Zooming in on the Actin gene:
Genetic interactions of diploid and haploid yeast.

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Stephanie DiPrima

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Chad Myers

March 2013

© Stephanie DiPrima 2013

Acknowledgements

I would like to thank everyone who assisted and supported me on this project, especially Dr. Chad Myers, who inspired me in a single class to join the field of bioinformatics and who has guided me throughout this journey. The other members of my committee, Dr. Judy Berman and Dr. Rui Kuang have been very flexible, encouraging, and patient throughout my work on this project. All the members of the Myers Lab have been incredibly helpful with getting me up to speed on projects and giving me guidance on many aspects of my thesis. I owe a particularly big thank you to Ben, Elizabeth, Raamesh, and Carles for all your guidance and work with me over the years. This project would not have been possible without all of the hard work from our collaborators in Dave Amberg's lab, especially Dr. Dave Amberg and Brian Haarer. I'd also like to thank the Boone lab for their contribution to the data as well, particularly Michael Costanzo, Anastasia Baryshnikova, Charlie Boone, and Brenda Andrews. Lastly, I am very grateful to all of friends and family for being my support and cheer team throughout this process and getting me to where I am today.

Dedication

To Patrick, for everything.

Abstract

Actin is one of approximately 1200 essential genes in yeast. It is a structural protein involved in a number of functions throughout the cell, including cell growth, division, motility, exocytosis, and many other important processes. However, since actin is an essential gene, the standard method of knocking out the gene in a haploid yeast to study mutant phenotypes or genetic interactions is not feasible. Instead, we used diploid yeast with one working copy of actin to study the genetic interactions within the cell. With the use of diploids with point mutations in actin, we can begin to look more specifically at the profile of the actin surface to greater understand its structure, function, and interactions as well as study the relationship between the abundant haploid data with the new diploid data. In collaboration with a yeast genetics lab, we studied complex haploinsufficient interactions between 32 actin point mutants and the genes across the yeast genome. We found that different alleles exhibit different genetic interactions, but that both the degree and the identity of the interaction partners are similar for alleles that affect amino acids near each other on the surface of actin. We also compared interactions between the actin mutants in diploid and haploid cells and found important differences in diploid interactions, suggesting diploid mutant genetic screens will be important for understanding complex genetics.

Table of Contents

Acknowledgements…………………………………………………………………..……. i

Dedication...………………………………………………………………..……….ii

Abstract……………………………………………………………………..………iii

List of Tables…………………………………………………………………….……vi

List of Figures…………………………………………………………………………..vii

Chapter 1: Introduction and Background

1.1 Why choose yeast for our study?...1

1.2 Yeast, a genetically simple eukaryote………………………….……….. 2

1.3 Actin, an interesting essential protein ..………………………….……… 3

1.4 Previous work on complex haploinsufficiency with actin ……………..... 6

Chapter 2: Studying Diploid Yeast with Actin Point Mutants

2.1 Mutant actin CHI experiment background………………………………12

2.2 Actin mutations represent locations along the surface of the molecule…………………………………………………………………….14

2.3 Mutant actin alleles exhibit high variation in the number of interactions…………………………………………………………………17

2.4 Comparisons of broad locations show hot spots on the front/top/bottom…………………………………………………………18

2.5 Hierarchical clustering of allele profiles relates to protein structure………………………………………………………………………20

2.6 Stacked distance measurements show correlation with interaction profile……………………………………………………………………….23

2.7 Conclusions……………………..……………………………………………26

Chapter 3: Comparison of Actin Genetic Interactions Between Haploids and Diploids

3.1 Advances in haploid technology have led to expanding haploid data……………..…………………………………………...……27

3.2 Creating a haploid set to study against the diploid set……………..…28

3.3 Haploid interactions show a broad enrichment for CHI genes……….30

3.4 Interaction profile comparisons reveal differences between haploid and diploid interactions…………………………….…………33

3.5 While haploids can tell us which genes, we need data from diploids to learn how the interactions occur ……..………..……...…..36

Bibliography ……………………………………………………………………39

List of Tables

Table 2.2.1 Allele Location Table …………………………………………………16

Table 3.3.1 Allele Specific Enrichment of CHI genes in Haploid Interactions 32

Table 3.4.1 Allele Specific Overlap of Profile Interaction Pairs ………………… 36

List of Figures

Figure 1.3.1 Actin Monomer and Filament Structure ……………………………4

Figure 1.4.1 Schematic of Genetic Arrangement in Yeast for
Null Actin Study ……………………………………………………8

Figure 1.4.2 Genetic Interaction Map of Genes that are CHI with act1 …………10

Figure 2.1.1 Schematic of Haploid Genetic Cross with Point Mutant Actin …….14

Figure 2.3.1 Histogram of the Number of Interactions for each Allele …………..17

Figure 2.3.2 Heat Map of Interactions on the Surface of Actin …………………..18

Figure 2.4.1 Mean Number of Interactions On Surface Areas of the
Actin Protein ……………………………………………………20

Figure 2.5.1 Clustergram of Interaction Profiles Showing Allele Clustering ……22

Figure 2.5.2 Sample Cluster Locations on Actin's Surface ………………………23

Figure 2.6.1 Stacked and Backed Configurations of Actin Monomers …………24

Figure 2.6.2 Scatter Plot of Profile Correlation and Stacked
Distance Measure ………………………………………………25

Figure 3.2.1 Example Construct of Haploid Double Mutant ……………………...28

Figure 3.2.2 Schematic of Haploid and Diploid Overlap …………………………30

Figure 3.3.1 Diagram of Broad Enrichment Data …………………………………31

Figure 3.4.1 Diagram of Local Specific Overlap …………………………………...34

Figure 3.4.2 Distribution of Specific Interaction Pair Overlap in
the Global Subset ………………………………………………35

Figure 3.5.1 Visualization of act1-112 ……………………………………………37

CHAPTER 1: INTRODUCTION AND BACKGROUND STUDY

1.1 Why choose yeast for our study?

The common baker's or brewer's yeast (*Saccharomyces cerevisiae*) is a member of the budding yeast family, all members of which reproduce by means of budding a new daughter cell from the cell wall of the mother cell (Yeong 2005). This yeast is a very scientifically interesting organism and is, in fact, the cornerstone of many biological studies. There are many reasons that geneticists choose yeast as an experimental organism. For one, yeasts are single celled organisms, and do not have the complexity of higher organisms. Their small size makes them very easy to house and grow large quantities while using very little space. They are also relatively inexpensive to feed and maintain. Yeast are among the simplest eukaryotes in terms of their genetics. In fact, in 1996 the yeast genome was the first eukaryotic genome to be sequenced (Goffeau et al. 1996). Many computational and experimental tools have been developed to help manipulate and study the genes within that genome with relative ease. The combination of the yeast's simplicity and the many tools available has paved the way for the creation of high throughput technology to study yeast. Because yeast are eukaryotes, like humans, many of their molecular systems may be similar to those in humans, and studies done in yeast can help us learn about our own biology. In fact, yeast and humans have nearly 2000 genes with significant homology (Botstein et al. 1997). Another incentive to use yeast is that it has proven itself to be indispensable in a number of industries, alcohol and bread

(Legras et al. 2007), as well as a number of biotechnology industries are just a few examples. Thus, studying the genetics of these organisms can help improve the efficiency of their use in industry while furthering our fundamental understanding of biology. By studying a gene or protein in yeast, we may be able to obtain essential information to help us to better understand human genetics in a relatively simple, cheap, and safe environment.

1.2 Yeast, a genetically simple eukaryote

Saccharomyces cerevisiae has 16 chromosomes, which contain just over 6000 genes (Goffeau et al. 1996). In comparison, humans have 23 chromosomes with approximately 20,000 genes (International Human Genome Sequencing Consortium 2004). In many organisms, including yeast, we have developed methods to modify the genome in order to “delete” a specific gene of interest. In creating these so-called deletion mutants, we can try to determine the function of the gene based on how the organism is affected when we take that particular gene away (Dixon et al. 2009). An example of this is the *white* gene in *Drosophila melanogastor* (fruit flies). A wild type fruit fly has bright red eyes. However, when the *white* gene is not present in the genome, the eyes are white. While this does not imply that the *white* gene directly encodes the red pigment, it does tell us that the *white* gene is somehow related to the pathway responsible for the red pigment in normal flies (MacKenzie et al. 1999). In yeast, geneticists have created deletion mutants for each of the 6000 known genes and

have found that approximately 20% of these genes are essential to life because without any one of those 20% the yeast is not able to survive (Giaever et al. 2002).

Another attribute of yeast that makes it genetically interesting is that it will readily exist in either a haploid or a diploid form. In its haploid form, yeast has only one copy of every chromosome, whereas diploid organisms have two. This is an advantageous state in which to study yeast because creating the series of null mutations becomes much simpler if one only needs to work with a single copy of a gene. This makes it much easier when performing things like double or triple knockouts, that is deleting two or three genes at a time, to determine if the genes have an effect on each other, because there are fewer functional copies of each gene that need to be deleted in order to study the phenotypic effects of a null mutation. If two haploid yeast mate, they join together to form one diploid yeast. This case is biologically interesting because many higher eukaryotes live mainly as diploids, humans included. While a haploid genome may be simpler, it may not necessarily be representative of how genes or genetic variants behave when they are in the context of a diploid organism. This is one point I explore in this thesis while examining the behavior of one specific gene, *act1*, the gene which codes for the protein actin.

1.3 Actin, an interesting essential protein

Actin is a structural molecule within the cell – when in its filamentous form

it can be considered similar to the rebar structure within a building because it forms the long rods that hold the shape of the cell around it. Actin functions in both a monomeric and filamentous form as well as other far less common configurations (Dominiguez and Holmes 2011). The figure below shows a single monomer as well as an actin filament. The filament is formed by backing together two stacked strands of actin against each other, one of which is shown darker and the other lighter (Figure 1.3.1). The pool of actin molecules is constantly fluctuating between the two major states. The fluid assembly and disassembly of the filaments in the cytoskeleton gives the cell the ability to perform a number of tasks necessary to life. For example, this property is required for the cytoskeleton remodeling that occurs in cell replication (Dominiguez and Holmes 2011).

