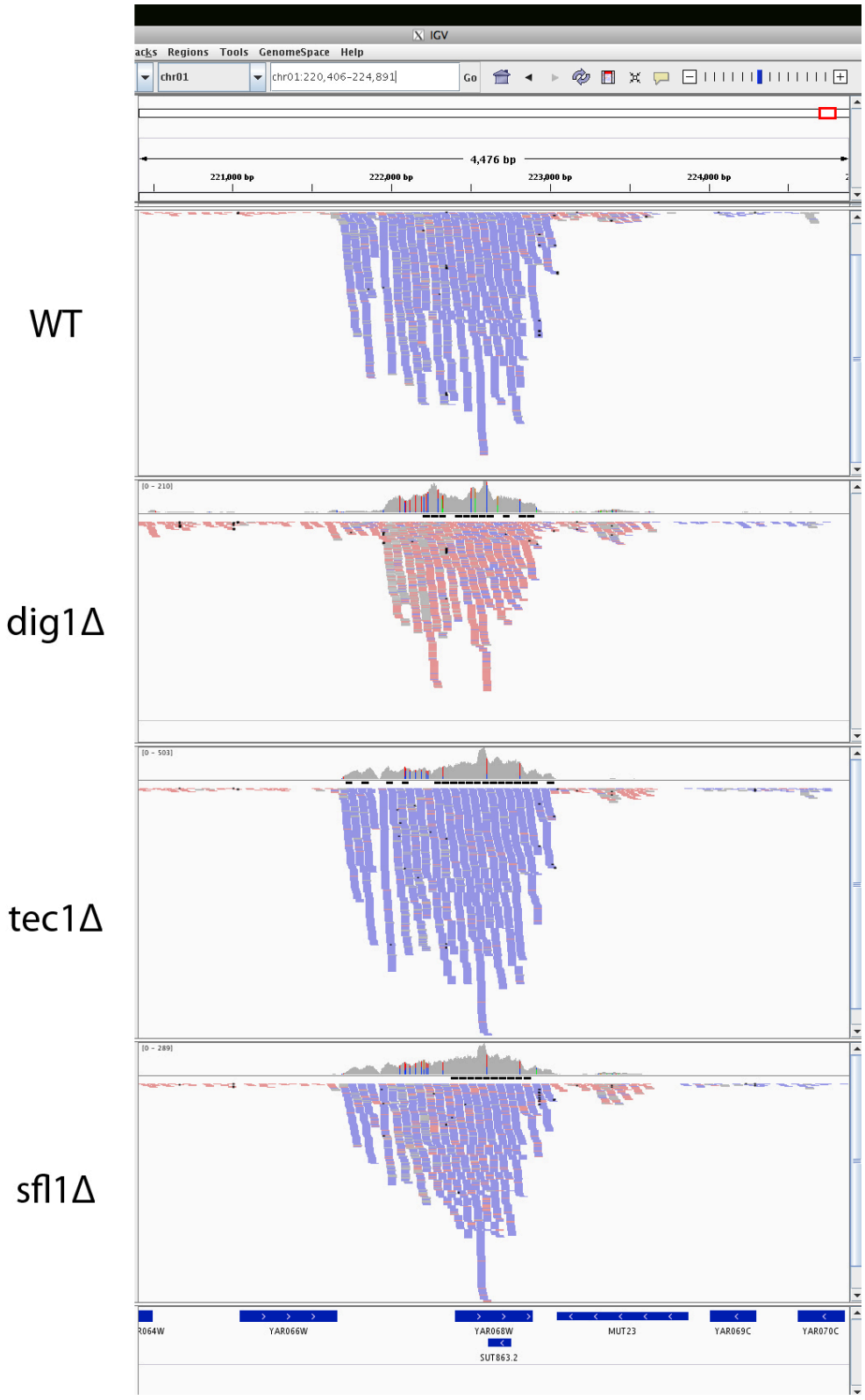


FIGURE B-9 - YAR068W (BLUE) AND SUT863.2 (RED) SHOW INTERESTING PROFIILES. THIS LOCI IS VERY SIMILAR TO THE PREVIOUS FIGURE.

