Identifying Genetic Causes of Specific Phenotypes by Whole-Genome Sequencing Analysis

by

Rong Huang

A thesis submitted in conformity with the requirements
for the degree of Master of Science

Molecular Genetics
University of Toronto

Identifying Genetic Causes of Specific Phenotypes by Whole-Genome Sequencing Analysis

Rong Huang

Master of Science

Molecular Genetics
University of Toronto

2015

Abstract

The advent of next-generation sequencing (NGS) makes it more practicable to extract genetic information from biological systems. Whole-genome sequencing analysis has become one of the major approaches to identify genetic causes of specific phenotypes, with the help of well-developed mapping and annotation tools. My project is aimed to identify causative mutations for some mutant *Saccharomyces cerevisiae* strains. One class of mutants belong to Ctf strains, with increased rate of chromosome loss in mitosis. Exploration of Ctf genes is helpful to understand cellular causes and consequences of chromosome instability (CIN), thereby elucidating mechanisms of some human diseases. Another class of mutants are strains that adapted for stronger growth in His- medium in a yeast two-hybrid assay ("papillation strains"). Papillation strains could harbor spontaneous genetic modifiers of protein-protein interactions (PPIs). By computational analysis of whole-genome sequencing data, I identified some candidate genes responsible for these two phenotypes, respectively, and validated some candidate Ctf genes experimentally.

Acknowledgments

First, I would like to express my sincere gratitude to my supervisor, Dr. Frederick Roth, for providing patient advice, supportive mentorship and an excellent work environment during my M.Sc. study. In addition to teaching me new knowledge and techniques, he has also trained me as a researcher with independent and critical thinking. It has been a great pleasure to work with him for the past two years.

I would also like to thank my committee members, Dr. Andrew Fraser and Dr. Zhaolei Zhang, for the helpful advice, invaluable suggestions and continuous support to complete my project. Additionally, I would like to thank Dr. Philip Hieter for providing Ctf mutants and Dr. Marc Vidal and Dr. Quan Zhong for offering papillation strains, as well as their friendly suggestions.

I am truly grateful for the help and support from my lab-mates and collaborators. I want to thank Dr. Atina Coté for sharing her wide-ranging knowledge of yeast biology and next generation sequencing. I am truly indebted to Takafumi Yamaguchi and Siyang Li for their patience of teaching me how to apply the computational pipeline and answering my questions. I benefited a lot from Dr. J. Javier Díaz-Mejía for his guidance during my rotation, which made me familiar with most yeast techniques quickly. He also offered much help when I had technical problems with my experiments. I am thankful to the rest of the Roth lab, especially Dr. Song Sun, Albi Celaj, Marinella Gebbia, Sunny Yang for warm support and many invaluable discussions. I also want to thank Tejomayee Singh from the Hieter lab for her helpful suggestion for Ctf project, and Christelle, a summer student, for performing some validation under Tejomayee's supervision, which supplements my experiments.

I would also like to show my appreciation to my friends for supporting me and providing warm friendships.

Finally and most importantly, I would like to express my heartfelt gratitude to my parents and Yunjiang for their endless love, understanding and support during my M.Sc. study.

Table of Contents

List of Tables

List of Figures

List of Appendices

1 Introduction

1.1 Next-generation sequencing technology helps to reveal genetic causes of yeast mutants

The advent of next-generation sequencing (NGS) makes it faster, more practicable and more economical to extract genetic information from biological systems by genome sequencing. The genome of *Saccharomyces cerevisiae* is about 12.2Mb (S288C genome assembled by Saccharomyces Genome Database (SGD)). It is one of the smallest genomes among model eukaryotic organisms, which makes it possible to sequence numerous different samples at one time through multiplexing in a single sequencing run with deep coverage enabling variant analysis. Whole-genome sequencing analysis has become one of the major approaches to identify genetic causes of specific phenotypes, with the help of well-developed mapping and annotation tools.

1.2 Chromosome transmission fidelity (Ctf) strains

1.2.1 Chromosome instability

Chromosome instability (CIN) indicates increased rates of whole chromosome missegregation in mitosis (Thompson et al., 2010). Loss of chromosome segregation fidelity could cause aneuploidy, which represents a chromosome number that is not an exact multiple of the haploid number in a cell. This chromosomal imbalance usually has a detrimental effect on cells, that results in many human diseases, including cancer (Lengauer et al., 1997) and developmental disorders like Down syndrome (Lejeune et al., 1959). A mutational environment created by CIN in early oncogenesis could facilitate necessary oncogenes or tumor-suppressor genes mutation as well as a malignant tumor phenotype (Stratton et al., 2009). In addition, CIN could increase genetic diversity among tumor cells, which contributes to poor patient prognosis (Heilig et al., 2010). Therefore, exploring underlying mechanisms of CIN is important to understand causes and consequences of CIN and improve cancer therapy. Some mechanisms causing CIN are already well explored, including defects in chromosome cohesion, mitotic checkpoint function, centrosome copy number, kinetochore-microtubule attachment dynamics and cell-cycle regulation (Thompson et al., 2010). However, identification of CIN genes is still ongoing.

1.2.2 Chromosome transmission fidelity screen

CIN detection is based on measurement of chromosome missegregation rates (Geigl et al., 2008), which requires a combination of chromosome counting and clonal cell assays. The haploid phase of *S. cerevisiae* life cycle offers highly efficient mutagenesis screens for nonlethal phenotypes because of direct phenotyping. Around 25 years ago, a chromosome transmission fidelity (Ctf) screen for mutants with defects in CIN was conducted by the Hieter lab using a visual colony color assay to monitor inheritance of a nonessential artificial chromosome fragment (CF) (Spencer et al., 1990). Each plasmid contains a restriction site between a telomeric Y' element and a unique chromosomal sequence, a centromere, a selectable marker (*URA3*), and *SUP11* (Figure 1). Through linearization and recombination, the telocentric chromosome fragment contains one short, plasmid-derived arm with a yeast telomere, and one long arm derived from yeast chromosomal DNA. *SUP11* gene encodes a suppressing tyrosyl-tRNA, which can suppress the ochre stop codon mutation in *ade2-101*. On medium with limited adenine, *ade2-10* cells form red colonies due to the accumulation of an adenine precursor, and this block of adenine production is relieved in the presence of *SUP11*. So cells with CF give rise to white colonies whereas cells without CF develop red colonies. Therefore, white colonies with red sectors from CF containing strains indicate an increased rate of chromosome loss. After light ethyl methanesulfonate (EMS) mutagenesis, this Ctf screen yielded 136 Ctf mutants, including 15 complementation groups and 39 singletons (one gene per complementation group). (Note however that complementation was only tested for strains of opposite mating type, so that the actual number of complementation groups may be less than reported, especially for singletons.)

1.2.3 Identification of Ctf genes

Within the 25 years after the first Ctf screen, 108 Ctf alleles have been mapped to mutations in 26 genes cloned by complementation testing or tiling microarrays, but 28 singletons have remained uncloned (Stirling et al., 2012). Exhaustive screening of systematic mutant collections has shown that ~350 genes are related to Ctf phenotypes (Measday et al., 2005; Stirling et al., 2011; Yuen et al., 2007). Genes identified in the original Ctf forward genetic screen are particularly interesting, as we know that these defects can arise by spontaneous mutation making them particularly relevant to cancer. Because it had been difficult to find causative mutations for the remaining uncloned Ctf strains by classical approaches, here I

identified candidate mutations of some singletons by whole-genome sequencing analysis and experimentally validated 3 genes.

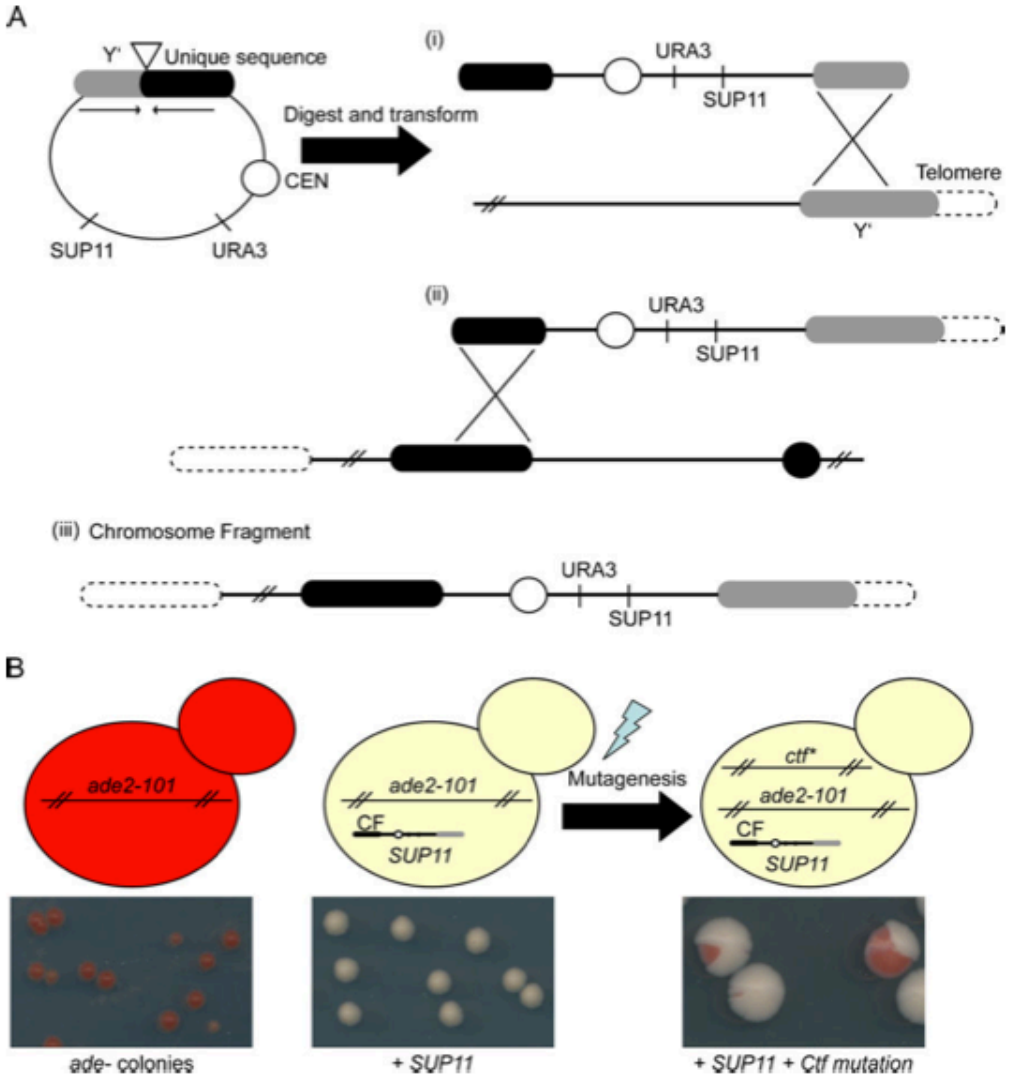


Figure 1. Features of the chromosome fragment and Ctf assay (Stirling et al., 2012). Each plasmid contains a restriction site between a telomeric Y' element and a unique chromosomal sequence, a centromere, a selectable marker (*URA3*), and *SUP11*. The telocentric chromosome fragment contains one short, plasmid-derived arm with a yeast telomere, and one long arm derived from yeast chromosomal DNA. *SUP11* gene encodes a suppressing tyrosyl-tRNA, which can suppress the ochre stop codon mutation in *ade2-101*. On medium with limited adenine, *ade2-10* cells form red colonies due to the accumulation of an adenine precursor, and this block of adenine production is relieved in the presence of *SUP11*. Cells with CF give rise to white colonies whereas cells without CF develop red colonies. Therefore, white colonies with red sectors from CF containing strains indicate an increased rate of chromosome loss.