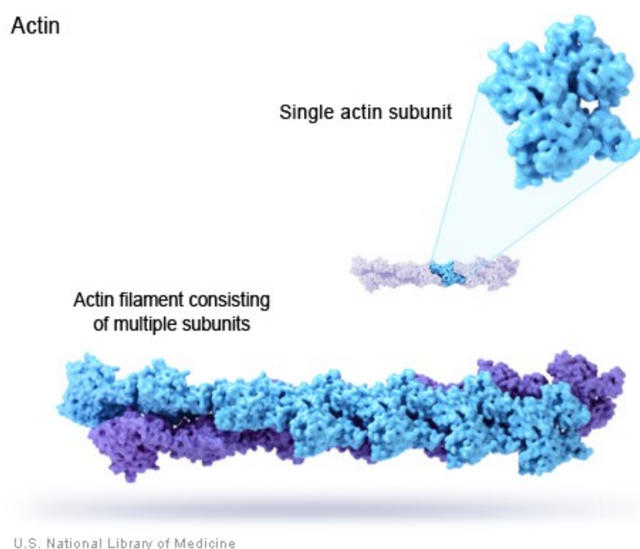

Figure 1.3.1 Actin Monomer and Filament Structure
This figure shows the arrangement of actin monomers in a filament form. The actin monomers form two strands by stacking and then twist together to form the final product. The dark and light represent the two strands. (U.S. National Library of Medicine, 2013) http://ghr.nlm.nih.gov/handbook/illustrations/actin

Actin is an important player in functions all over the cell. It is well known as a key component in structural roles such as cell motility and maintenance of cell shape and polarity (Dominguez and Holmes 2011). However, it also is very important in transcription regulation. Actin proteins have been found in a number of nuclear complexes that have implicated its involvement not only in transcription but also in chromatin remodeling and nucleocytoplasmic trafficking as well (Bettinger et al. 2004). With all of its functions, it is not surprising that actin is the most abundant protein in most eukaryotic cells and that is has the more protein-protein interactions than any other known protein (Dominguez and Holmes 2011).

The actin cytoskeleton has shown to be quite sensitive to perturbations in the sequence (Wertman et al. 1992). Because of this, actin is a highly conserved gene from species to species. Humans and yeast share a ~80% sequence similarity in actin gene sequence and ~95% protein structure similarity (Vandekerckhove and Weber 1978). In fact, the actin amino acid sequences between any two eukaryotic sequences are all generally ~90% identical. One theory for why it is so highly conserved stems from its high level of functionality within the cell. The surface of actin interacts with so many other proteins within the cell that one small difference can lead to many primary and secondary effects on function. In the process of evolution it is easier to change the binding partner than to change actin (Alberts et al. 2002).

When two different genes are mutated, a genetic interaction is defined to

be the case in which the effect of both mutations is different than what we would expect based on phenotypes resulting when each gene is deleted individually. Genetic interactions can also tell us a lot about functions within a cell. By looking at what other genes actin interacts with, we can learn about its function and role in cellular processes. Large sets of interaction data also allow for creations of genetic networks that can be clustered to create functional maps of the cell. Functional maps have been used to infer function of previously uncharacterized nonessential genes in the genome, simply by comparing its interactions to the interactions of other genes (Costanzo et al. 2010).

1.4 Previous work on complex haploinsufficiency with actin

Since actin is an essential gene, we must study it from a different angle than non-essential genes, because true null alleles are impossible to study in haploids. Because of this, we aim to study its genetic interactions in diploid yeast. In diploid yeast with one null copy and one fully functional copy of actin, we see haploinsufficiency, which means that diploid cells with only one functional actin allele exhibit a significant growth defect. This is specific to diploids and cannot be completely recreated in haploids. When studying interactions with actin, this property must be kept in mind.

Instead of looking at heterozygous null actin, we can examine a number of point mutations. By looking at point mutations on actin, we are able to look for interactions with other genes that are affected by the part of the molecule that is

mutated. However, to do this study we would need to do a combination of each point mutant by all genes to find which genes interact. The caveat of going the "point mutant by all genes" route is that even with just 30 point mutations, you would need construct over 190,000 yeast strains to create double mutant combinations with every other gene, a majority of which would not be interesting.

To narrow down the list of potential interesting double mutants, Haarer et al. first performed an experiment with diploid yeast strains that were heterozygous null for actin and heterozygous null for one other gene. From this, they created a list of genes that had complex haploinsufficiencies with actin, meaning that they had a secondary genetic effect caused by the decrease in actin in addition to the haploinsufficiency effect itself. By doing so, they were able to select a majority of the genes that they would expect to show interactions with any of the point mutants.

To build these diploid yeast strains, they mated a haploid which was wild type for actin and null for some "GeneX" (see A in Figure 1.4.1) with another haploid which was wild type for that "GeneX" and "null" for the actin gene (See B in Figure 1.4.1. This "null" allele was created by introducing a low copy number plasmid into the yeast that would keep the yeast without an actin gene alive long enough to mate. Then the plasmid would be removed to ensure the plasmid actin was not interfering with the results. The cross can be seen below in Figure 1.4.1. This genetic arrangement was tested on ~4800 heterozygotes to look for possible interactions between the mutant and the missing gene.

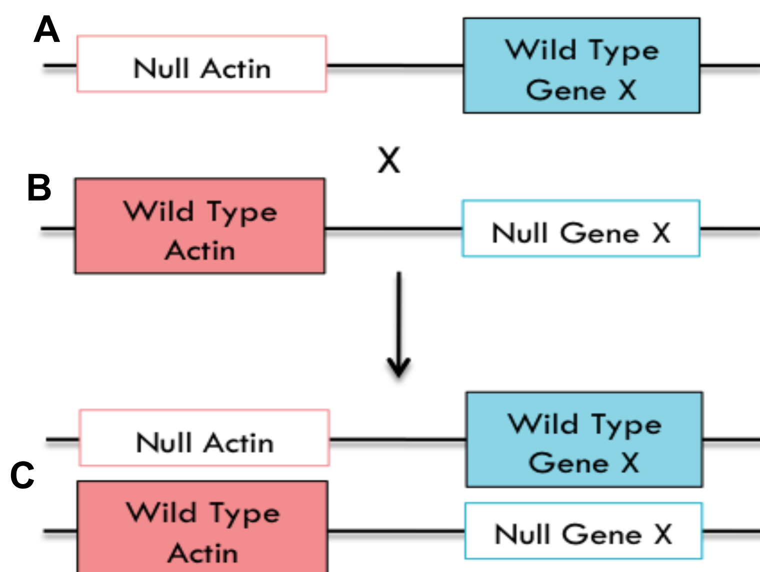

Figure 1.4.1 Schematic of Genetic Arrangement in Yeast for Null Actin Study
This diagram shows constructs of haploid yeast that are mated to form the diploid double mutants for the CHI study. A represents the haploid that has the null gene X and a wild type actin. B represents the haploid that has a wild type gene X and a null actin. This yeast also has an additional extra chromosomal plasmid that holds an actin gene that is removed after mating to all for the haploid to persist without the essential gene. C represents the diploid after mating A and B that is now heterozygous null for gene X and actin. While the genes in this diagram are shown on the same chromosome, this is not always the case for each pair.

The resulting diploid yeast strains (See C in Figure 1.4.1) were not only haploinsufficient for the level of actin, but in some cases, they also showed fitness deficiencies through interactions between actin and "GeneX." Fitness in this experiment was measured by colony size. When paired together, a strain that had an interaction would appear sicker than expected based on the sicknesses of the two individually. The interaction with haploinsufficient actin is called a complex haploinsufficiency (CHI) interaction. In this study, Haarer et al. identified 242 nonessential genes that showed CHI interactions with heterozygous null actin. Later, five of these genes were removed from this set

because it was found that the interactions were results of other unrelated factors. The remaining 237 genes span across a wide variety of cellular functions which reflects the many roles that actin plays. The majority of the genes' major functions fall into cellular organization and biogenesis, which is expected given that actin is a structural molecule. Figure 1.4.2 shows a genetic network of these 237 genes, which we will call CHI genes, each of which is color coded by function (Haarer et al. 2007).

The set of CHI genes included a number of known actin binding proteins and many of them also show deficiencies in actin organization when they are deleted. Haarer et al. also saw that when a member of a protein complex is within the CHI set, often a number of the other complex members were in the CHI set as well. In a number of cases, the interactions were explainable by localized defects on the actin molecule, for example, act1-105 and act 1-111 which are near each other on the surface shared a large number of their interactions.