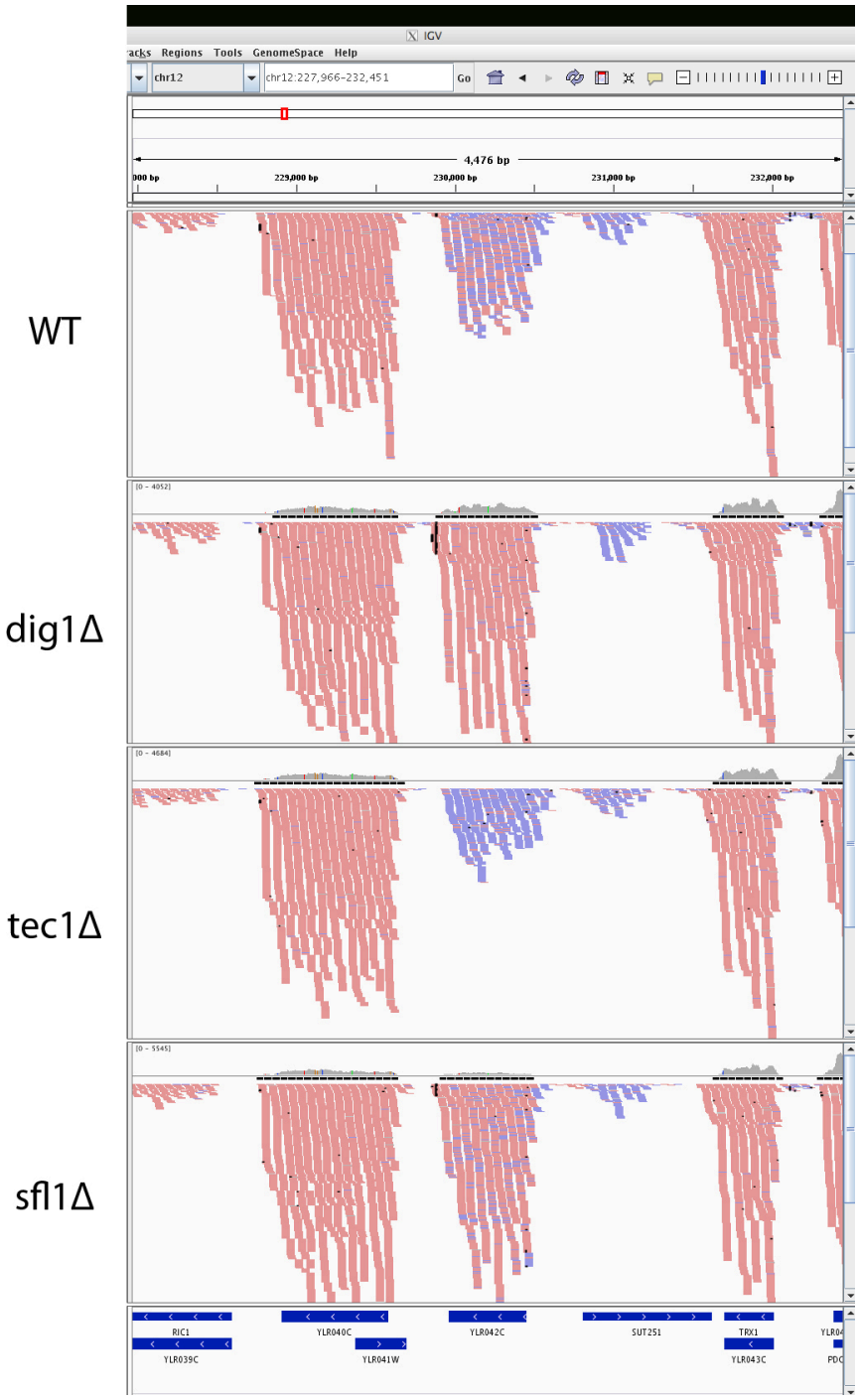


FIGURE B-10 - YLR042C (RED) SHOWS INTERESTING PROFILES. NOT ONLY IS THE ANTISENSE TRANSCRIPT UNANNOTATED, IT SEEMS TO BE COMPLETELY LOST IN A *DIG1* MUTANT, WHILE EXPRESSION OF *YLR042C* SEEMS TO BE COMPLETELY LOST IN A TEC1 MUTANT. *YLR042C* IS HIGHLY INTERESTING BECAUSE IT IS DIFFERENTIALLY EXPRESSED IN OUR STUDY IN CHAPTER 5, AND IS OFTEN DIFFERENTIALLY EXPRESSED IN CONDITIONS THAT INDUCE FILAMENTATION, THOUGH NO FUNCTION HAS BEEN ASCRIBED TO IT.