1.3 "Papillation" strains adapted for stronger yeast two-hybrid protein interaction signal

1.3.1 Yeast two-hybrid screen is a powerful tool to map protein-protein interactions (PPIs)

Proteins interact with each other to fulfill their functions within the cell, and the dynamics of interactions regulates cell behaviors. So PPIs play essential roles in many biological processes. Therefore, many techniques have been developed to study physical protein-protein interactions (PPIs), which have contributed to our understanding of cellular levels of an organism.

The yeast two-hybrid (Y2H) method is one of the most powerful techniques for analyzing PPIs (Figure 2). A transcription factor (usually GAL4) is split into two separate parts, the binding domain (DB) and activation domain (AD), responsible for binding to the upstream activating sequence (UAS) and activating transcription of downstream reporter gene, respectively (Fields and Song, 1989). These two parts are fused to proteins of interest, X (also known as "bait" protein) and Y (also known as "prey" protein). When haploid cells with DB-X or AD-Y plasmids are mated, the diploid cells contain both DB-X and AD-Y plasmids. So physical interaction between DB-X and AD-Y could reconstitute the transcription factor and induce transcription of the reporter gene; thereby the diploid cells could grow on selective medium.

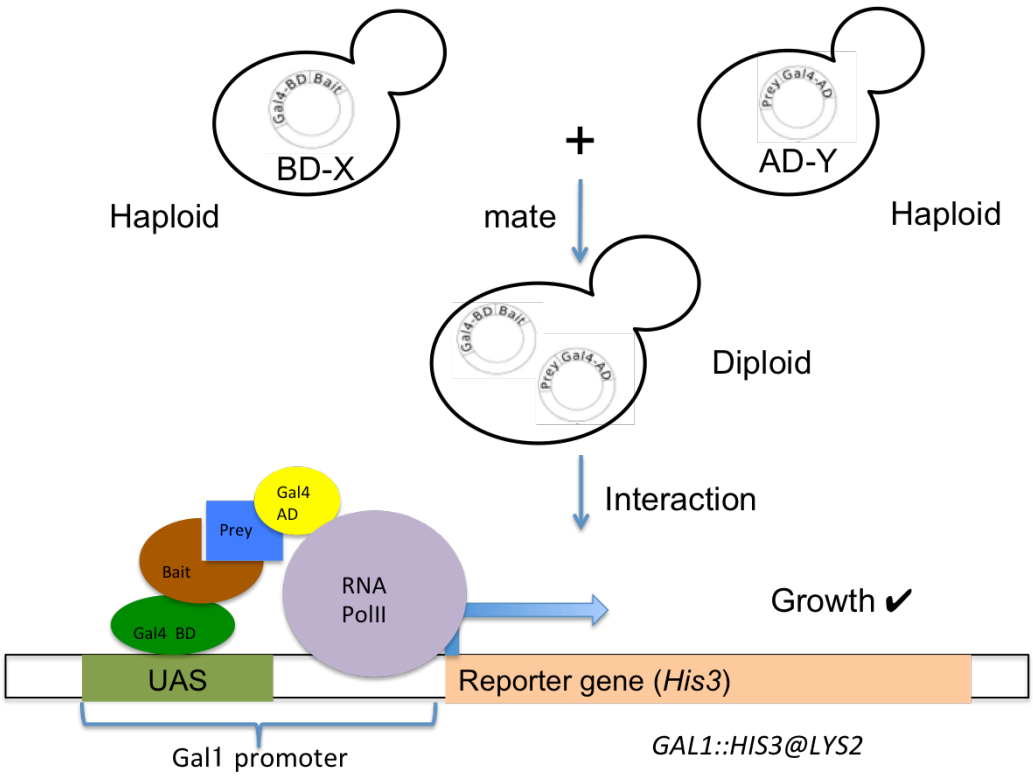


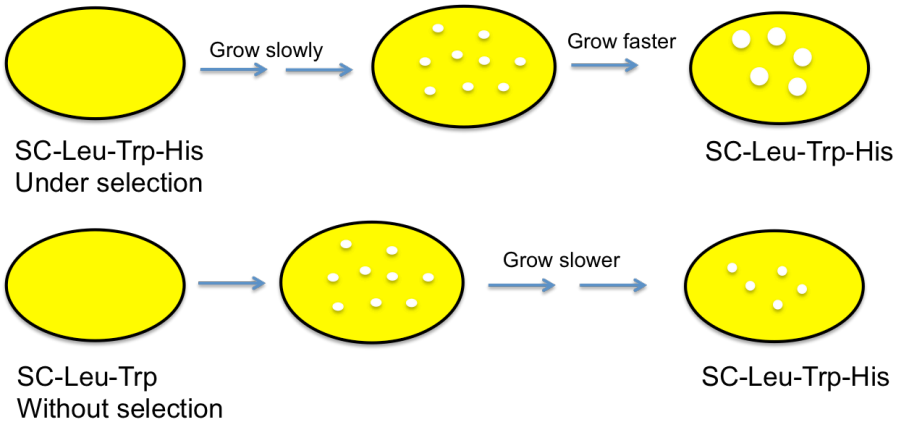
Figure 2. Yeast two-hybrid screen. A transcription factor (usually GAL4) is split into two separate parts, the binding domain (DB) and activation domain (AD), responsible for binding to the upstream activating sequence (UAS) and activating transcription of downstream reporter gene, respectively (Fields and Song, 1989). These two parts are fused to proteins of interest, X (also known as "bait" protein) and Y (also known as "prey" protein). When haploid cells with DB-X or AD-Y plasmids are mated, the diploid cells contain both DB-X and AD-Y plasmids. So physical interaction between DB-X and AD-Y could reconstitute the transcription factor and induce transcription of the reporter gene; thereby the diploid cells could grow on selective medium.

1.3.2 Y2H strains with and without 'papillation'

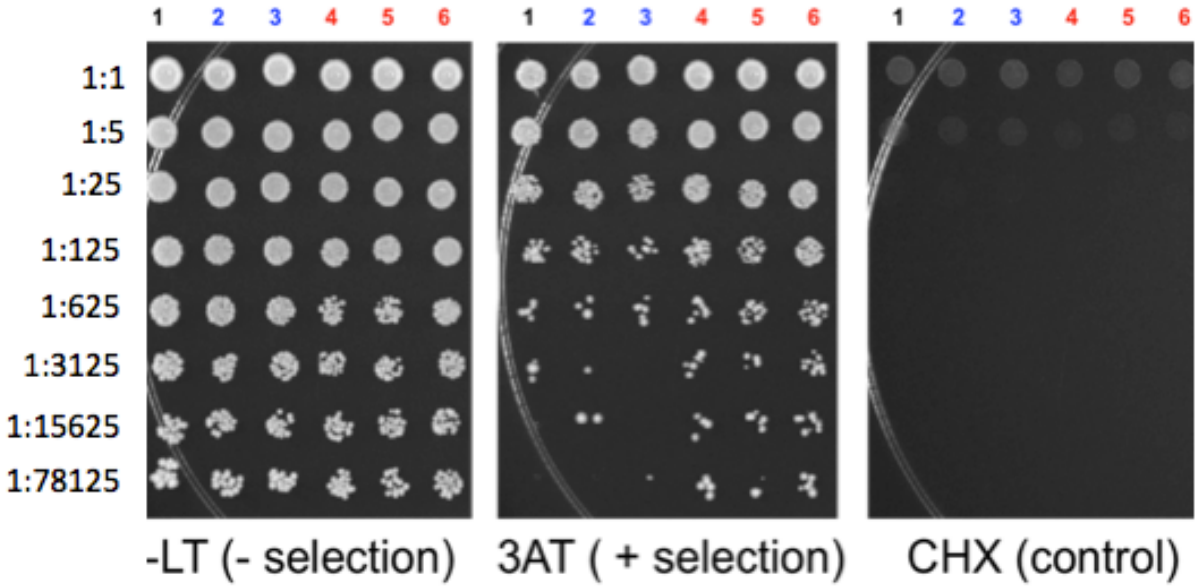
By tradition, papillation indicates the phenomenon of producing secondary colonies on the surface of the main colony. For example, colonies of *Bacillus anthracis* Sterne allow appearance of papillae with mutations after incubation for 6 days. The papillae resulted from spontaneous mutations that allowed the cells to continue growth. Papillation assay is also used as a tool for finding 'mutator' mutations that increase the mutation rate and generate elevated levels of papillation (Yang et al., 2011).

In our study, papillation has a different definition, but with a similar concept. When Y2H experiments were conducted using some specific pairs of proteins and with *HIS3* as the reporter gene, some colonies came out slowly after several days on His- medium under selection. But when these colonies were re-plated to another His- medium, they grew much faster than colonies from His+ medium without selection (Figure 3A). Similar to papillation in *Bacillus anthracis*, some spontaneous mutations might lead to stronger growth on His- medium. In Figure 3B, lane 1 represents parent diploid strains, lanes 2 and 3 represent colonies isolated from SC-Leu-Trp medium without selection (pre-selection populations), lanes 4-6 represent colonies from SC-Leu-Trp-His medium with selection (post-selection populations). On His- medium, pre-selection populations grew weaker than post-selection populations. The difference in growth rate is possibly related to some genetic differences. We hypothesized that parental strains contained mixed genotypes, some of them confer better growth, but others do not. When we isolated single colonies from His- medium, we basically picked from the former population; but when we isolated colonies from non-selective medium (His+, pre-selection populations), it is more likely that we picked from the dominant population in the parental mix. Therefore, we would like to find out what genotype gives rise to better growth on His- medium.