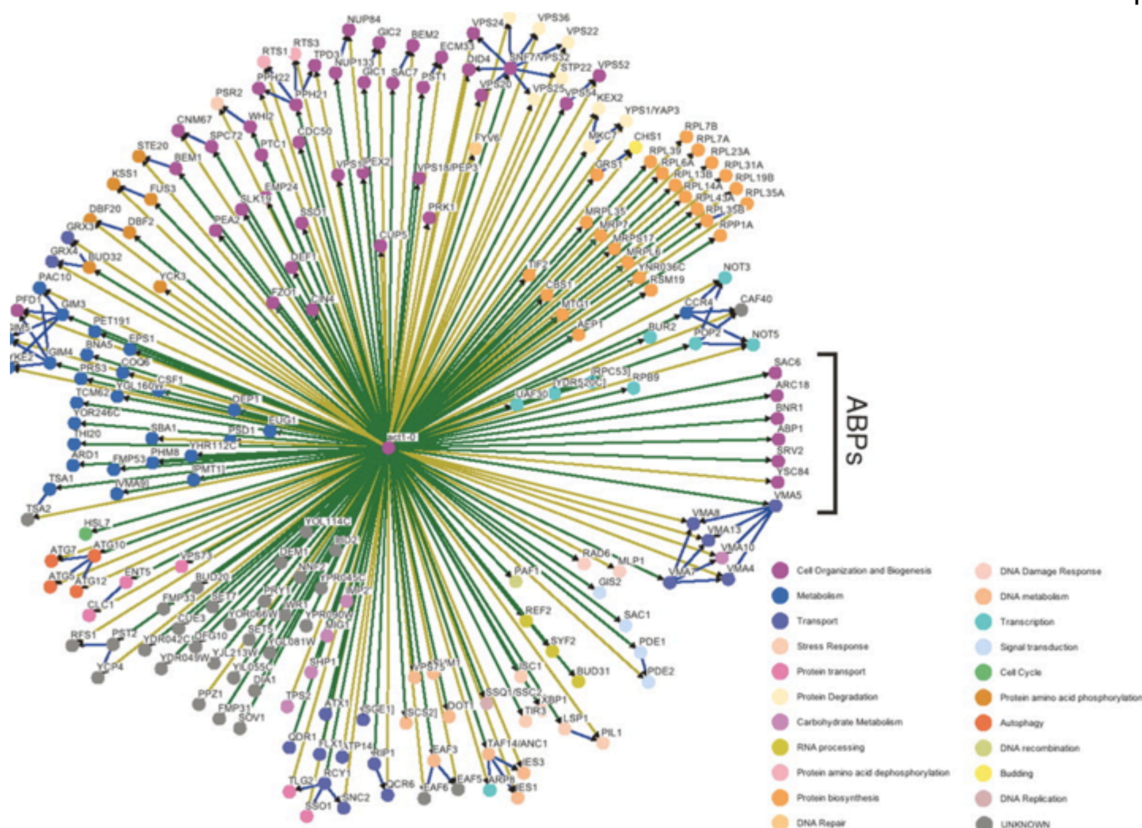

Figure 1.4.2 Genetic Interaction Map of Genes that are CHI with act1
A genetic interaction map is shown which shows act1 in the center surrounded by the 242 genes that showed CHI interactions. The outer genes are clustered by functions and coded in the legend on the right. (Haarer et al. 2007)

One interesting case that the Haarer et al. found was an actin-related functional difference of two ribosomal paralog pairs. Paralogs are highly similar genes that result from an ancient duplication event and are generally considered to be functionally similar to each other. However, in each of these cases only one of the two paralogs in each of the pairs showed a CHI interaction.

For a long time it was expected that the haploinsufficiency phenotypes seen in yeast that are heterozygous null for actin was due to the reduced level of actin not making as much filament as is necessary to sustain normal function.

Instead, they found that the haploinsufficiency effects actually come from having too low a concentration of monomeric actin. The majority of the available actin is shifted to the creation of excess filaments instead of what was originally expected (Haarer et al. 2007).

As a part of the experiments for their 2007 paper, the Amberg lab also conducted a pilot study that serves as a basis for this research. They operated under the hypothesis that some of the CHI interactions may not be the result of the reduced level of actin, but instead the result of the loss of certain actin functions, such as binding capacity, within the cell. They determined that the use of mutated actin sequences in place of the null sequences would allow them to see how the different locations on the surface interact with other genes in the cell. By studying six different mutants they were able to see many interesting results, such as correlation of severity of the fitness defect with number of interactions. They also saw that some of the interactions were localized to certain areas on the actin molecule. These results encouraged them to expand the study with the hope of learning more about actin interactions.

CHAPTER 2: Studying Diploid Yeast with Actin Point Mutants

2.1 Mutant actin CHI experiment background

While studying pairs of heterozygous null mutations or null mutations in a haploid genome can give interesting functional insights, it is not directly relevant to understanding the genetics of human disease. The 1000 Genomes Project Consortium found that on average, humans have approximately 100 true null alleles in protein coding genes (MacArthur et al. 2012). With over 20,000 genes in the human genome (National Human Genome Research Institute 2011), the chances that 2 of those 100 null alleles are in the two different alleles of the same gene is very unlikely. Furthermore, for essential genes like actin, homozygous null mutations would result in inviability and thus, individuals with these combinations would not exist in the population. Even having two heterozygous null mutations with a significant interaction in a heterozygous state is likely to be rare. However, single nucleotide polymorphisms (SNPs), changes of a single nucleotide in the coding sequence, which may or may not affect the protein, are not altogether uncommon. In humans, SNP's occur at a rate of about one in 100-300 base pairs (U.S. Department of Energy Genome Program 2008). Given that rate and the fact that the human genome is approximately 3 billion base pairs long (National Human Genome Research Institute 2011), there are a high number of gene A null – gene B SNP combinations within the human genome, many of which may lead to genetic interactions. These combinations have shown to be very important in studying human genetic diseases.

Geneticists have found that very few human diseases are truly Mendelian – caused by a mutation or loss of a single gene. There are, more often than not, many genes contributing to a single disease phenotype (Altshuler et al. 2008). The specific combination of mutated genes and the severity of mutations can both greatly affect the course of the disease (Ralston 2008).

By using strains with point mutations in actin and null deletions elsewhere, we hope to model those circumstances, which are arguably more realistic in human genetics than two heterozygous null genes. Our collaborators, Dave Amberg et al., have created an experiment that models a simple point mutation in one copy of actin and a deletion mutation in a second gene. These other genes that were chosen are the 237 genes that were shown to have interactions with the null actin allele in their previous study, greatly narrowing down the possible pairs and scope. By using yeast strains with point mutations, which have a much higher fitness level than the heterozygous null actin yeast, we can study a wider range of fitness and explore more specific interactions and their relationship to the structure of the actin molecule. When compared with wild type, we can look at the effect of a given point mutant when mutated actin is present at a 50% level with respect to another heterozygous null mutation of one of the 237 genes.

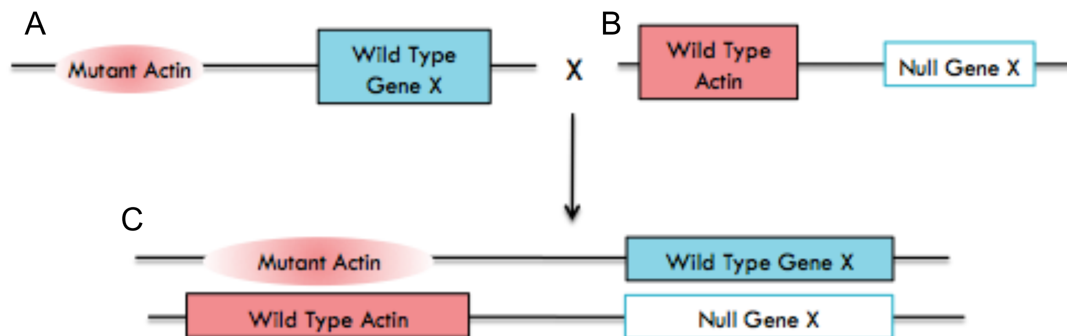

Figure 2.1.1 Schematic of Haploid Genetic Cross with Point Mutant Actin
This diagram shows constructs of haploid yeast that are mated to form the diploid double mutants for the CHI study. (A) represents the haploid that has the wild type gene X and a mutated actin. (B) represents the haploid that has a wild type actin and a null gene X. (C) represents the diploid after mating (A) and (B) that is now heterozygous null for gene X and heterozygous mutant actin. While the genes in this diagram are shown on the same chromosome, this is not always the case for each pair.

The layout is similar to that of the previous experiment except that, as you can see in Figure 2.1.1 above, a mutant actin has been substituted for the null actin.

Another motivation for this experiment was to gain information about protein binding partners of actin based on the allele specificity with each of the CHI genes. By looking at which of the genes interact with which alleles, we hope to build a meaningful residue-level mapping of genetic and physical interactions across the surface of actin.

2.2 Actin mutations represent locations along the surface of the molecule

The Amberg lab chose 32 different point mutant actin alleles for the CHI screen, 31 of which are on the surface of the protein. The remaining one, act1-159, lies in the ATP-binding site and is thought to be involved in maintaining filament stability (Haarer et al. 2007). The point mutants were chosen from a

collection of alanine-scan mutants made by Wertman et al. (Wertman et al. 1992). The Amberg lab verified and reconstructed the mutations in the study to work within the CHI construct and corrected those that were found to have additional mutations in preparation for the CHI study (Haarer et al. 2010). Table 2.2.1 shows a list of the 32 mutations, the amino acid substitution, and the position of the mutations on the actin molecule. This set of mutations was chosen because of their locations on the exterior of the molecule; this allows us to specifically study the genes that have interactions with the surface of actin. They were also chosen because the overall integrity of the molecule was not affected too strongly by the mutations, making them less detrimental to the yeast.

Table 2.2.1 Allele Location Table.
This table gives the list of alleles studied in the diploid CHI experiments. It also gives the residue that are mutated, the corresponding amino acid replacement, and the location call that was assigned to it. Act1-136 does not have a location call because the residue it mutates is not available on the pdb structure for imaging. Act1-159 also does not have a location call because it is inside of the inside the binding domain and is not on the surface.