B.4 REFERENCES

1. Bumgarner, S.L., Dowell, R.D., Grisafi, P., Gifford, D.K. & Fink, G.R. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc Natl Acad Sci U S A* **106**, 18321-18326 (2009).
2. Cook, J.G., Bardwell, L., Kron, S.J. & Thorner, J. Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast Saccharomyces cerevisiae. *Genes Dev* **10**, 2831-2848 (1996).
3. Xu, Z. et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033-1037 (2009).
4. Lardenois, A. et al. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc Natl Acad Sci U S A* **108**, 1058-1063 (2011).
5. van Dijk, E.L. et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114-117 (2011).

Appendix C - QUANTITATIVE TRAIT LOCI MAPPING EFFORTS

C.1 FLUFFY IS A **COMPLEX** TRAIT

There is no question that fluffy is a complex genetic trait. Investigators that have performed systematic deletion screens have found numerous genes that are involved in the modulation of the trait[1-3]. These genes cover a broad range of functions and pathways, and the effects of their deletion also very in strength. A recent study that mapped QTLs for colony morphology between 2 non-fluffy strains that show ~2% fluffy progeny concluded that genetic interactions between multiple genes contribute to the trait, and that multiple genotypes are able to specify the same phenotype[4].

While it is obvious that any trait in which hundreds of genes can modulate it would be highly complex, another reason for this complexity could be because one of the genes that is necessary for fluffy, *FLO11*, contains one of the largest promoters (~3kb) with binding sites for numerous transcription factors[5]. These transcription factors coordinate their responses from a multitude of nutritional and environmental signals, leading to the highly complex regulation of *FLO11* expression levels. Not only that, "…two copies of the *FLO11* locus in *Saccharomyces cerevisiae* switch between a silenced and competent promoter state in a random and independent fashion"[6], making its expression heterogeneous at the single-cell level. The gene length is also unstable, with the intragenic tandem repeats present in its sequence being able to rapidly expand or contract, leading to functional phenotypic differences[7, 8]. Unsurprisingly too, there is evidence that *FLO11* is post-transcriptionally regulated[9]. As all this complexity lies within just one of the genes responsible for fluffy morphology, one can see why this would be a highly complex trait to dissect genetically.

In our own studies, we too see complex genetics at play. In a similar cross between 2 non-fluffy strains (YO486 X YO502), fluffy progeny were observed at a frequency of ~5%. If combinations of single genes led to the complete gain or loss of fluffy, this would indicate that a combination of 4-

5 genes would lead to the gain of fluffy. However, when 2 progeny that were fluffy were crossed, fluffy progeny were observed at a frequency of 37-50% (**Table C-1**, YCR12 and YCR14). F2 crosses were also created, and fluffy progeny were observed at a frequency of 43-69% (**Table C-1**, YCR21 and YCR22). This indicated much more complex genetics underlying the trait, and we are attempting to tease out the underlying genetics of this trait by performing classic QTL mapping. In order to improve the statistical power of our analysis, we decided to remake the initial cross and obtain a much larger number of fluffy progeny through random spore dissection. Also, as YO486 was aneuploid and harbored an extra chromosome I, possibly confounding the numbers, the choice was made to create a strain that had the extra chromosome and had returned to an euploid state (**Figure C-1**).