A



B



1. Parent diploid strain (Y8800+Y8930)
2. Colony isolated from -Leu-Trp prior to selection
3. Colony isolated from -Leu-Trp prior to selection
4. Colony isolated from 3AT after selection (-His)
5. Colony isolated from 3AT after selection (-His)
6. Colony isolated from 3AT after selection (-His)

Figure 3. Y2H strains with and without 'papillation'. Lane 1 represents parent diploid strains, lanes 2 and 3 represent colonies isolated from SC-Leu-Trp medium without selection, and lanes 4-6 represent colonies from SC-Leu-Trp-His medium with selection.

1.3.3 Priority regions for papillation strains

The stronger growth is likely caused by higher expression of the Y2H reporter gene. In these experiments, the reporter gene is *HIS3* driven by *GAL1* promoter located downstream of *LYS2* gene. There are some common regions where mutations might increase Y2H signals without specifically changing the dissociation constant of the target proteins X and Y. For instance, mutations in the origin region of the plasmids may increase plasmid copy number; mutations in promoter region of the plasmids could increase expression level of DB-X or AD-Y proteins; modified DB or AD protein sequence increases PPI; modified DB protein sequence recruits RNA PolII without AD-Y; mutations in X or Y sequence increases PPI; mutations related to *HIS3* increase expression of reporter gene by promoter mutation, stabilized mRNA or protein, translation efficiency, etc. Therefore, it is more likely that mutations in these regions could generate papillation strains.

1.3.4 Papillation strains contain specific pairs of prey/bait proteins

Although false positives in Y2H screens exist, this explanation is unlikely because papillation strains appear X/Y-dependent. One pair is DB-Ubc9 and AD-Smt3, involved in SUMO pathway (Hay, 2005). The Smt3/SUMO precursor is catalyzed by a SUMO specific protease to reveal the C-terminal dyglycine, which is activated by the E1 enzyme, Aos1p/Uba2p. After transesterification, Smt3 interacts with Ubc9, the E2 SUMO conjugation enzyme. Then the substrate is ligated with SUMO by an E3 ligase. Smt3/SUMO and Ubc9 are required for efficient proteolysis mediated by anaphase-promoting complex cyclosome (APCC) in yeast (Dieckhoff et al., 2004). So there might be some modifiers that change the extent of SUMO ligation, which may enhance Ubc9 and Smt3 interaction, and thus further increase *HIS3* expression.

Therefore, we are looking for some mutations that may regulate X/Y interaction pattern (more details see Section 3.2). This could reveal new genetic pathways for modulation of these protein-protein interactions.

2 Materials and Methods

2.1 Media and Yeast strains

Yeast haploid Ctf strains used in my study are from the Hieter laboratory (Spencer et al., 1990). The corresponding parental strains, YPH277, YPH278, YPH362, YPH363 and their genotypes are listed in Table 1. Yeast strains were grown in YPD (1% yeast extract, 2% peptone, and 2% glucose with or without 2% agar) or SC-Ura (0.2% Drop-out mix, 2% glucose and 0.67% yeast nitrogen base without amino acids). Heterozygous diploids were selected in SC-Ura-His-Trp medium. Pre-sporulation medium contains 10% glucose, 0.3% bacto peptone and 0.8% yeast extract. Sporulation medium contains 0.5% potassium acetate and 0.02% amino acids supplement (His+, Trp+).

Amino acids supplement powder mixture for complete synthetic medium contains 3g adenine, 2g uracil, 2g myo-inositol, 0.2g p-aminobenzoic acid, 2g alanine, 2 arginine, 2g asparagine, 2g aspartic acid, 2g cysteine, 2g glutamic acid, 2g glutamine, 2g glycine, 2g histidine, 2g isoleucine, 10g leucine, 2g lysine, 2g methionine, 2g phenylalanine, 2g proline, 2g serine, 2g threonine, 2g tryptophan, 2g tyrosine and 2g valine (Tong and Boone, 2006). Drop-out mixture contains all above components except the appropriate supplements.

Yeast diploid papillation strains are from the Vidal laboratory (personal communication), deriving from mating Y8800 and Y8930 (Table 1). Yeast strains were gown in SC-Leu-Trp-His or SC-Leu-Trp medium.

Vitamin supplement ponder mixture contains 0.4g calcium pantothenate, 0.02g folic acid, 0.4g niacin, 0.4g pyridoxine hydrochloride, 0.02g biotin.

Complete synthetic medium for inositol auxotrophy contains 2% glucose, 0.67% Difco

vitamin-free yeast nitrogen base, 20mg/L vitamin supplement mix, 0.2% amino acid drop-out mix without inositol, with or without myo-inositol (20mg/L).

Strain	Genotype	Source
YPH277	*MATa ura3-52 lys2-801 ade2-101 trp1Δ1 leu2Δ1 CFVII (RAD2.d. YPH277) URA3 SUP11*	Hieter laboratory
YPH278	*MATα ura3-52 lys2-801 ade2-101 his3Δ200 leu2Δ1 CFIII (CEN3.L.YPH278) URA3 SUP11*	Hieter laboratory
YPH362	*MATa ura3-52 lys2-801 ade2-101 trp1Δ1 leu2Δ1 CFVII (RAD2.d. YPH362) URA3 SUP11*	Hieter laboratory
YPH363	*MATα ura3-52 lys2-801 ade2-101 his3Δ200 leu2Δ1 CFIII (CEN3.L.YPH363) URA3 SUP11*	Hieter laboratory
Y8800	*MATa, leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ GAL2-ADE2 LYS2::GAL1-HIS3 MET2::GAL7-lacZ cyh2R can1Δ::pCMV-rtTA-KanMX4*	Boone laboratory
Y8930	*MATα, leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ GAL2-ADE2 LYS2::GAL1-HIS3 MET2::GAL7-lacZ cyh2R can1Δ::tADH1-tetO2-Cre-tCYC1-KanMX4*	Boone laboratory

Table 1. Genotypes for parental strains.

2.2 Illumina Sequencing libraries construction and whole-genome sequencing

After growing on YPD plates at 30°C for 2 days, single colonies from 14 Ctf strains were picked and placed into liquid YPD medium for another day of inoculation. Genomic DNA (gDNA) from each strain was extracted using YeaStar Genomic DNA Kit by ZYMO Research, and quantified by Qbit fluorimeter. Genomic DNA (~6ng) was sheared and tagmented to ~300bp fragments with adapters at each end by an engineered variant of the Tn5 transposon's *transposase* enzyme (Figure 4). Then each tagmented library was PCR-amplified to add a unique molecular barcode and the sequencing primers. Libraries were purified by AMPure beads, quantified by qPCR (KAPA Library Quantification Kit) and mixed to ensure similar representation of each sample. The whole-genome sequencing was performed on an Illumina HiSeq 2500 instrument according to the manufacturer's recommended guidelines.

Twenty-four papillation strains and two non-papillation strains (details see section 3.2.1) were picked to get whole-genome sequencing with similar steps. To get rid of DNA fragments less than 300bp, the gel was cut out between 300bp-1000bp to get more specific amplification. The whole-genome sequencing was performed on Illumina Nextseq platform.

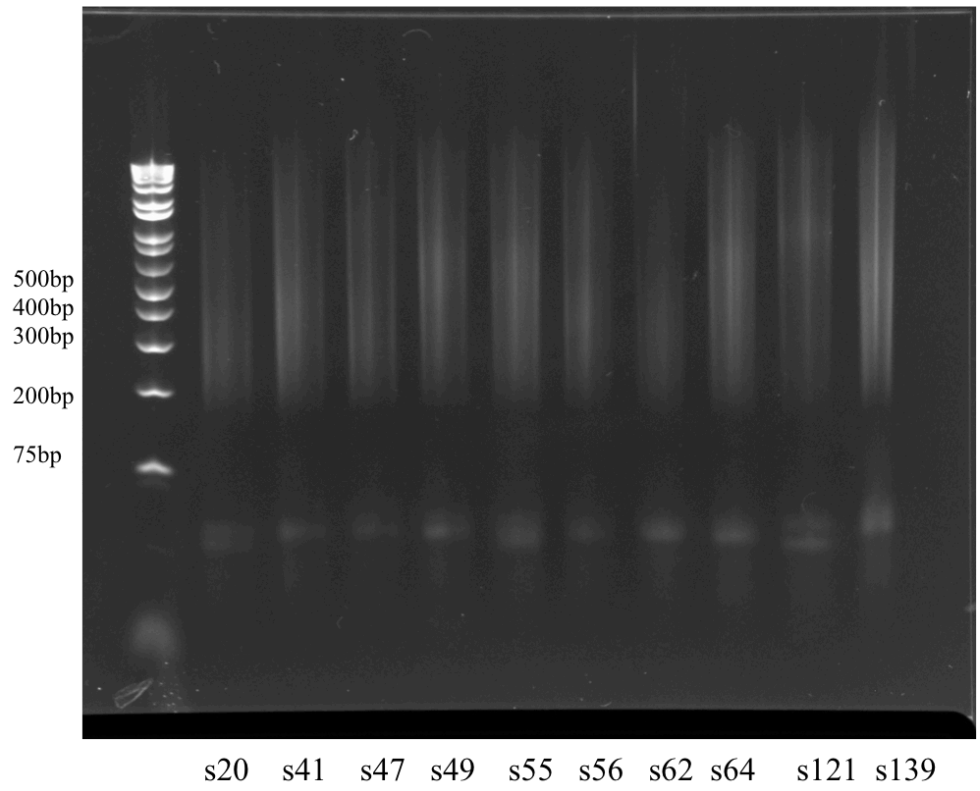


Figure 4. Genomic libraries of some Ctf strains visualized by gel. Genomic DNA (gDNA) from each strain was extracted using YeaStar Genomic DNA Kit by ZYMO Research, and quantified by Qbit fluorimeter. Genomic DNA (6ng) was sheared and tagmented to ~300bp fragments with adapters at each end by an engineered variant of the Tn5 transposon's *transposase* enzyme. Then each tagmented library was PCR-amplified to add a unique molecular barcode and the sequencing primers.