Actin Allele	Residue	Replacement	Location Call
act1-101	363-364	D363A, E364A	Side
act1-102	359, 361	K359A, E361A	Side
act1-103	334-336	E334A, R335A, K336A	Front Face
act1-104	315-316	K315A, E316A	Side
act1-105	311-312	E311A, R312A	Front Face
act1-106	290-292	R290A, K291A, E292A	Side
act1-107	286, 288	D286A, D288A	Top/Bottom
act1-108	256, 259	R256A, E259A	Back Face
act1-109	253-254	E253A, R254A	Front Face
act1-110	237-238	E237A, K238A	Top/Bottom
act1-111	222, 224, 226	D222A, E224A, E226A	Side
act1-112	213-215	K213A, E214A, K215A	Front Face
act1-113	210-211	R210A, D211A	Front Face
act1-115	195-196	E195A, R196A	Top/Bottom
act1-116	187, 191	D187A, K191A	Back Face
act1-117	183-184	R183A, D184A	Back Face
act1-119	116-118	R116A, E17A, K118A	Back Face
act1-120	99-100	E99A, E100A	Side
act1-121	83-84	E83A, K84A	Side
act1-122	80-81	D80A, D81A	Side
act1-123	68, 72	R68A, E72A	Back Face
act1-124	56-57	D56A, E57A	Front Face
act1-125	50-51	K50A, D51A	Side
act1-127	270, 275	E270A, D275A	Back Face
act1-128	241, 244	E241A, D244A	Top/Bottom
act1-129	177, 179	R177A, D179A	Back Face
act1-131	61-62	K61A, R62A	Top/Bottom
act1-132	37-39	R37A, R39A	Back Face
act1-133	24-25	D24A, D25A	Front Face
act1-135	4	E4A	Front Face
act1-136	2	D2A	-
act1-159	159	-	-

2.3 Mutant actin alleles exhibit high variation in the number of interactions

Each double mutant yeast strain was constructed, and colony size was manually inspected to derive a fitness score. These values were then compared to the fitness of each of the single mutants and the differences were given scores between 1 and 3, where 1 is the strongest and 3 is the weakest interaction effect. The number of interactions, of any strength, for each allele varied from over a hundred interactions to only a few. Only a few of the alleles had a very high number of interactions, while the majority of the alleles had fewer than twenty. As clearly shown in Figure 2.3.1, some mutations affect actin's ability to interact with CHI genes more than others.

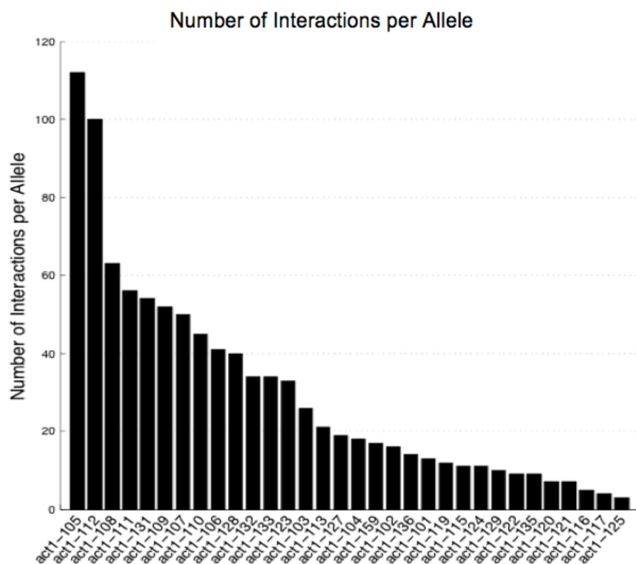

Figure 2.3.1 Histogram of the Number of Interactions for each Allele
This is a histogram showing the number of interactions that each allele has with the CHI genes. The alleles are sorted on the bottom axis beginning with the highest number of interactions on the left.

Figure 2.3.2 shows a heat map of the alleles along the surface. Alleles with the most interactions are colored in black and alleles with the fewest

interactions are colored in light grey. By looking at this figure we can begin to see interaction level differences between different areas on the molecule.

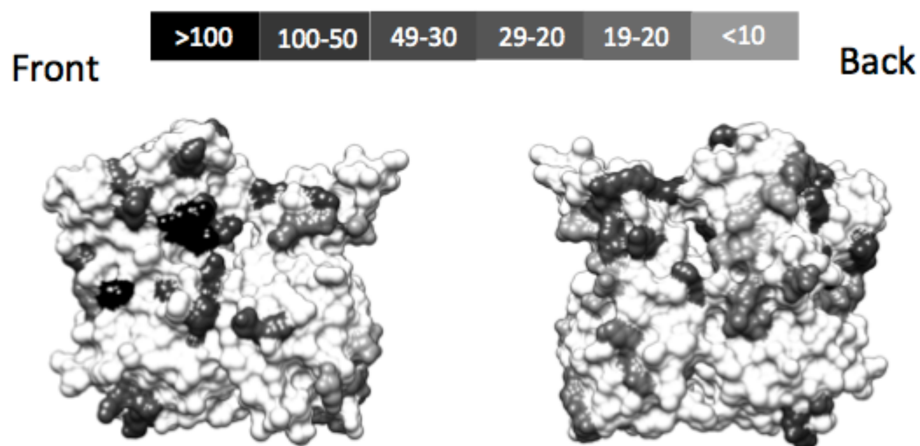

Figure 2.3.2 Heat Map of Interactions on the Surface of Actin
This figure shows a heat map of the interactions on the surface of the molecule. The alleles with the highest number of interactions are in black and the alleles with the lowest in light grey. Actin structure visualized in UCSF Chimera (Pettersen et al. 2004).

2.4 Comparisons of broad locations show hot spots on the front/top/bottom

Studying the location differential on the surface could allow us the ability to find the active binding areas that external molecules use to bind with the surface of actin. It could also help us to understand the flexibility of actin, how the filament forming process can be changed, and how that affects other processes within the cell. Since actin forms filaments, we divided the alleles into the following four location groups to measure and compare the interaction levels of each group: the front face, the top/bottom, the back face, and the sides of the molecule. The front face is the area on the front of the molecule that is the most exposed in filament form. The top/bottom area of the molecule affects actin’s ability to stack and mutations here could cause problems in filament stability.

The back face is obscured from the environment of the cell when actin is in a filament, but it may have some small effect on filament stability if it is involved in the backing process. The sides are considered to be mostly obscured as well because there is little space between molecules. Because some mutations are in the border areas of the molecule, judgment calls were made as to which class the mutation belonged, and these calls can be seen in table 2.2.1 above. In agreement with our expectations, alleles with exposed mutations generally exhibited more interactions than those with changes that were more obscured. However, these differences where slight and not significant due to a lack of statistical power. We therefore combined the categories to into two broader groups: front/top/bottom, which we expected to be highly active because they are exposed or involved in filament stacking and would therefore be more sensitive to changes, and the back/sides, which is expected to be tucked inside the filament or otherwise harder to access. The 15 front/top/bottom alleles have higher mean number of interactions than the 17 back/sides. The median number of interactions for front/top/bottom and back/sides are 37 and 11 respectively ($p<0.024$, ranksum). Figure 2.4.1 shows the mean number of interactions with standard error. Note that for this analysis, act1-159 and act1-136 were not included because act1-159 is known to be inside the active cleft and therefore is not on the outer surface and act1-136 is not available on visualization.

As we expected, the outward-facing surface would have more significant interactions than the back, which is inaccessible when actin is in filamentous

form. This analysis confirmed that there was a higher level of interactions on the outward-facing surface. We also see that when the stacking ability was possibly impaired or altered there was also an increase in interactions. It is possible that some point mutations actually affect the structure of the molecule far from the region near the mutated amino acid. Thus, further study into the structure of each mutant would be necessary before direct conclusions of a defined binding site could be drawn.

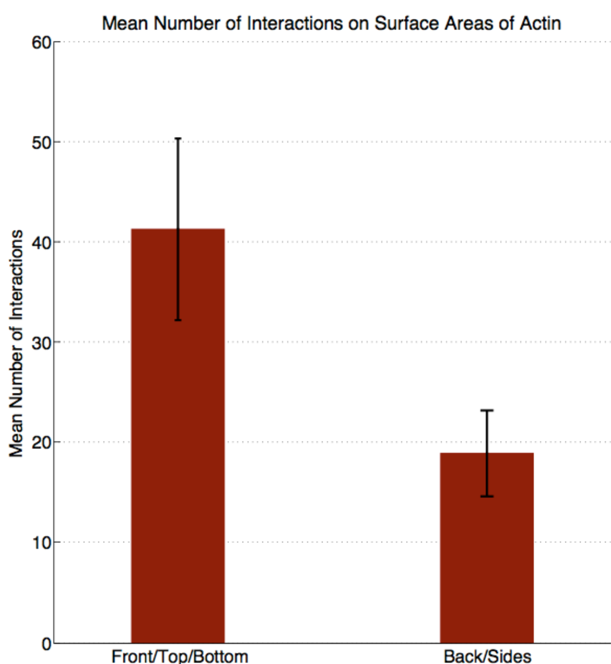

Figure 2.4.1 Mean Number of Interactions On Surface Areas of the Actin Protein
This figure shows the average number of interactions for the two location categories, Front/Top/Bottom and Back/Sides. The former is thought to be the more exposed or essential part to the function of actin, while the latter is thought to be hidden and more flexible. Error bars show the standard error within the number of interactions for each allele set.

2.5 Hierarchical clustering of allele profiles relates to protein structure

We clustered the actin mutant interaction profiles to find meaningful patterns between the mutations and the genetic interactions. We used Cluster

3.0 (de Hoon et al. 2004) to perform hierarchical clustering for both alleles and genes with which they interact. Similarity was measured using uncentered Pearson correlation with average linkage. The clustering algorithm required the weight of the scores to be reversed so that 3 became the strongest score and 1 became the weakest so that the weak scores were closest to non-interactions, shown in black, and the strongest would be the brightest white.

Actin alleles that cluster together are expected to affect the actin molecule in similar ways, because a similar set of CHI genes interacts with those alleles. When the interaction profiles are clustered, the alleles do form informative groups. We find that nearly all the allele clusters have mutations that are members of structural neighborhoods on the surface of the molecule. That is, the point mutations are either very near each other, an example of this is cluster 7 shown in Figure 2.5.2 (A), or affect a single domain of the protein, as shown in cluster 3. This supports the theory that certain areas of the molecule are important for certain aspects of actin's function.