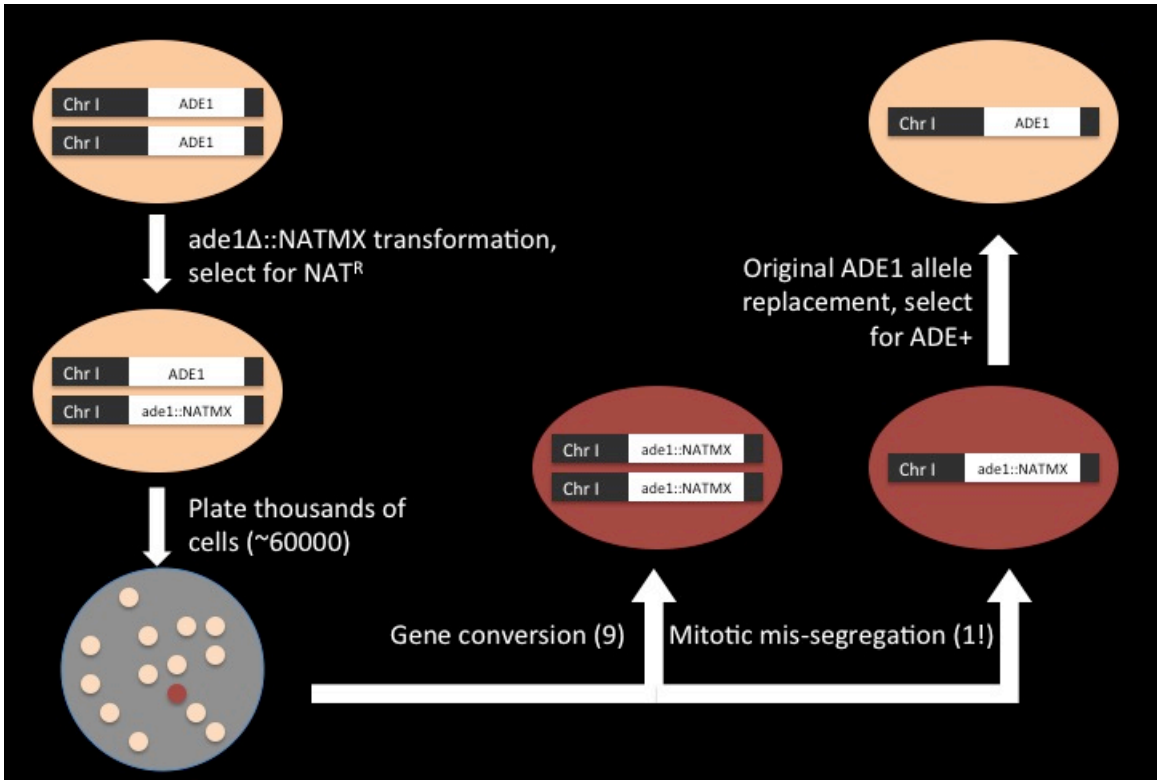


FIGURE C-1 CARTOON DIAGRAMMING THE METHODS USED TO REMOVE THE EXTRA CHROMOSOME I

After the strain was made, and having PCR verified the loss of chromosome I, the progeny were created (888 fluffy and 384 smooth) and RAD-tag sequenced. Analysis revealed the shocking discovery that while YO486 did lose the extra chromosome I, it also picked up an extra chromosome XII, thus remaining aneuploid. Something interesting to note is that the percentage of progeny that are fluffy remained quite similar at ~6%. Since we do have a very large number of progeny, an attempt will be made to dissect this trait with the aneuploidy present. This might even shed light on how aneuploidy can affect the segregation of a complex trait.

C.2 A RANDOM STAB

It has been a struggle to account for the results and percentages I have obtained with these crosses. But to take a random stab at a theory for the complexity of this trait, I feel that fluffy is a complex binary trait where a threshold character is set on an underlying continuous distribution of *FLO11* expression levels, i.e. the expression of *FLO11* is needed at a certain level, below which strains will not turn fluffy. The complex regulation and variegated expression of *FLO11* will make this underlying distribution a difficult one to dissect. However, I also think that this trait is even more complex because it will be dependent on how fast the colony expands. The higher the growth rate for a strain, the larger the colony, the more Flo11 is needed to maintain colony structure. If this is true, what this essentially translates into is a threshold that can move up and down the continuous distribution depending on the growth rate of the strain. Dissecting this trait and understanding how unique colony patterns are formed would thus require the dissection of 2 separate traits that are highly complex on their own: *FLO11* expression levels and growth rate.