2.3 Computational analysis of Illumina sequencing data

Sequencing data of Ctf strains generated by the Illumina HiSeq platform was paired-end reads with inserts between 200-400bp. Data was processed using a computational pipeline assembled by T. Yamaguchi (Figure 5). It began with base-called reads with quality scores from the Illumina pipeline. First, the sequencing reads was mapped to the S288C reference genome (SacCer3) from SGD by Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010). Then sequence Alignment/Map (SAM) tools (Li et al., 2009) was applied to identify SNVs and micro-indels with a relatively stringent threshold (Phred score>17, read depth>=10, and >=90 of high-quality aligned reads support the mutations), and Pindel was applied to find large deletions, medium sized insertions, inversions, tandem duplications and other structural variations (Ye et al., 2009). Any common mutations identified in more than one strains were removed, since all strains were selected for adaptation independently making common mutations unlikely.

Furthermore, under light mutagenesis condition in the primary screen, it is assumed that Ctf mutants resulted from mutations in single genes, which makes follow-up process easier. The candidate mutations were narrowed down based on two approaches. The first basis is the previous association of gene products with the Ctf phenotype. I checked if some candidate genes on my list had already been identified as Ctf genes. Also, amino acid substitutions were analyzed and deleterious SNVs were predicted according to BLOSUM80 score (Henikoff and Henikoff, 1992), Polyphen2 (Adzhubei et al., 2010), and Provean (Choi et al., 2012) (Table 2).

Copy number variation (CNV) was estimated based on coverage for each strain. After normalization, the median coverage across 200bp windows in each strain is 1, and the median coverage of each window across all strains is also 1. Therefore, the expected coverage for the 200bp window within a single copy gene, a deleted gene, and a duplicated gene will be 1, 0 and 2, respectively.

For papillation strains, since the pipeline described above only applies to haploid genome, I adjusted some steps. The stringent threshold filled out all heterozygous mutations, so I extracted mutations with a relatively permissive threshold (Phred score >17, read depth >= 10 and >= 20% of high-quality aligned reads support the mutations). In case there are some hotspots where mutations seem more likely *a priori* to increase Y2H interaction signal, I extracted not only unique mutations but also mutations with incidence ≤ 2/26 to capture recurrent mutations.

Then I used ANNOVAR (Wang et al., 2010) to annotate all mutations to distinguish heterozygous mutations from homozygous mutations.

Finally, genome tracks were generated from the output of this pipeline and visualized by ChromoZoom (Pak and Roth, 2013).

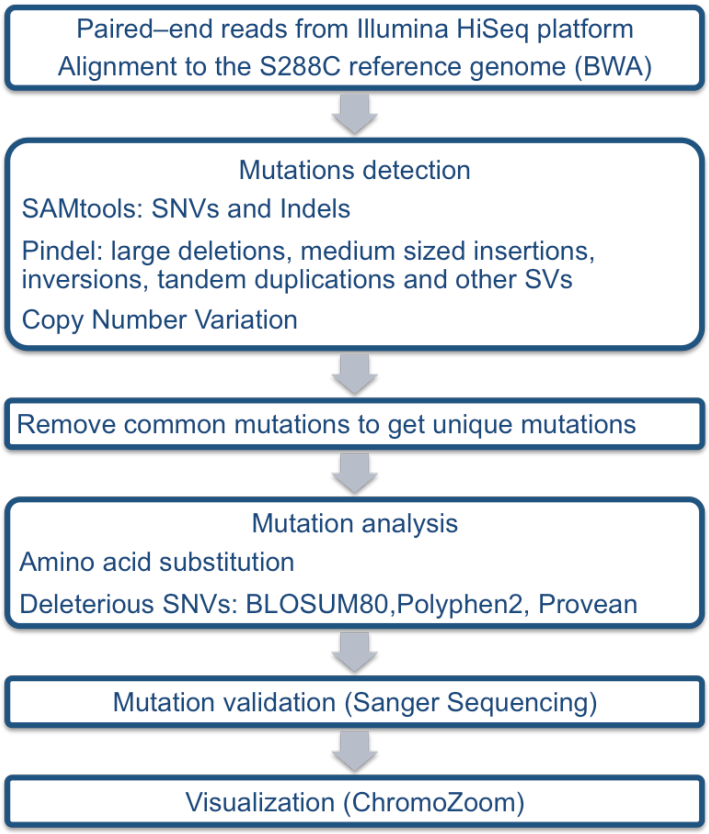


Figure 5. Computational workflow for processing Illumina sequencing reads (adapted from Takafumi Yamaguchi). Sequencing data generated by the Illumina HiSeq platform was paired-end reads on inserts between 200-400bp. First, the reads were mapped to the S288C reference genome from SGD by Burrows-Wheeler Aligner (BWA). Then sequence Alignment/Map (SAM) tools was applied to identify SNVs and micro-indels with a relatively stringent threshold, and Pindel was applied to find large deletions, medium sized insertions, inversions, tandem duplications and other structural variations. Any common mutations identified in more than one strains were removed. After identifying unique mutations, amino acid substitution was analyzed and deleterious SNVs were predicted. Copy number variation (CNV) was estimated based on coverage for each strain. Finally, genome tracks were generated and visualized by ChromoZoom (Pak and Roth, 2013).

2.4 Sanger sequencing

Sanger sequencing was performed to validate candidate SNV/micro-indels identified by whole-genome sequencing analysis. UCSC Genome Browser and primer3 were used to design PCR primers to ensure specific amplification from primers flanking the site of the putative mutation. The same purified gDNA used for Illumina sequencing was used to amplify the target mutations. PCR products were purified by Zymo DNA clean&concentration Kit, and the DNA concentrations were measured by Qbit. The Sanger sequencing of PCR samples was performed by the TCAG DNA Sequencing Facility.

For validation by tetrad analysis, colony PCR from individual spores was performed to amplify the DNA fragment covering the mutations. A full size colony was dissolved in 200µl ddH_2O and 5µl of mixture was added to 20µl Zymolyase lysis buffer (1% 5U/µl in sodium phosphate buffer). Cells were incubated at 37°C for 3 hours followed by 95°C for 15min. 75ul ddH_2O was added to dilute mixture. After spining down, supernatant lysates were used for PCR amplification.

2.5 Experimental validation: tetrad analysis and plasmid transformation rescue

First, tetrad analysis was applied to validate some Ctf candidate genes. Ctf mutant haploids were mated to the corresponding parental haploids. According to the genotypes, heterozygous diploids were selected on SC-Ura-His-Trp medium.

To induce sporulation, heterozygous diploids were incubated in presporulation medium for 6 hours starting at OD_{600}=1 when cells were at exponential phase. After spinning down, cells were resuspended in sporulation medium. Then the culture was diluted to OD_{600}=1, and was incubated at room temperature for 1 week.

After asci represented more than 10% of the population of cells/asci, tetrad dissection was performed to separate spores (Figure 6). Centrifugation of 500µl sporulation medium for 1min was performed to spin down cells, then 50µl Zymolyase 20 (0.25mg/ml) was added without mixing. Cells were incubated at 30°C for 10min without shaking to digest asci. Then 150µl H_2O was added and mixed very gently in the middle without touching cells, and 10µl cells was moved to a YPD plate to perform tetrad dissection under microscope. This could

separate mutations between the spores randomly. Phenotype and genotype for the specific variants for colonies from each spore were checked to see which mutation always tracked with the Ctf phenotype.

An alternative validation strategy was ‘plasmid transformation rescue’. In the original screen, only 20 of 136 mutants showed detectable but incomplete dominance, and others showed recessive behavior, which means the ectopically expressed wild-type gene could provide at least some rescue of the phenotype caused by endogenous mutations in all cases. So a 2-micron plasmid (MoBY-ORF library (Ho et al., 2009)) with the wild-type allele of the candidate Ctf gene was transformed into the Ctf mutant and red sectoring phenotype was scored on SC+1/5th Ade medium. A strain with Ctf phenotype shows red sectors in white colonies at a frequency greater than the average of a wild-type strain on SC+1/5th Ade medium (usually 2-3 sectoring colonies out of 100). Also, once the colonies have grown big enough, it helps to keep the plate at 4°C for at least 5 days before scoring the Ctf phenotype, which facilitates red color development and visibility. If transformation with the wild-type allele of the candidate gene could rescue the Ctf phenotype, this indicates that the candidate gene is a Ctf gene.

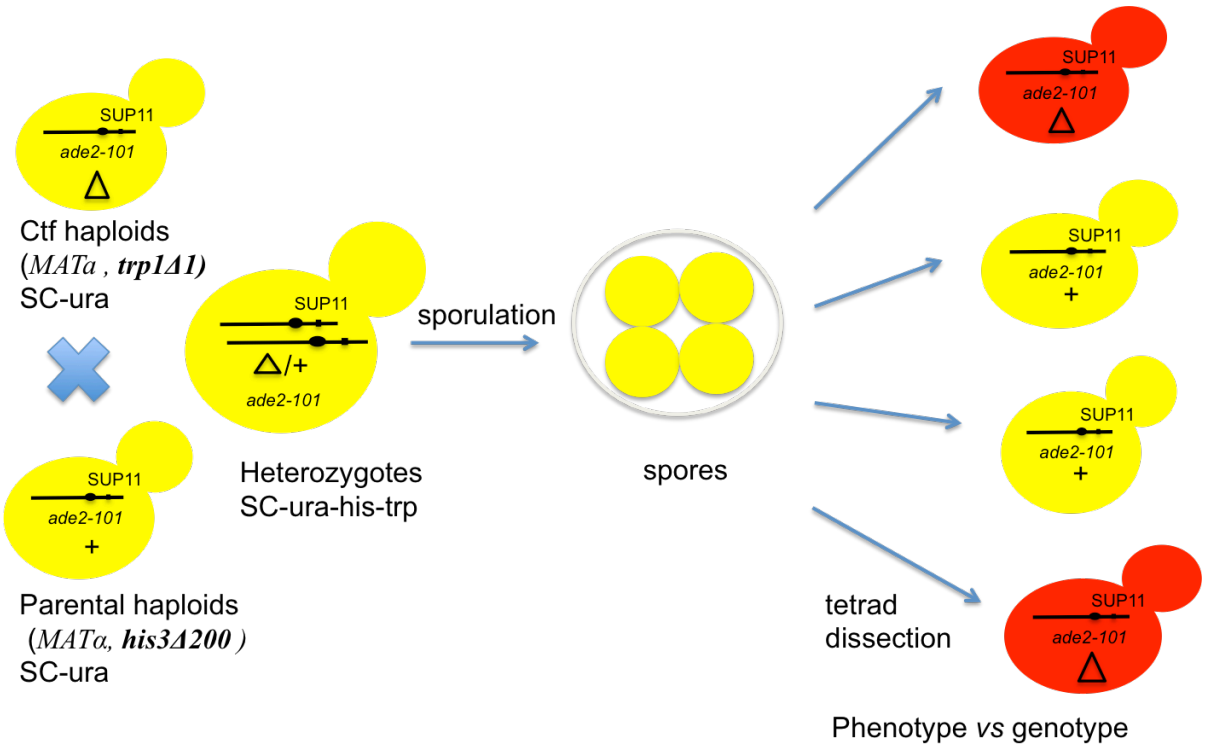


Figure 6. Tetrad analysis to validate Ctf candidate genes. Ctf mutant haploids were mated to the corresponding parental haploids. According to the genotypes, heterozygous diploids were selected on SC-Ura-His-Trp medium. Then sporulation was induced and tetra dissection was performed to separate spores. This could separate mutations between the spores. Phenotype and genotype for the specific variants for colonies from each spore were checked to see which mutation always tracked with the Ctf phenotype.