One interesting case is cluster 3, labeled on Figure 2.5.1, and is arranged on the surface as shown in Figure 2.5.2 (A). At first glance, the affected residues appear to be far apart and on opposite ends of the molecule. However, when actin is in its filamentous form, these residues are incredibly close together on two adjacent stacked monomers. In fact, both are located on the salt bridge that holds the stacked dimer together. The clustering by profile was able to distinguish this significant group as well as the other allele clusters. While cluster

4, labeled on Figure 2.5.1, did not map within a close neighborhood, it does contain act1-159, which contributes to the stabilization of the molecule's ability to stack (Haarer et al. 2007) and we could therefore hypothesize that perhaps the other alleles in the cluster have similar functional defects based on the similarity of the interaction partners. While the genes with which the actin alleles interact are clustered as well, they were all within a very specialized subset of genes that are CHI with actin. This very specific background made searching gene enrichment in Gene Ontology terms (Gene Ontology Consortium 2004) yield very few significant results.

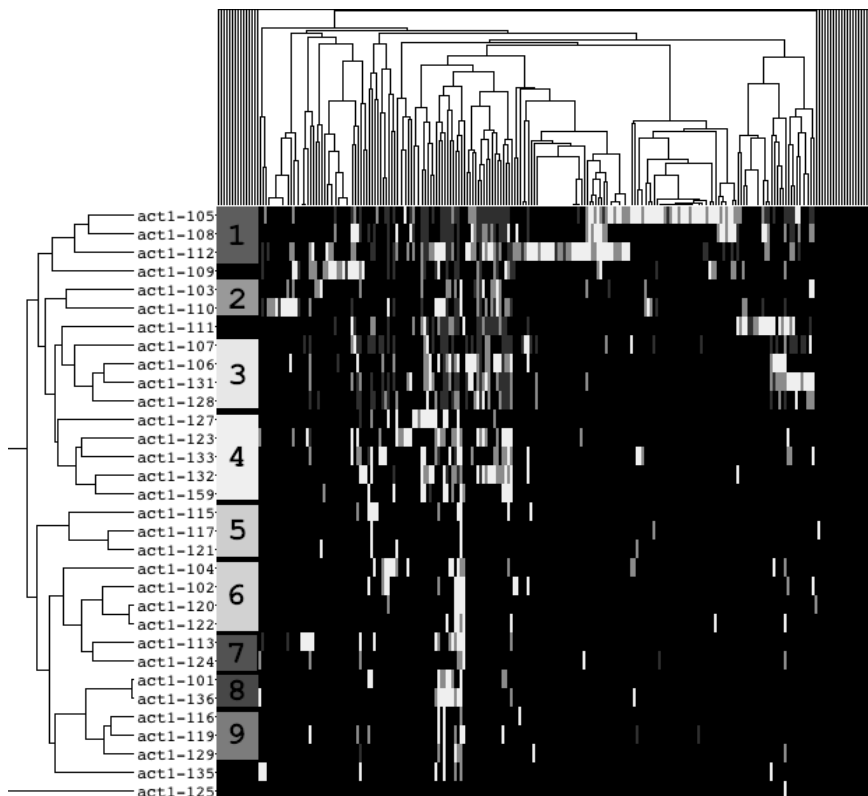

Figure 2.5.1 Clustergram of Interaction Profiles Showing Allele Clustering
This clustergram, visualized using Java Treeview (Saldanha 2004), shows hierarchical clustering of the genes and alleles. Interactions are shown in white, brighter white is stronger interactions, while black is no interaction. Along the left hand side clusters of alleles are numbered for reference.

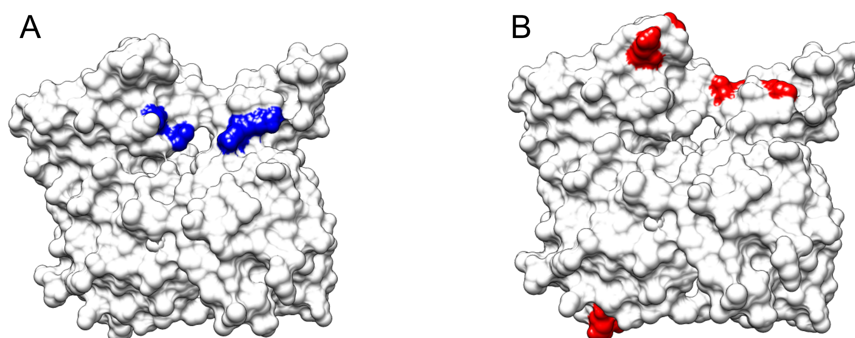

Figure 2.5.2 Sample Cluster Locations on Actin's Surface
(A) represents cluster 7 from Figure 2.5.1. It contains two alleles, act1-113 and act1-124. (B) represents cluster 3 from Figure 2.5.1 and contains 4 alleles, act1-106 , act1-107, act1-131 and act1-128. Actin structures visualized in UCSF Chimera (Pettersen et al. 2004).

2.6 Stacked distance measurements show correlation with interaction profile

Since we have shown that some areas of actin's surface are more interactive than others, we could also ask if mutations that are close in proximity are more likely to have common interaction partners than distant mutations. We hypothesized that the closer the mutations are on the structure, the more their interaction profiles should be correlated. To study this, we examined three sets of distance measurements: monomer, stacked dimer, and backed dimer. The monomer distance gives the shortest straight-line distance between the alpha carbons of the two mutated amino acids on a single protein. When actin forms filaments, it stacks monomers to form a strand and backs two of those strands together, so we also considered how the mutations may interact in a backed dimer configuration and a stacked dimer configuration. Figure 2.6.1 illustrates a few filament configurations. (A) shows two stacked monomers, while (B) shows

the two stacked monomers with a third backed up to them, and (C) shows the same as (B) but from a side angle.

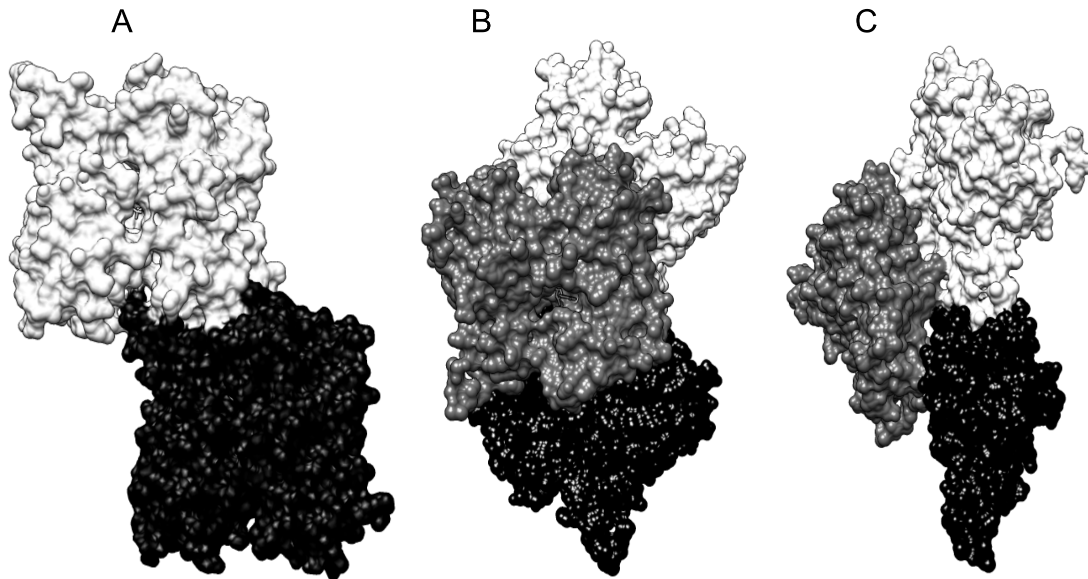

Figure 2.6.1 Stacked and Backed Configurations of Actin Monomers
(A) shows two actin molecules in stacked formation from the front. (B) shows the back of the two actin molecules in (A) with a third actin monomer backed against it. (C) shows the arrangement in (B) from the side to show the contact that occurs between the two strands. Actin structures visualized in UCSF Chimera (Pettersen et al. 2004).

These distances were found by measuring the shortest straight-line, Euclidean, distance from one mutation on one protein to another mutation on the other protein.

We measured the interaction profile similarity using Pearson correlation and measured the correlation between the distance profile and the interaction profile by using Pearson correlation as well.

When we compared the interaction profile similarity to the distance on a stacked dimer, we find a significant negative correlation ($r=-0.1424$, $p=0.004$). This suggests that the smaller the distance between stacked dimer, the more

similar the profiles tend to be. Backed and monomer distances did not show significant correlations. This may imply that the mutations in general are more detrimental to how actin functions as a filament than as the monomer, and that between stacking and backing, stacking has more influence over the filament stability. This was not surprising because actin is very specialized in its structure to ensure for very fast stacking so that building and destroying of filaments is a fluid process. When the stacking is not exact, the filament cannot be formed properly and, thus, it is probable that it would be the arrangement that is most affected by the mutations. Figure 2.6.2 shows a scatter plot of the profile correlation and the stacked distance.