C.3 MORE INFORMATION

More information on the crosses, e.g. tetrad viability, drug resistance, etc. can be found **Table C-1**, and in:

/dudley6/LabShare/Zhihao/Fluffy/Info for crosses

TABLE C-1 - LIST OF CROSSES MADE FOR THE STUDY OF COLONY MORPHOLOGY

Cross	Parent A	Parent α	Number of progeny	Notes
YCR8	YPG746 (F47)	YPG453 (F13)	28	Pilot F1 inter-cross between progeny of Vida's cross; Both parents are fluffy
YCR9	YPG746 (F47)	YPG586 (F29)	28	Pilot F1 inter-cross between progeny of Vida's cross; Both parents are fluffy
YCR12	YPG746 (F47)	YPG453 (F13)	144	F1 inter-cross between progeny of Vida's cross; Both parents are fluffy; Progeny are created through manual tetrad dissection; 72/144 (50%) progeny are fluffy; Non-viable progeny were given YPG numbers, sequenced plates have non-continuous YPG numbers
YCR14	YPG746 (F47)	YPG453 (F13)	204	F1 inter-cross between progeny of Vida's cross; Both parents are fluffy; Progeny are created through manual tetrad dissection; 76/204 (37.3%) progeny are fluffy; Non-viable progeny were given YPG numbers, sequenced plates have non-continuous YPG

				numbers
YCR21	YPG3783	YPG2077	35	F2 inter-cross (F13 X F47) between progeny of Vida's cross; Both parents are fluffy; Progeny are created through manual tetrad dissection; 15/35 (42.9%) progeny are fluffy; Non-viable progeny were given YPG numbers, sequenced plates have non-continuous YPG numbers
-	YPG3783	YPG2147	61	F2 inter-cross (F13 X F47) between progeny of Vida's cross; Both parents are fluffy; Progeny are created through manual tetrad dissection; 42/61 (68.9%) progeny are fluffy; Non-viable progeny were given YPG numbers
YCR23	YO1862	YO502	960	Re-made Vida cross; YO1862 is aneuploid for chromosome XII; Both parents are non-fluffy; Progeny are created through random spore dissection, strains were hand-picked for fluffy phenotype; ~6% progeny are fluffy; Rearrayed into YCR25 after diploids were removed following drug testing
YCR24	YO1862	YO502	480	Re-made Vida cross; YO1862 is aneuploid for chromosome XII;

				Both parents are non-fluffy; Progeny are created through random spore dissection, strains were hand-picked for smooth phenotype; ~6% progeny are fluffy; Rearrayed into YCR26 after diploids were removed following drug testing
YCR25	YO1862	YO502	888	Rearrayed from YCR23; Plates that are actually sequenced; A new set of YPG numbers were given to the strains in this format, therefore YPG numbers are in running order
YCR26	YO1862	YO502	384	Rearrayed from YCR24; Plates that are actually sequenced; A new set of YPG numbers were given to the strains in this format, therefore YPG numbers are in running order

Images of the development of the tetrads of each cross can be found in:

/dudley2/imaging/Zhihao/Vida Cross

/dudley2/imaging/Zhihao/Mating/progeny

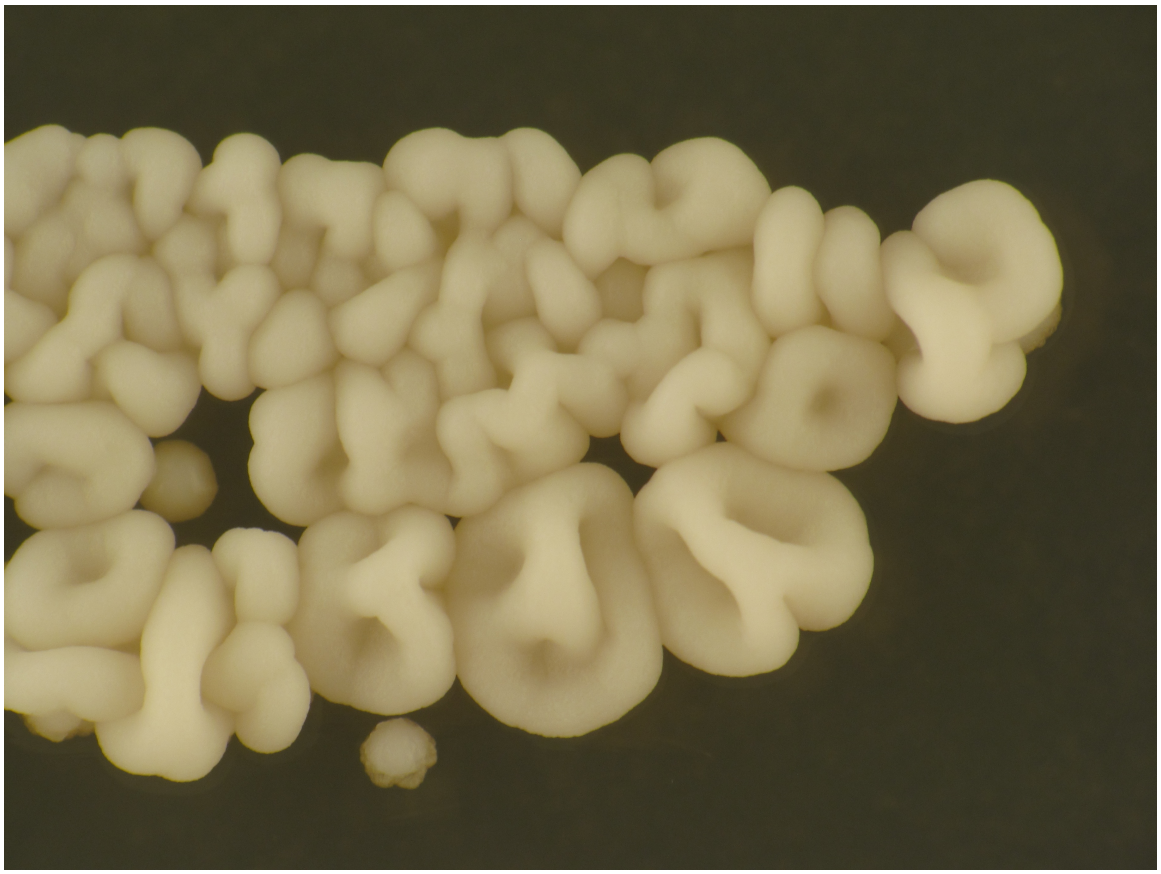
C.4 REFERENCES

1. Granek, J.A. & Magwene, P.M. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823 (2010).
2. Voordeckers, K. et al. Identification of a complex genetic network underlying Saccharomyces cerevisiae colony morphology. *Mol Microbiol* **86**, 225-239 (2012).
3. Ryan, O. et al. Global gene deletion analysis exploring yeast filamentous growth. *Science* **337**, 1353-1356 (2012).
4. Taylor, M.B. & Ehrenreich, I.M. Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet* **10**, e1004324 (2014).
5. Rupp, S., Summers, E., Lo, H.J., Madhani, H. & Fink, G. MAP kinase and cAMP filamentation signaling pathways converge on the unusually large promoter of the yeast FLO11 gene. *Embo J* **18**, 1257-1269 (1999).
6. Octavio, L.M., Gedeon, K. & Maheshri, N. Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet* **5**, e1000673 (2009).
7. Verstrepen, K.J., Jansen, A., Lewitter, F. & Fink, G.R. Intragenic tandem repeats generate functional variability. *Nat Genet* **37**, 986-990 (2005).
8. Zara, G., Zara, S., Pinna, C., Marceddu, S. & Budroni, M. FLO11 gene length and transcriptional level affect biofilm-forming ability of wild flor strains of Saccharomyces cerevisiae. *Microbiology* **155**, 3838-3846 (2009).
9. Fischer, C. et al. Posttranscriptional regulation of FLO11 upon amino acid starvation in Saccharomyces cerevisiae. *FEMS Yeast Res* **8**, 225-236 (2008).

Appendix D - IT'S SO FLUFFY!

"It's so fluffy I'm gonna die!"

-Agnes, *Despicable Me* [1]



[1] https://www.youtube.com/watch?v=82utG7Q3G_k