2.6 Spotting assay

To test the hypothesis of an interdependent relationship between inositol level and SUMO, inositol auxotrophy of strains with or without papillation containing DB_*UBC9* and AD_*SMT3* vectors was checked: E01_1, E02_1, E03_1, E04_1, E05_1, E06_1, F01_1, F02_1, F03_1, F04_1, F05_1, F06_1 by a standard spotting assay. I also picked a set of strains with DB_*GRR1* and AD_*SKP1* for negative control. In addition, *ino1Δ* deletion strain was used for positive control.

All strains grew overnight in 5ml liquid medium (papillation strains in SC-Leu-Trp medium, and *ino1Δ* strain in YPD medium) at 30°C. Then I measured OD_{600} and diluted cultures to OD_{600} = 1 with the same medium. Fivefold dilution series were made in a 96-well plate. Then 5µl dilutions were dropped onto inositol+ or inositol- plates using a multipipette. After the plates dried completely, the plates were incubated at 30°C. Photographs were taken after 2 days of incubation.

3 Results

Through computational analysis of whole genome sequencing data and following-up experimental validation, I identified some candidate causal mutations for both Ctf strains and papillation strains. I will describe the results for Ctf strains firstly, and then I will describe papillation strains in the second part.

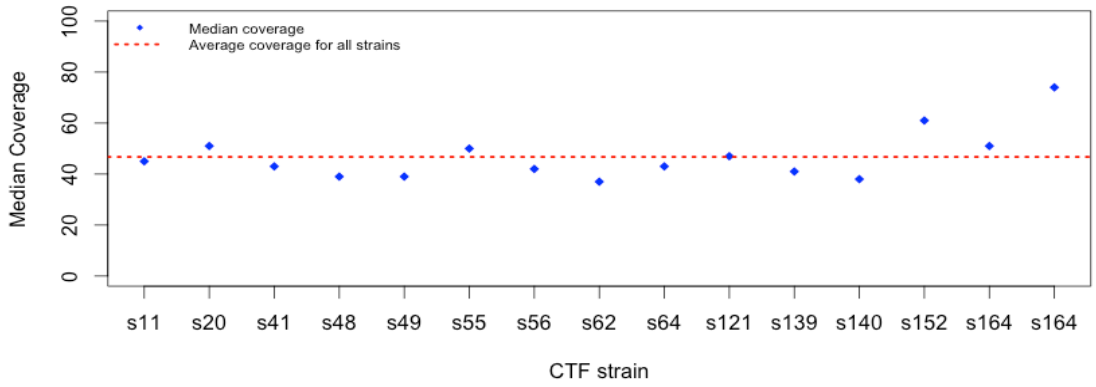
3.1 Ctf strains

The original Ctf mutagenesis screen was preformed by the Hieter lab. I received 14 Ctf mutant yeast strains from his lab. We aimed to identify causal mutations for the remaining uncloned mutants.

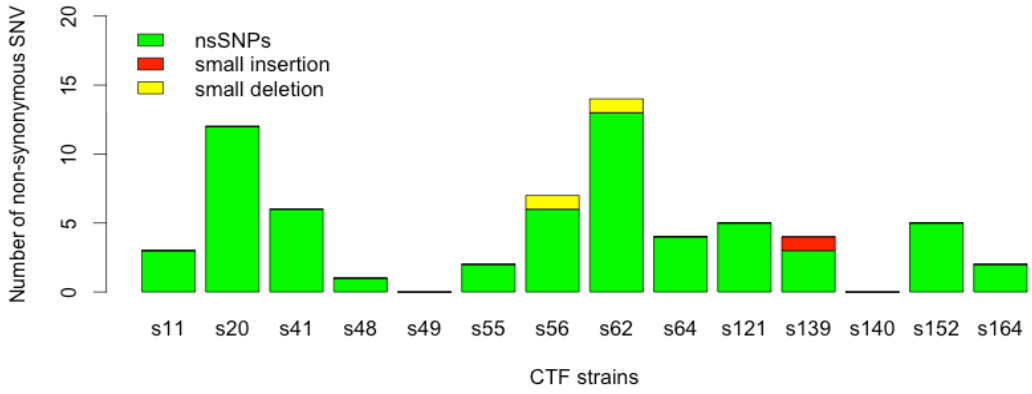
3.1.1 Mutations detected by whole-genome sequencing analysis

After mapping to the reference genome, the mean of median coverage of sequenced genomes for each strain was 46.7, which was sufficient for analysis of variations based on previous experience (Figure 7A). After removing all common SNVs, number of unique non-synonymous SNV for each strain ranged from 0 to 14, at a mean of 4.6 (Figure 7B). Number of unique structural variation detected by Pindel ranged from 1 to 14, at a mean of 6.4 (Figure 7C). Most SVs in coding regions located within single genes, except 4 large duplications (Figure 7D). Genome tracks were displayed by Chromozoom (URL see Table 2). For these mutations, I prioritized them based on functions of gene products and deleterious prediction. Based on previous studies including both forward-genetic and systematic reverse-genetic screens, a list of 354 mutants with Ctf phenotypes has been identified, covering 237 essential genes and 117 nonessential genes (Measday et al., 2005; Stirling et al., 2011; Yuen et al., 2007). Thus, I identified some candidate Ctf genes based on computational analysis of whole-genome sequencing.

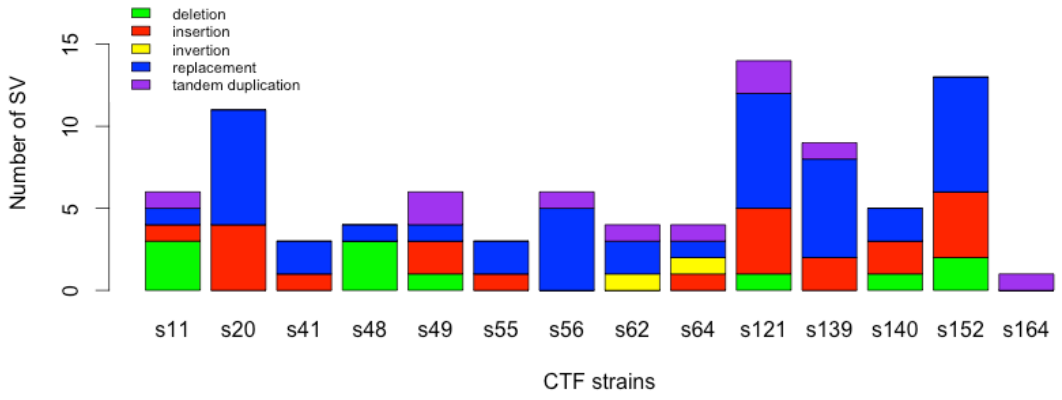
A


Median coverage
Average coverage for all strains
Median Coverage
0
20
40
60
80
100
s11
s20
s41
s48
s49
s55
s56
s62
s64
s121
s139
s140
s152
s164
s164
CTF strain

B


nsSNPs
small insertion
small deletion
Number of non-synonymous SNV
0
5
10
15
20
s11
s20
s41
s48
s49
s55
s56
s62
s64
s121
s139
s140
s152
s164
CTF strains

C



D

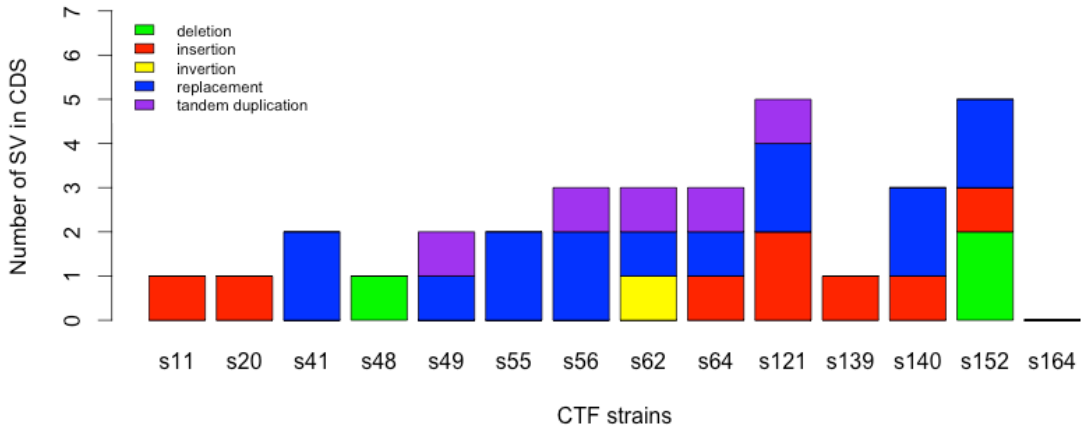


Figure 7. Mutations detected by the computational pipeline. 7A: The mean of median coverage of sequenced genomes for each strain was 46.7. 7B: Number of unique non-synonymous SNV for each strain ranged from 0 to 14, at a mean of 4.6. 7C: Number of unique structural variation detected by Pindel ranged from 1 to 14, at a mean of 6.4. 7D: Most SVs in coding regions located within single genes, except 4 large duplications.

Ctf strain	Chromozoom track
s11	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s11.txt
s20	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s20.txt
s41	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s41.txt
s48	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s48.txt
s49	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s49.txt
s55	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s55.txt
s56	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s56.txt
s62	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s62.txt
s64	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s64.txt
s121	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s121.txt
s139	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s139.txt
s140	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s140.txt
s152	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s152.txt
s164	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/CTF/s164.txt

Table 2. Web links to genome tracks for Ctf strains.