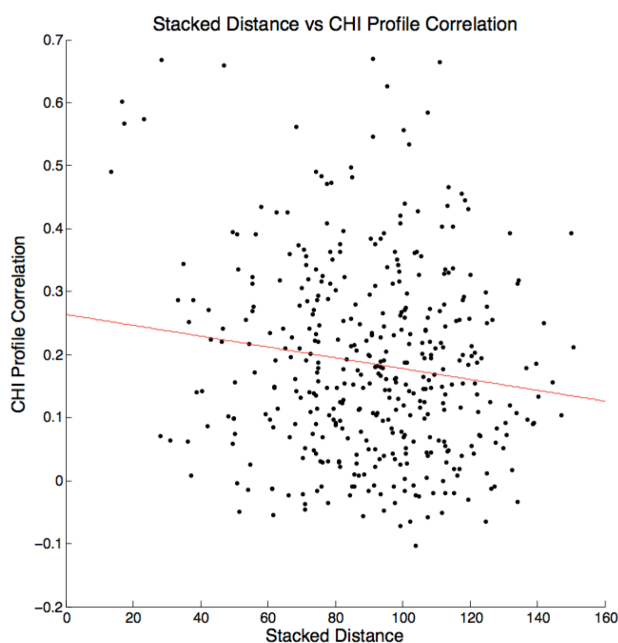

Figure 2.6.2 Scatter Plot of Profile Correlation and Stacked Distance Measure
This graph shows stacked dimer distance versus the CHI profile correlation. The distance measure is on the bottom axis and the profile correlation is on the left. The correlation is measured by Pearson correlation. The distance is measured as the shortest straight-line distance between the mutations on the two molecules. The best-fit line is also shown.

Unless it is otherwise stated, all statistical analyses were performed and statistical figures were created using MATLAB (R2011b, The MathWorks, Natick, MA).

2.7 Conclusions

By looking at the distribution of interactions among the alleles, we could initially see that some mutations had much more impact on the cell than others. When we compared areas on the surface, we were able to see an interaction differential in different areas on the molecule, showing that the front face/top/bottom were more interactive than the sides/back face. Hierarchical clustering analysis of the alleles proved a meaningful way to see that the profiles gather into clusters that could be assigned to specific regions of the actin surface. The gene clustering was less revealing due to an already specialized background of actin interacting genes, but introduction of a broader background or fine-tuning the tools may help elucidate more findings on this. We could also see that the stacking distance correlates with the profiles of the alleles, which showed an interesting connection between the stacking ability and interactions.

CHAPTER 3: Comparison of actin genetic interactions between haploids and diploids

3.1 Advances in haploid technology have led to expanding haploid data

There have been significant advances in genome-scale genetic interaction mapping in haploid yeast. As of 2010, 5.4 million double mutant strains have been constructed to study synthetic gene interactions (Costanzo et al. 2010), and nearly every pair-wise combination will soon be available. This 5.4 million pair set accounts for ~30% of all possible gene pairs across the yeast genome and has been extremely informative for haploid genetics. However, no studies have evaluated how these interactions translate to diploids. These double mutant pairs were constructed by the Boone Laboratory in Toronto using Synthetic Genetic Array (SGA) methodology. SGA utilizes an automated mating and selection process to rapidly produce arrays of double mutant yeast colonies. This high-density data is relatively easy to create because of the fairly simple genetics of haploid yeast and advances in technology, including robotic manipulation. A genetic interaction in this data set is defined to be a pair of genes that, when mutated in the same strain, show a significantly different fitness from the expected fitness based on the multiplied fitnesses of the two single-mutant strains (Baryshnikova et al. 2010). We are particularly interested in negative interactions, which are instances where the fitness is much lower than expected. Using these genetic interactions, Costanzo et al. built a genetic network illustrating how the genes are connected. Clustering this network resulted in a functional map of the yeast cell, organizing genes and their

functions. In this way, Costanzo et al. were able to reveal connections between pathways and complexes that previously had not been understood. Using a guilt by association approach, they were able to infer the roles of some uncharacterized genes based on which other genes have similar profiles. When Costanzo et al. compared the data against protein-protein interaction networks and other data sets; they found that many of the sets significantly complemented each other (Costanzo et al. 2010). This data reveals much about haploid genetics, but it begs the question, “How can we relate these findings to the genetics of diploid organisms such as humans?”

3.2 Creating a haploid set to study against the diploid set

To investigate this, the same technique that was applied in the haploid yeast can also be used on actin mutants. The creation of a haploid version of the diploid experiment was also completed by the Boone Lab at the University of Toronto as part of a larger set of screens. Since these haploid strains have only one copy of every chromosome, they have only a single copy of actin to mutate and a single copy of each Gene X to mutate. For this test, the Boone lab constructed a collection of yeast that has a mutated copy of actin and a null copy of the gene X. The arrangement for this can be seen below in Figure 3.2.1

Figure 3.2.1 Example Construct of Haploid Double Mutant
This is a construct showing the genetic arrangement in the haploid yeast. Here they are shown on the same chromosome, but this is not true of all pairs.

This is a more extreme version of the diploid study because the yeast will only have actin proteins made from the mutated gene instead of half mutated half wild type as in the case of heterozygous diploids. The second protein, made from gene X, will be completely missing. We expect those gene pairs that caused sick phenotypes in the CHI study to have much sicker phenotypes in haploids and we expect not to see any growth in strains containing some of the most severe mutations, such as the ones near the salt bridge, as they interfere with stacking. We hypothesize that in the CHI experiment, there is enough normal actin that the stacking mechanism can compensate for the mutated portions, but when all of the actin are mutated, the whole stack is too unstable to operate. Because of this, we are only able to make the haploid/diploid comparison for some of the alleles and genes. As shown in Figure 3.2.2 below, the overlap between the two data sets is 9 alleles and 128 genes. While the subset is too small for some analyses, we are able to make some significant comparisons.

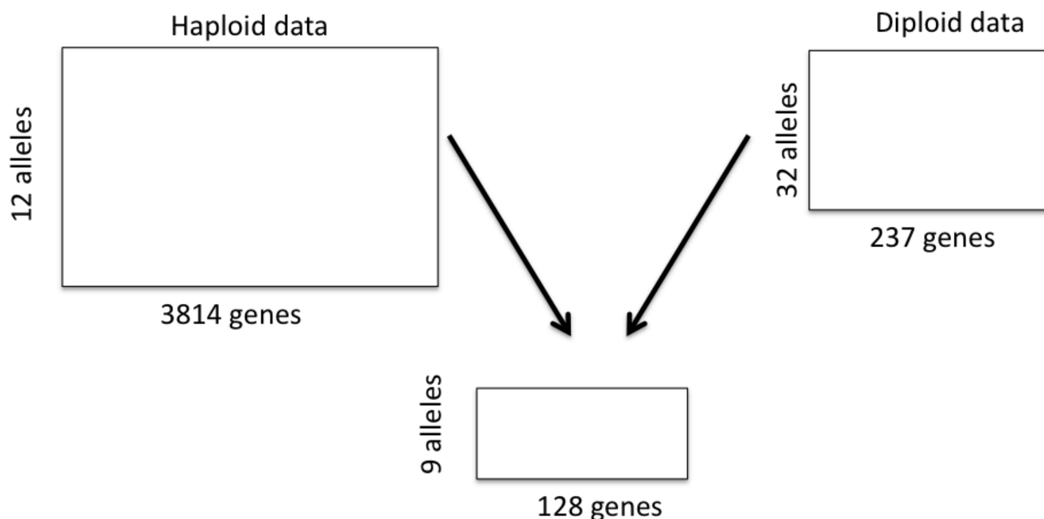

Figure 3.2.2 Schematic of Haploid and Diploid Overlap
This diagrams the data sets and the subset of common alleles and genes. The left hand side is the haploid data which consists of 12 alleles and 3814 genes. The diploid set on the right contains 32 alleles and 237 genes. There are only 9 alleles that are in common between the two and only 128 genes.

3.3 Haploid interactions show a broad enrichment for CHI genes

To begin comparing the two data sets, we started by looking for a meaningful connection on a very broad level, such as whether or not the genes within the CHI set, the genes that have interactions with the heterozygous WT/null actin mutant, account for a significant number of the negative interactions with the actin mutants in haploids. In Figure 3.3.1, we see an example of the enrichment measure that is computed for each allele in haploid set. In the example below, one out of two of the shown interactions for act1-101 are on CHI gene list.

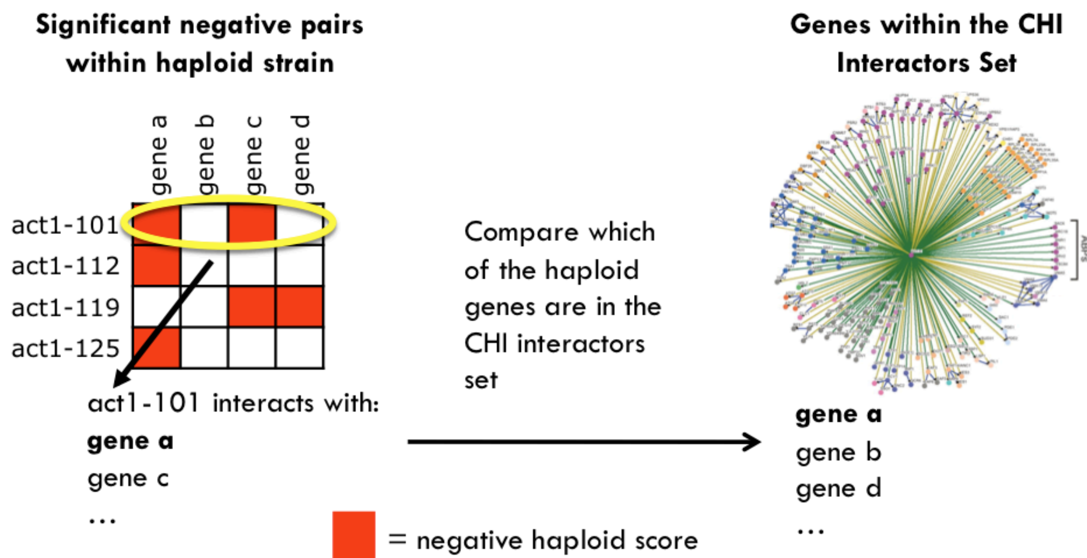

Figure 3.3.1 Diagram of Broad Enrichment Data
The matrix represents a portion of the haploid data where a colored square is a significant negative genetic interaction. Below this shows a list of genes that we see interact with the chosen allele act1-101. The right shows a portion of the CHI gene list. We compare these two lists to see of the genes that act1-101 interacts with in a haploid, how many of those are also in the CHI gene list.