3.1.2 Candidate Ctf genes

Seven out of 14 strains carried mutations in known Ctf genes (Table 3). For instance, strain s11 carried a predicted-deleterious (according to Provean) point mutation in *BIM1*, whose product is a microtubule-binding protein; s20 had a nonsense mutation in *CHL4*, and its product is a component of kinetochore; and s164 had a probably-damaging (according to Polyphen2) point mutation in *MCM5*, related to DNA replication. The result was consistent with the previous conclusion that most Ctf genes are related to microtubule, kinetochore or DNA replication. The mutations we discovered in Ctf genes were predicted mostly as 'probably deleterious' (87.5%, 7/8), as compared with 45% (26/58) of mutations in genes not known to be Ctf genes (according to Provean).

For other strains without mutations in known Ctf genes, it represented chance to find new Ctf genes. I explored all variants including non-coding region and synonymous SNPs, and found evidence supporting the Ctf phenotype. For example, there was a deleterious point mutation, G to A, at *MCM7* in s152 strain (according to Provean). This gene is reported to cause chromosome loss (Stirling et al., 2011), and its product is a component of the hetero-hexameric MCM2-7 complex, which primes origins of DNA replication in G1. Since *MCM7* is very similar to *MCM5*, a known Ctf gene, the former may be a novel Ctf gene. Finally, there were several candidate genes for each unexplained strain (Table 4). However, there was no validated mutation for two strains, s48 and s49. For these strains, Ctf phenotype may result from copy number variation or large structural variations, although we did not identify specific strong candidate positions of structural variation.

Strain	Position	Alt/Ref	Amino acid (Ref.Pos.Alt)	BLOSUM80	Polyphen2 (cutoff=0.5)	Provean (cutoff=-2.5)	Gene	Function or cellular complex
s11	chrV:188373	A/G	G33R/345	-3	1	-7.696	*BIM1*	Microtubule/cohesion
s20	chrIV:965788 chrXIV:346977	T/C T/C	Q226*/459 P222L/410	-6 -3	- 0.002	- -0.716	*CHL4* *INN1*	Outer kinetochore Actomyosin ring
s41	chrXVI:799508	A/G	R92K/928	2	0.009	-2.742	*CTF4*	Cohesion/DNA replication
s56	chrVI:209552	A/G	V173M/552	1	1	-2.98	*CDC14*	Protein phosphatase/mitotic
s62	chrVII:337406	T/C	Q841*/918	-6	-	-	*SPC105*	Kinetochore-microtubule
s121	chrXII:287288	T/G	E157*/285	-6	-	-	*SIC1*	G1/S transition regulator
s164	chrXII:692951	A/G	G466D/776	-2	1	-6.483	*MCM5*	DNA replication

Table 3. 7 strains with mutation in known Ctf genes.

Strain	Position	Alt/Ref	Type	BLOSUM80	Polyphen2 (cutoff=0.5)	Provean (cutoff=-2.5)	Candidate genes	Function or cellular complex
s55	chrVII: 447980	T/C	nsSNP(V186I)	3	0.181	-0.827	*TRP5*	Tryptophan synthase
s64	chrIX: 241364	T/C	nsSNP(V138I)	3	0.002	-0.359	*FIS1*	Mitochondrial membrane
	chrXII:205725	T/C	sSNP	-	-	-	*RAD5*	DNA replication
s139	chrIV:1288791	A/C	nsSNP(K23N)	0	0.01	-1.254	*ADE8*	Purine nucleotide biosynthetic
	chrXIV:170994	A/G	nsSNP(S326N)	0	0.765	-1.231	*TEX1*	mRNA export
s140	chrXIII: 90860	TTTTCGG/ TGTGA	RPL	-	-	-	*YML090W*	Unknown
s152	chrXII: 188576	G/C	S1569*/1873	-6	-	-	*UBR2*	Ubiquitin-protein ligase
	chrXII: 369920	A/C	nsSNP(T60N)	0	1	-4.592	*CCW12*	Cell wall
	chrII: 626622	A/G	nsSNP(C284Y)	-3	1	-10.962	*MCM7*	DNA replication

Table 4. 5 strains without mutation in known Ctf genes.

3.1.3 Three candidate genes were validated experimentally

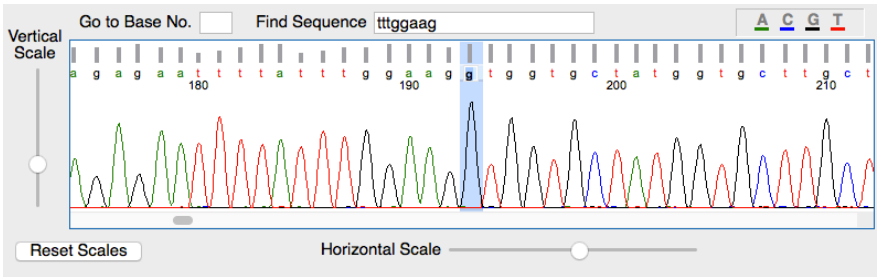
Meanwhile, 29 mutations were assessed by Sanger sequencing. Among them, 17 mutations were validated. Eight mutations were proved to be false positives (7 false positive structural variations came from Pindel detection and one SNV was from permissive threshold in pipeline (Phred score > 10, read depth >= 3 and >= 50% of high-quality reads support the mutations)), 2 mutations failed to be amplified because of repeat sequence regions, and 2 mutations were ruled out because they were the same as each other (for details see section 7.1). After mating to corresponding parental strains, heterozygous mutations was also checked by Sanger sequencing, and 100% were confirmed to agree with the original call (Figure 8).

Strains s11, s62, and s121 did not grow on SC-Ura medium, indicating they had lost the *URA* marker. After tetrad dissection of the remaining strains, some strains did not generate any colonies with Ctf phenotype. For example, after growing on YPD medium, colonies from s41 spores displayed two pure (non-sectoring) phenotypes: red and white. In Figure 9, each column represents 4 colonies derived from one tetrad. But while white colonies with red sectoring indicate increased chromosome loss rate, red colonies only represent that these colonies do not carry the artificial chromosome. The artificial chromosome loss could have happened, for example, during meiosis. The second column from the left was discarded due to death of half of the spores. After re-plating these colonies and Sanger sequencing for *CTF4*, colonies marked as circle and diamond carried *CTF4* mutation and displayed typical Ctf phenotype, and colonies marked as circle had *CTF4* mutatation but displayed either pure red or white phenotypes. All of the rest four colonies with a white phenotype initially carried wild-type *CTF4* and did not have sectoring phenotype.

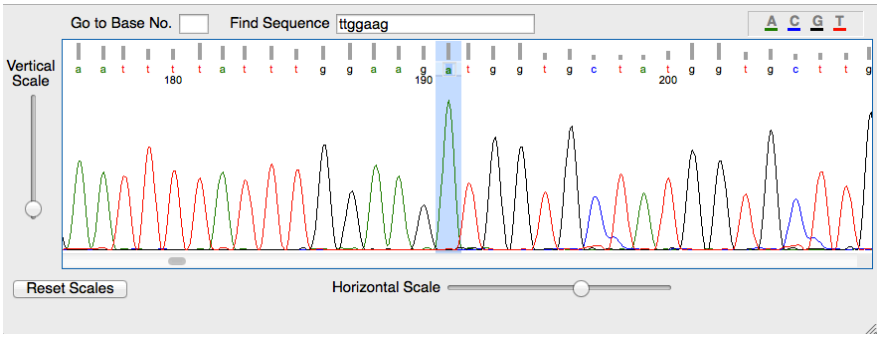
Colonies from s164 spores displayed the typical Ctf phenotype (Figure 9B). By Sanger sequencing, *MCM5* varients were validated in 7 out of 8 colonies with red sectoring (cell patches marked with circle), and one colony with red sectoring has wild-type *MCM5* (colony marked with diamond).

When attempting plasmid transformation rescue there were 2 positive hits (cases where expression of the wild type gene rescues), *CTF4* for s41 and *SIC1* for s121.

A



B



C

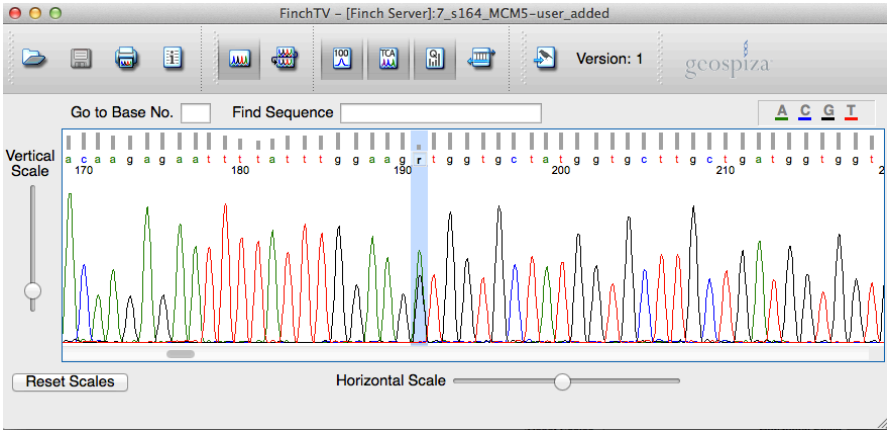
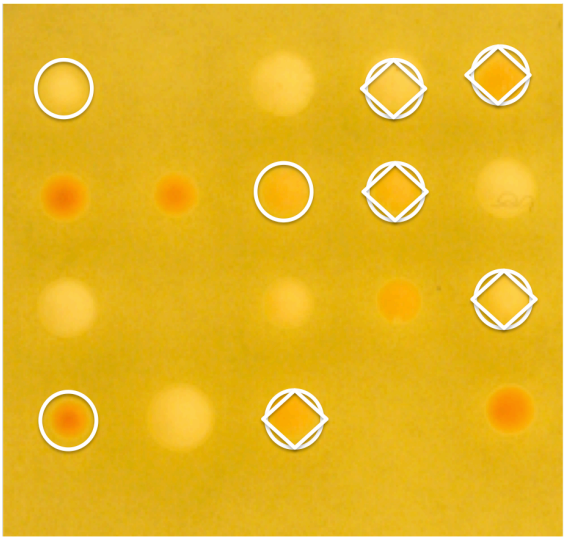


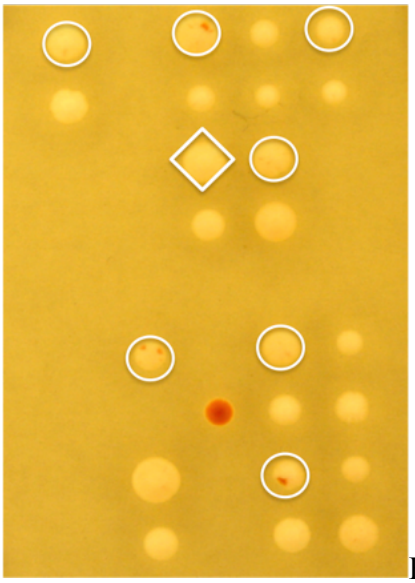
Figure 8. One example of Sanger sequencing. The strains s164 has a G→A point mutation in *MCM5*. A. reference; B. mutant in a haploid strain; C. heterozygous mutant in a diploid strain.