When measured using a hypergeometric test, we find that nearly every allele has interactions that are significantly enriched for CHI genes. Table 3.3.1 shows the overlap and significance of each allele. Only one allele is not significant (act1-101). This tells us that the interactions of the mutant actin alleles in haploids overlap significantly with the set of genes that are complex haploinsufficient with actin.

Table 3.3.1 Allele Specific Enrichment of CHI genes in Haploid Interactions
This table shows the allele level broad enrichment of CHI genes within the significant negative haploid interactions. It includes the location call of the mutation, the number of significant negative interactions in the haploid (SGA) set, the number of CHI genes available (this is the same for all alleles), the number of genes that are shared between the two, and the hypergeometric p value (this is shaded when p<0.05).

Allele	Location	Haploid Interactions	CHI Interactors	Negative Overlap	Hyper-geometric P value
act1-120	Side	127	128	17	7.08E-07
act1-125	Side	143	128	17	3.85E-06
act1-132	Back Face	392	128	30	8.72E-06
act1-121	Side	227	128	20	5.03E-05
act1-129	Back Face	153	128	15	1.47E-04
act1-124	Front Face	192	128	15	0.0017
act1-155	NA	328	128	20	0.0058
act1-119	Back Face	164	128	12	0.0082
act1-3	NA	421	128	23	0.0115
act1-4	NA	211	128	13	0.0233
act1-112	Front Face	39	128	4	0.0402
act1-101	Side	94	128	5	0.2068

3.4 Interaction profile comparisons reveal differences between haploid and diploid interactions

To evaluate the agreement between actin genetic interactions in haploids and diploids more specifically, we further compared the specific profiles of the two screens. To do this, we looked at profile similarity of the actin mutant alleles as indicated by the two different sets of genetic interactions. We built a measure of how similar an allele's profile is to every other allele's profile by correlating each allele against every other allele resulting in a 9 by 9 symmetrical matrix of Pearson correlation coefficients. Comparing the profiles allowed us to measure

the similarity between alleles in either haploids or diploids as opposed to looking at specific interactions of the alleles. When we looked at the Pearson correlation of the profiles, we found that there is not a significant correlation between the two data sets. This tells us that we are unable to say that alleles that behave similarly in haploids are the same as alleles that behave similarly in diploids. The difference may be in part because in the CHI diploid yeast, there is a mixture of healthy actin and mutant actin, and the cell may be able to compensate for the mutant actin with the use of the wild type. However, when the mutant is the only allele present, as it is in the haploid set, the effects are either much more pronounced, for example when the mutation is on the salt bridge, or easier to work around because the actin molecules are homogenous (i.e. only the mutant allele is expressed), such as when the mutation is on the front of the protein.

Another way to more specifically compare the subsets is to directly compare their interactions among the set of genes screened in common. To do this, we looked for enrichment of diploid interactions within the haploid interactions for each mutant allele. Note that the diploid mutant actin screens were only conducted against the set of 243 CHI genes and only 128 of these were screened in the haploid genetic interaction screens. Thus, our test for direct overlap of interactions for each mutant actin allele was restricted to these 128 genes. As illustrated in Figure 3.4.1, we are now comparing the number of darker boxes in the haploid set that line up with lighter boxes in the diploid set to see which pairs match within each of the sets. If we break it down by allele and

look specifically at act1-119, the haploid and diploid rows have one of its two interactions in common, act1-119 and gene c.

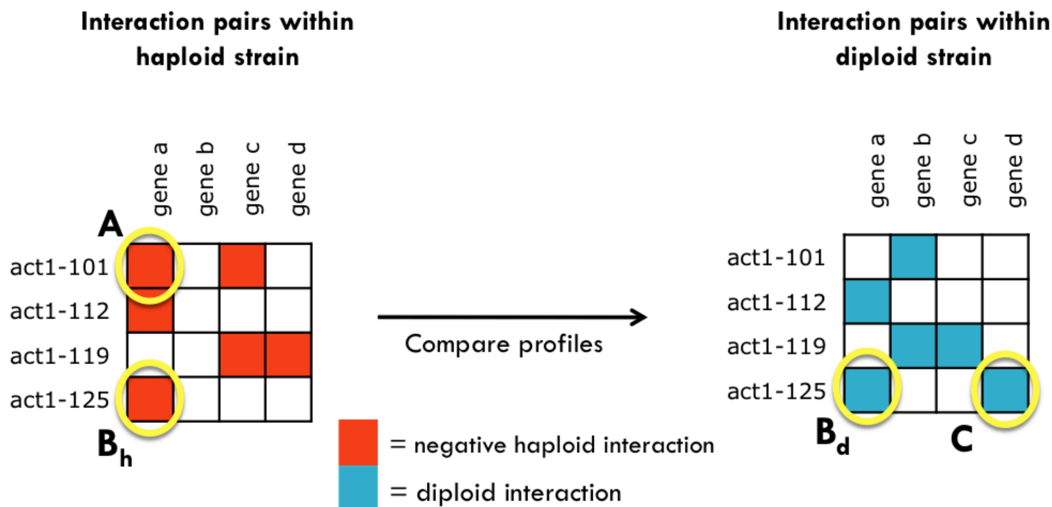

Figure 3.4.1 Diagram of Local Specific Overlap
The left matrix represents a portion of the haploid data where a darkened square is a significant negative genetic interaction. The right matrix represents a portion of the diploid data set. In the specific enrichment there are three possibilities. (A) shows an interaction that is in the haploid set but not the diploid set. (B_h) and (B_d) show an interaction that is found in both the haploid and the diploid set. Finally, (C) shows an interaction that is only in the diploid set. We are most interested in the B, the interactions that are in both sets.

There are 1152 possible pairs in the haploid subset (9 alleles by 128 genes), of which there are 135 significant negative pairs. There are 72 CHI scoring pairs within the diploid subset. Between the two, there are only 7 that overlap. Figure 3.4.2 illustrates the overlap.

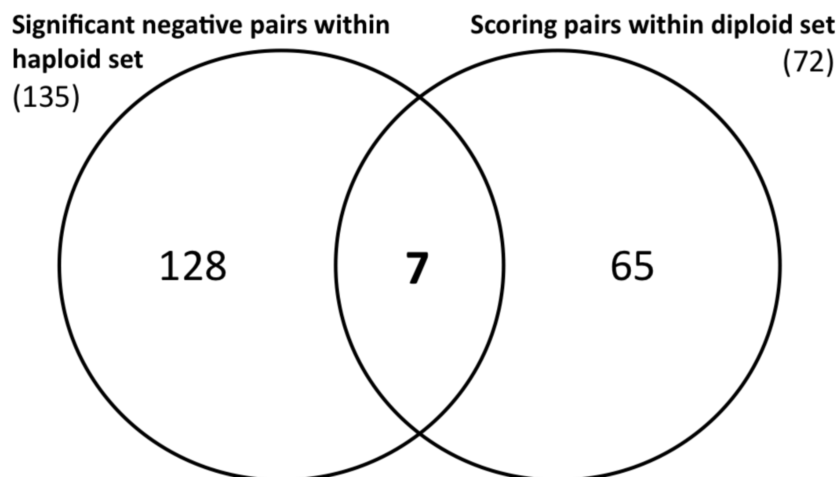

Figure 3.4.2 Distribution of Specific Interaction Pair Overlap in the Global Subset
This Venn diagram illustrates the distribution of interactions for specific enrichment. On the left we have the number of significant negative pairs in the subset (135) and on the right we have the number of CHI gene/allele scoring pairs (72). Of those, there are only 7 pairs in common. There are 1152 possible CHI gene/allele pairs in the subset haploid data. To the hypergeometric p value of 7 of the pairs being shared given the numbers in each set is shown in the bottom right.

In this case, the overlap is not significant with a hypergeometric p-value of 0.7620. When we break down the data into the individual alleles, the significance is again very poor for nearly all alleles (see Table 3.4.1). Note that in both cases, we compared only for the presence/absence of significant interactions, not the magnitude of the interactions.

Table 3.4.1 Allele Specific Overlap of Profile Interaction Pairs
This table shows the allele level specific profile enrichment of CHI interactions within the significant negative haploid interactions. It includes the location call of the mutation, the number of significant negative interactions in the haploid (SGA) set, the number of diploid interactions, the number of genes that are shared between the two, and the hypergeometric p value (this is shaded when p<0.05).

Allele	Location	Haploid Interactions	CHI Interactors	Negative Overlap	Hypergeometric P value
act1-132	Back Face	30	8	4	0.0865
act1-124	Front Face	15	1	1	0.1172
act1-112	Front Face	4	51	2	0.5234
act1-101	Side	5	6	0	1
act1-119	Back Face	12	3	0	1
act1-120	Side	17	1	0	1
act1-121	Side	20	1	0	1
act1-125	Side	17	0	0	1
act1-129	Back Face	15	1	0	1

3.5 While haploids can tell us which genes, we need data from diploids to learn how the interactions occur.

The non-significant overlap in the results tells us something quite interesting. The interactions that we see in the haploid set reflect the genes that were found in the CHI null experiment with very high significance. Through haploid or diploid experiments, we are able to find a set of genes that functionally relate and interact with actin. However, when we look more specifically at how each mutant actin allele interacts in a diploid cell and to what level those interactions are conserved in the haploid, we are unable to find a relation other than that expected by chance. This suggests, at least in the case of actin, that there are fundamental differences between genetic interactions observed in haploid versus diploid cells. An obvious instance of this is where a mutation in

one allele can be compensated by a wild-type copy of actin in the diploid. In such a case the mutant allele may only result in subtle defects in the diploid but could be highly deleterious in a haploid. This seems to be true in the area along the top/bottom of the actin molecule, for which the mutants were not even screened for genetic interactions in haploids because they result in very strong fitness defects. In the above table 3.4.1, we see that each of the haploid alleles has more interactions than its diploid counterpart, suggesting that in general, the mutations in a haploid environment are more deleterious, which is consistent with the expectation that most mutations will have stronger defects in a haploid context.

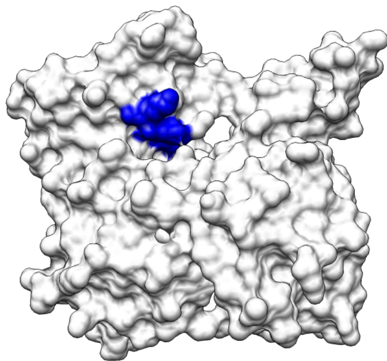

Figure 3.5.1 Visualization of act1-112
This shows a front facing image of an actin monomer. Shaded on the structure are the residues that are mutated in the act1-112 allele. These are residues 213-215. Actin structure is visualized in UCSF Chimera (Pettersen et al. 2004).

There is one interesting exception, however, act1-112. Act1-112 has 51 diploid interactions but only 4 significant negative haploid interactions. This allele is on the front face of the molecule as shown in Figure 3.5.1. The reason for this difference is unclear, but we can make a reasonable hypothesis. For

example, if this is an important location in binding proteins and enabling them to walk along the filament, then the bound protein may be able to compensate for the mutation when it is uniform from monomer to monomer as in the haploid case (the mutant allele is the only one expressed). However, when the filament is irregular and switches between mutant and wild type, the interacting protein may be more disrupted, resulting in a larger set of downstream interactions than the diploid context.

In summary, the haploid data is able to point us to a set of genes that is likely to be enriched with interactions in diploids as indicated by the enrichment for CHI genes. However, our analysis indicates that haploid screens are limited in their ability to predict the specific interactions of mutant forms of actin in heterozygous diploids. Thus, understanding complex genetics in diploid organisms will require more focused screening efforts.

Bibliography

Alberts B, A. Johnson, J. Lewis, et al. "The Self-Assembly and Dynamic Structure of Cytoskeletal Filaments." *Molecular Biology of the Cell*. 4th ed. New York: Garland Science, 2002. <http://www.ncbi.nlm.nih.gov/books/NBK26862/>.

Altshuler, David, Mark J. Daly, Eric S. Lander. “Genetic Mapping in Human Disease.” *Science*. 322.5903 (2008): 881-8

Baryshnikova, Anastasia, Michael Costanzo, Yungil Kim, Huiming Ding, Judice Koh, Kiana Toufighi, Ji-Young Youn, Jiongwen Ou, Bryan-Joseph San Luis, Sunayan Bandyopadhyay, Matthew Hibbs, David Hess, Anne-Claude Gingras, Gary D. Bader, Olga G. Troyanskaya, Grant W. Brown, Brenda Andrews, Charles Boone, and Chad L. Myers. "Quantitative Analysis of Fitness and Genetic Interactions in Yeast on a Genome Scale." *Nature Methods* 7.12 (2010): 1017-024.

Bettinger, Blaine T. , David M. Gilbert & David C. Amberg. “Actin up in the nucleus.” *Nature Reviews Molecular Cell Biology* 5 (2004): 410-15.

Botstein, David, Steven A. Chervitz, and Michael Cherry. "Yeast as a Model Organism." *Science* 277.5330 (1997): 1259-269.

Costanzo, M., A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L.Y. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St. Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.-Y. Lin, W. Liang, M. Marback, J. Paw, B.-J. San Luis, E. Shuteriqi, A. H. Y. Tong, N. Van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pal, F. P. Roth, G.

Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A.-C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J. Andrews, and C. Boone. "The Genetic Landscape of a Cell." *Science* 327.5964 (2010): 425-31.

de Hoon, M. J. L., S. Imoto, J. Nolan, and S. Miyano. “Open Source Clustering Software.” *Bioinformatics* 20.9 (2004): 1453-54.

Dixon, Scott J, Michael Costanzo, Anastasia Baryshnikova, Brenda Andrews, and Charles Boone. “Systematic Mapping of Genetic Interaction Networks.” *Annual Review of Genetics* 43 (2009): 601-625.

Dominguez, Roberto, and Kenneth C. Holmes. "Actin Structure And Function." *Annual Review of Biophysics* 40.1 (2010): 169-86.

Gene Ontology Consortium. “The Gene Ontology (GO) database and informatics resource.” *Nucleic Acid Res.* 32 (2004): D258-D261

Giaever, Guri, Angela M. Chu, Li Ni, Carla Connelly, Linda Riles, Steeve Véronneau, Sally Dow, Ankuta Lucau-Danila, Keith Anderson, Bruno André, Adam P. Arkin, Anna Astromoff, Mohamed El Bakkoury, Rhonda Bangham, Rocio Benito, Sophie Brachat, Stefano Campanaro, Matt Curtiss, Karen Davis, Adam Deutschbauer, Karl-Dieter Entian, Patrick Flaherty, Francoise Foury, David J. Garfinkel, Mark Gerstein, Deanna Gotte, Ulrich Güldener, Johannes H. Hegemann, Svenja Hempel, Zelek Herman, Daniel F. Jaramillo, Diane E. Kelly, Steven L. Kelly, Peter Kötter, Darlene LaBonte, David C. Lamb, Ning Lan, Hong Liang, Hong Liao, Lucy Liu, Chuanyun Luo, Marc Lussier, Rong Mao, Patrice Menard, Siew Loon Ooi, Jose L. Revuelta, Christopher J. Roberts, Matthias Rose, Petra Ross-Macdonald, Bart Scherens, Greg Schimmack, Brenda Shafer, Daniel D.

Shoemaker, Sharon Sookhai-Mahadeo, Reginald K. Storms, Jeffrey N. Strathern, Giorgio Valle, Marleen Voet, Guido Volckaert, Ching-yun Wang, Teresa R. Ward, Julie Wilhelmy, Elizabeth A. Winzeler, Yonghong Yang, Grace Yen, Elaine Youngman, Kexin Yu, Howard Bussey, Jef D. Boeke, Michael Snyder, Peter Philippsen, Ronald W. Davis, and Mark Johnston. "Functional Profiling of the Saccharomyces Cerevisiae Genome." *Nature* 418.6896 (2002): 387-91.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. "Life with 6000 Genes." *Science* 274.5287 (1996): 546-67.

Haarer, B., S. Viggiano, M. A. Hibbs, O. G. Troyanskaya, and D. C. Amberg. "Modeling Complex Genetic Interactions in a Simple Eukaryotic Genome: Actin Displays a Rich Spectrum of Complex Haploinsufficiencies." *Genes & Development* 21.2 (2007): 148-59.

International Human Genome Sequencing Consortium. "Finishing the Euchromatic Sequence of the Human Genome." *Nature* 431.7011 (2004): 931-45.

Legras, J.-L., D. Merdinoglu, J.-M. Cornuet, F. Karst. “Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history.” *Molecular Ecology*. 16 (2007): 2091–2102.

MacArthur, Daniel G. , S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J. K. Pickrell, S. B. Montgomery, C. A. Albers, Z. D. Zhang, D. F. Conrad, G. Lunter, H. Zheng, Q. Ayub, M. A. DePristo, E. Banks, M. Hu, R. E. Handsaker, J. A. Rosenfeld, M. Fromer,

M. Jin, X. Jasmine Mu, E. Khurana, K. Ye, M. Kay, G. Saunders, M. Suner, T. Hunt, If H. A. Barnes, C. Amid, D. R. Carvalho-Silva, A. H. Bignell, C. Snow, B. Yngvadottir, S. Bumpstead, D. N. Cooper, Y. Xue, I. Gallego Romero, 1000 Genomes Project Consortium, J. Wang, Y. Li, R. A. Gibbs, S. A. McCarroll, E. T. Dermitzakis, J. K. Pritchard, J. C. Barrett, J. Harrow, M. E. Hurles, M. B. Gerstein, C. Tyler-Smith. “A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes.” *Science* 335.6070 (2012): 823-828.

Mackenzie SM, MR Brooker, TR Gill, GB Cox, AJ Howells, GD Ewart. “Mutations in the white gene of Drosophila melanogaster affecting ABC transporters that determine eye colouration.” *Biochim Biophys Acta*. 1419.2 (1999): 173-85

National Human Genome Research Institute. "A Brief Guide to Genomics." *A Brief Guide to Genomics*. National Human Genome Research Institute, 19 Oct. 2011. Web. 04 Feb. 2013. <http://www.genome.gov/18016863>.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. “UCSF Chimera--a visualization system for exploratory research and analysis.” *J Comput Chem* 25.13 (2004): 1605-12.

Ralston, A. “Gene interaction and disease.” *Nature Education* 1.1 (2008)

Saldanha, Alok J. “Java Treeview—extensible visualization of microarray data” *Bioinformatics* 20.17 (2004): 3246-48

U.S. Department of Energy Genome Program. "SNP Fact Sheet." *SNP Fact Sheet*. U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Human Genome Program, 19 Sept. 2008.

Web. 04 Feb. 2013.
<http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml>.

U.S. National Library of Medicine. "Actin." *Genetics Home Reference*. U.S. National Library of Medicine, 28 Jan. 2013. Web. 04 Feb. 2013. <http://ghr.nlm.nih.gov/handbook/illustrations/actin>.

Vandekerckhove, J., and Klaus Weber. "At Least Six Different Actins Are Expressed in a Higher Mammal: An Analysis Based on the Amino Acid Sequence of the Amino-terminal Tryptic Peptide*1." *Journal of Molecular Biology* 126.4 (1978): 783-802.

Wertman, K. F., D. G. Drubin, and D. Botstein. "Systematic Mutational Analysis of the Yeast ACT1 Gene." *Genetics* 132.2 (1992): 337-50.

Yeong, F. M. “Severing all ties between mother and daughter: cell separation in budding yeast.” *Molecular Microbiology*. 55 (2005): 1325–31.