CTF strain	Growth in SC-Ura	Growth in YPD	Parental strains	CTF phenotype
s11	NO	YES	YPH278	YES
s20	YES	YES	YPH278	WEEK
s41	YES	YES	YPH278	YES
s48	YES	YES	YPH278	YES
s49	YES	YES	YPH278	YES
s55	YES	YES	YPH278	YES
s56	YES	YES	YPH278	YES
s62	NO	YES	YPH278	YES
s64	YES	YES	YPH278	YES
s121	NO	YES	YPH277	YES
s139	YES	YES	YPH362	NO
s140	YES	YES	YPH362	YES
s152	YES	YES	YPH362	YES
s164	YES	YES	YPH362	YES

Table 5. Ctf phenotype rescoring on obtained Ctf strains.



A



B

Figure 9. Experimental validations by tetrad analysis. Each column represents 4 colonies derived from one tetrad. 9A: After re-plating these colonies and Sanger sequencing for *CTF4*, colonies marked as circle and diamond carried *CTF4* mutation and displayed typical Ctf phenotype, and colonies marked as circle had *CTF4* mutatation but displayed either pure red or white phenotypes. All of the rest four colonies with a white phenotype initially carried wild-type *CTF4* and did not have sectoring phenotype. 9B: Colonies from s164 spores displayed the typical Ctf phenotype. By Sanger sequencing, *MCM5* varients were validated in 7 out of 8 colonies with red sectoring (cell patches marked with circle), and one colony with red sectoring has wild-type *MCM5* (colony marked with diamond).

3.1.4 Network profile

Among the original 136 Ctf mutants, >65% of the mutants were mapped to 13 genes, and 12 of these 13 genes were involved in sister chromatid cohesion and/or kinetochore functions. This functional enrichment suggests that these genes are more easily mutable to yield the Ctf phenotype. Examination of the 26 genes that were previously identified from original Ctf screen in GeneMANIA showed that these genes (black nodes in Figure 10) formed a highly connected network through physical interaction, genetic interaction, co-expression, or other relationships, providing further evidence of shared functions amongst Ctf genes. Hence, genes in this network that have not yet been identified in the Ctf forward mutagenesis screen (grey nodes in Figure 10) represent candidate Ctf genes. Indeed some of these have been seen in reverse genetic studies (for example, *BUB1*, *BUB3*, *MAD2* (Measday et al., 2005; Stirling et al., 2011; Yuen et al., 2007)).

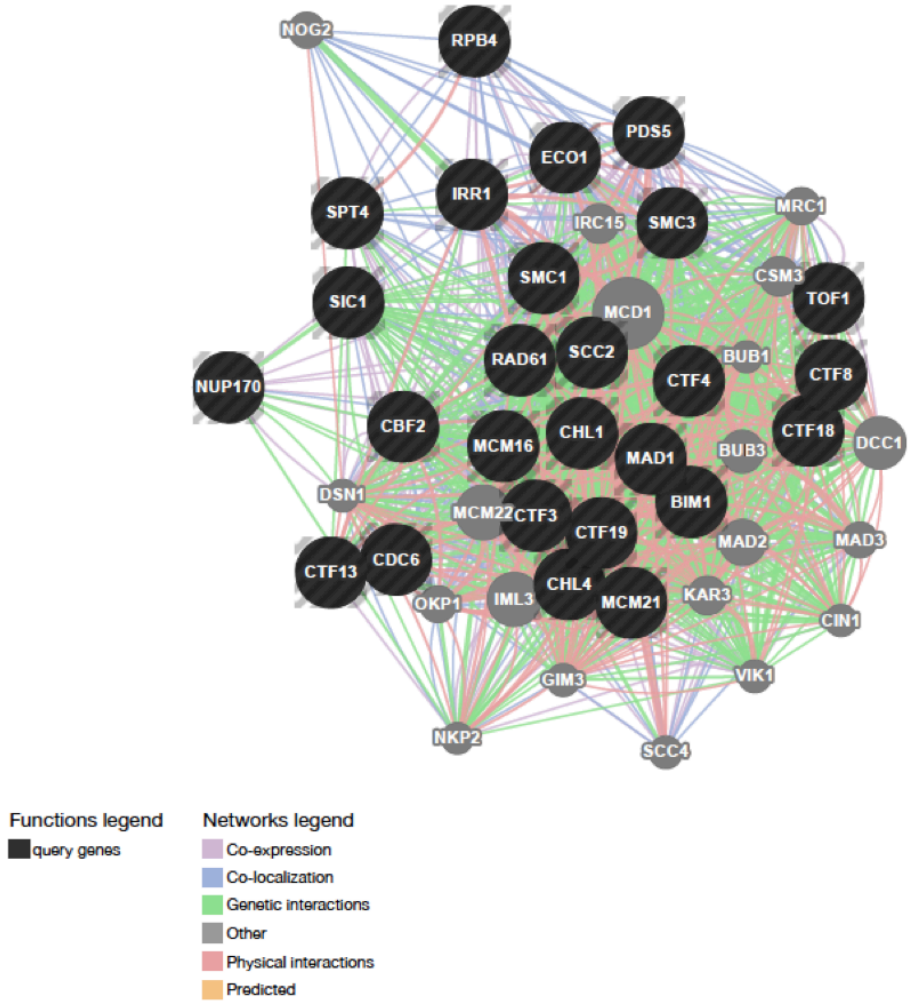


Figure 10. Ctf genes network profile. Network profile from GeneMANIA showed that Ctf genes formed a highly connected network. Black nodes represent genes identified from original Ctf screen (query genes), and grey nodes represent genes that have interactions with query genes. GeneMANIA analysis was performed with default options.

3.2 Papillation strains

Another class of mutants have a papillation phenotype that are adapted for stronger growth on selection medium in Y2H experiments. This stronger growth is likely due to higher expression of the reporter gene, which indicates a higher Y2H signal. Therefore, we tried to identify potential genetic modifiers of these PPIs.

3.2.1 Two growth types of papillation strains

When grown on SC-His medium, there are two distinct growth phenotypes of papillation strains. For type I, pre-selection populations tend to be no-growth, while for type II pre-selection populations tend to be weaker growth, similarly to parent strains (Figure 11). For each type, we picked up 3 post-selection populations from 4 groups, and 1 pre-selection population from 1 group (each group has unique pair of vectors). Each strain was named after their original coordinate: #row_#column_#plate. In total, we sequenced 26 populations to get whole-genome sequencing data, including two pre-selection populations. These strains contain 8 pairs of different vectors, corresponding to 7 distinct pairs of X and Y proteins.

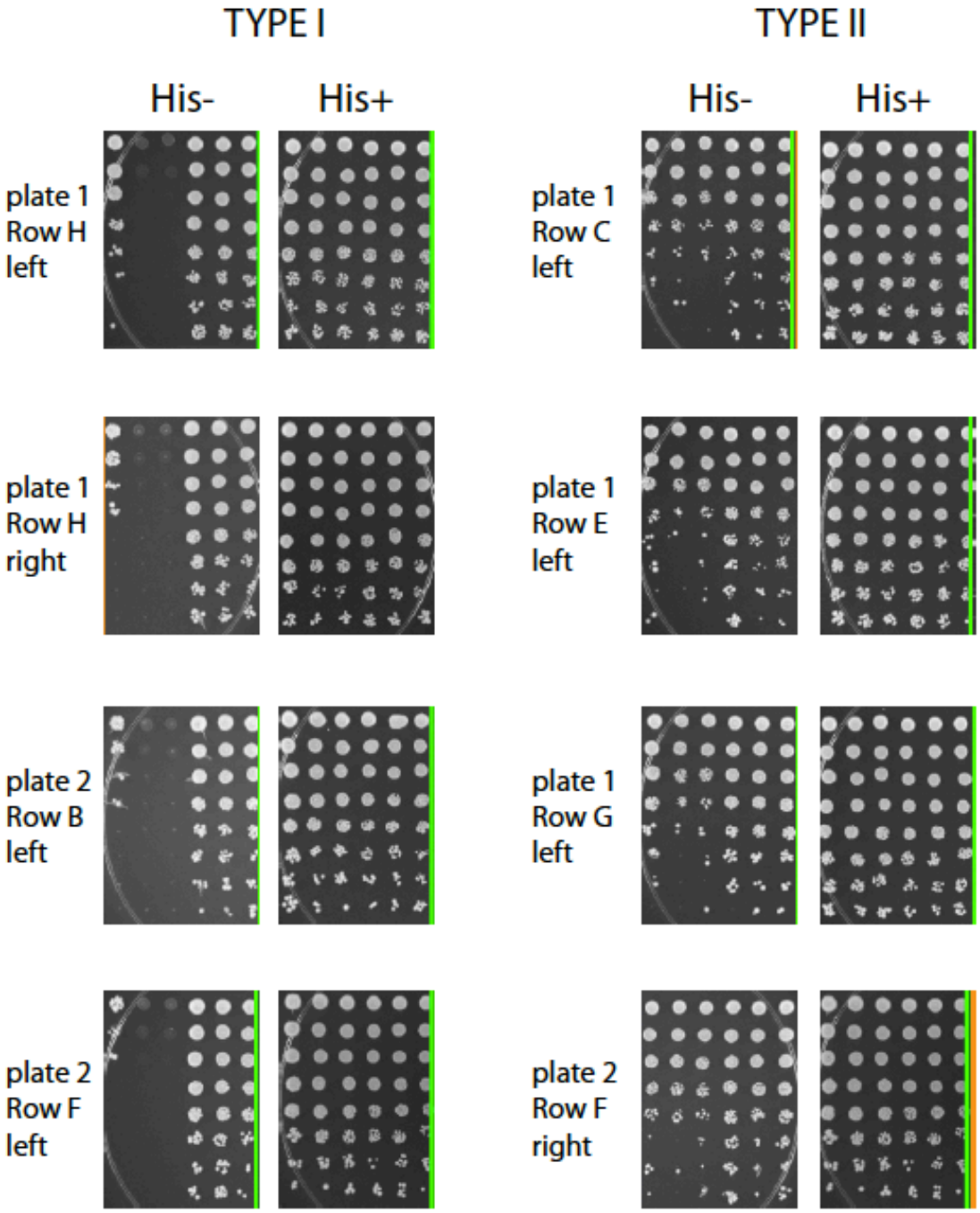


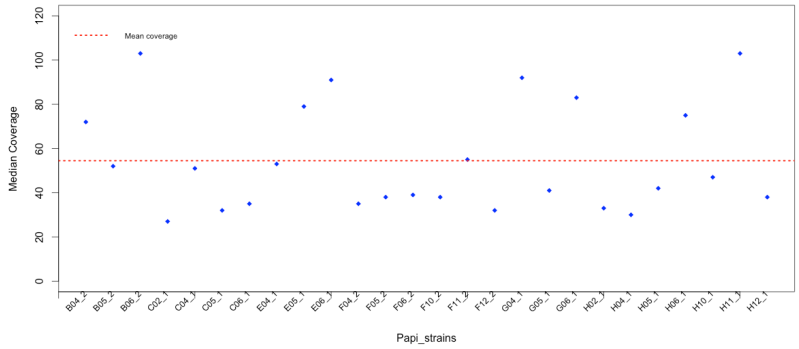
Figure 11. Two growth phenotypes of papillation strains. When grown on SC-His medium, there are two distinct growth phenotypes of papillation strains. For type I, pre-selection populations tend to be no-growth, while for type II pre-selection populations tend to be weaker growth. (Column and row labels are the same as Figure 3B).

3.2.2 Whole-genome sequencing analysis

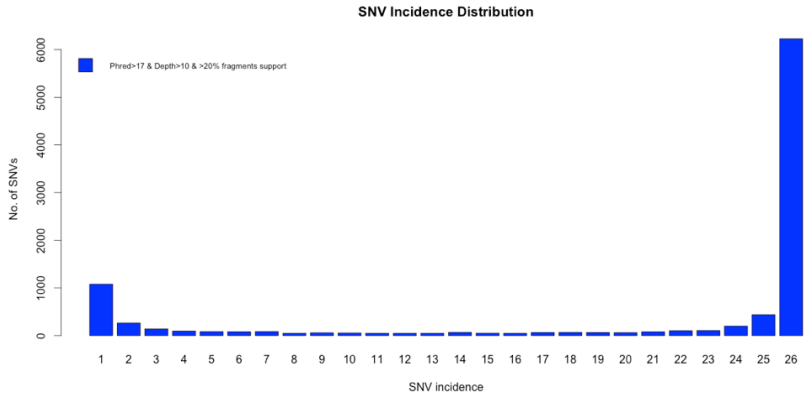
Similarly to analysis of Ctf strains, I mapped paired-end reads to reference genome S288C. The mean of median coverage of sequenced genomes was 54.4, with a range from 27 to 103 (Figure 12A). Although there were thousands of SNVs for each strain relative to the reference genome, most were shared and came from parental strains (Figure 12B). In case there are some hotspots where mutations are more likely to increase growth rate, I extracted mutations with incidence ≤ 2/26, Phred score > 17, read depth >= 10 and >= 20% fragments that support the mutation. After ANNOVAR annotation (Wang et al., 2010), there was an average of 13.5 unique non-synonymous mutations for each strain, varying from 6 to 31.

Genome tracks were generated as well (URL see Table 6).

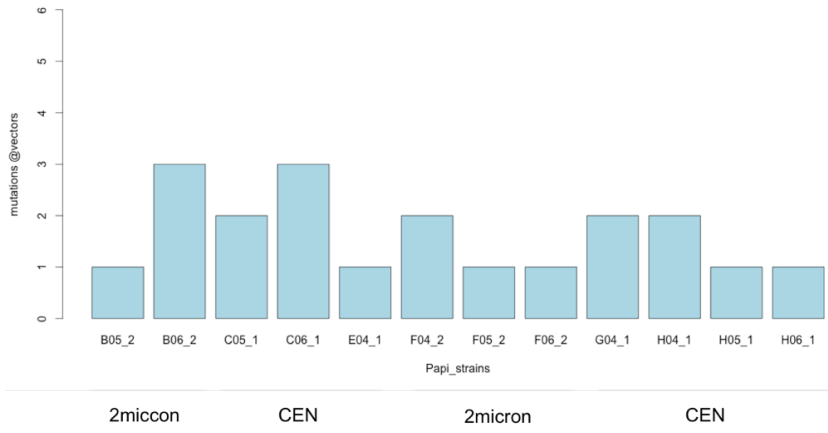
There were two sets of backbone plasmids used in these experiments. Some pairs have the low-copy DB/AD vectors (CEN); other pairs have the high-copy pVV212/pVV213 vectors (2-micron). After mapping sequencing reads to plasmid backbones, I also detected some unique mutations in vectors in 12 out of 26 strains (Figure 12C).



A



B



C

Figure 12. Whole-genome sequencing analysis of papillation strains. The mean of median coverage of sequenced genomes was 54.4, with a range from 27 to 103 (A). Although there were thousands of SNVs for each strain relative to the reference genome, most were shared and came from parental strains (B). There were unique mutations in vectors in 12 out of 26 strains (C).

Papillation strains	Chromozoom track
B04_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_B04_2.txt
B05_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_B05_2.txt
B06_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_B06_2.txt
C02_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_C02_1.txt
C04_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_C04_1.txt
C05_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_C05_1.txt
C06_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_C06_1.txt
E04_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_E04_1.txt
E05_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_E05_1.txt
E06_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_E06_1.txt
F04_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_04_2.txt
F05_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_05_2.txt
F06_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_06_2.txt
F10_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_10_2.txt
F11_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_11_2.txt
F12_2	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_12_2.txt
G04_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_G04_1.txt
G05_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_G05_1.txt
G06_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_G06_1.txt
H02_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H02_1.txt
H04_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H04_1.txt
H05_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H05_1.txt
H06_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H06_1.txt

H10_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H10_1.txt
H11_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H11_1.txt
H12_1	http://chromozoom.org/?db=sacCer3&customTracks=http://dalai.mshri.on.ca/~rhuang/ucsc_tracks/Papi3/Papi_H12_1.txt

Table 6. Web links to genome tracks for papillation strains.

3.2.3 Two classes of mutations

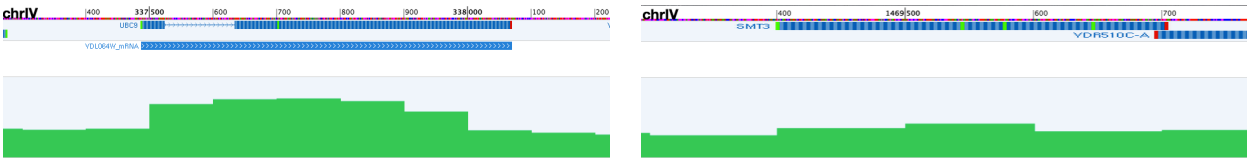
Some genetic changes could result in higher Y2H signals, such as mutations in GAL4 DB or AD domain, mutations located in the reporter gene, mutations in prey or bait protein, as well as higher expression of prey or bait protein. Also some mutations that could regulate prey/bait interaction can also alter Y2H signal. Since we are more interested in genetic modifiers of PPIs *in vivo*, we mainly focused on mutations that did not fall within genes encoding components of the Y2H system.

3.2.3.1 Copy Number Variations

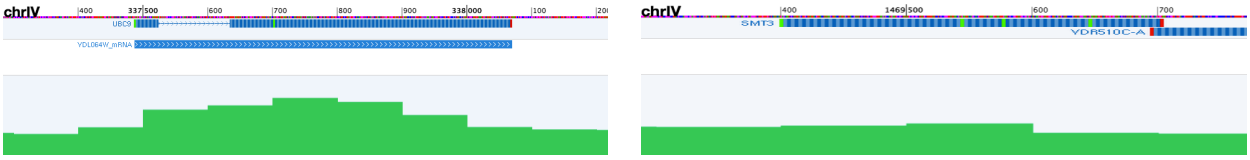
Displayed by Chromozoom, the papillation phenotype of 5 Type II strains with a CEN plasmid might be explained by a high copy number of the CEN plasmid, including E04_1, E06_1, G04_1, G05_1, and G06_1. For these 5 strains, the CNV of X_ORF is higher than its neighboring region, while the Y_ORF remains the same (Figure 13). So it is possible that high copy number of the DB_X vector increases expression level of DB-X protein, which in turn enhances DB-X and AD-Y interaction by mass action, thus increasing the expression of *HIS*3.

DB_ORF AD_ORF

E04_1



E06_1



G04_1

G05_1

G06_1

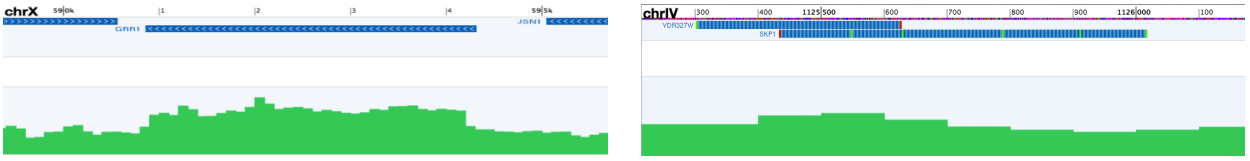


Figure 13. Copy Number Variations. The papillation phenotypes of 5 Type II strains with a CEN plasmid might be explained by a high copy number of the CEN plasmid, including E04_1, E06_1, G04_1, G05_1, and G06_1. For these 5 strains, the CNV of X_ORF is higher than its neighboring region, while the Y_ORF remains the same.

3.2.3.2 Mutations modify X and Y protein-protein interaction pattern

Because papillation strains occurred in Y2H experiments with specific pairs of X and Y proteins, the mutations we identified could represent potential genetic modifiers of PPIs *in viv*o.

3.2.3.2.1 DB-Ubc9p/AD-Smt3p

There are 6 strains carrying the X/Y protein pair DB-Ubc9p/AD-Smt3p: F04_2, F05_2, and F06_2 (2-micron, Type I), as well as E04_1, E05_1, E06_1 (CEN, Type II). As described in the Introduction, Ubc9p is an E2 SUMO-conjugating enzyme, which participates in the Smt3p conjugation pathway (Hay, 2005). In general, we found that distinct strains harbored mutations related to inositol. This is interesting given previous findings that inositol can modulate SUMO protein conjugation *in vivo* (Felberbaum et al., 2012). Even though the mechanism by which inositol modulates SUMO conjugation remains unknown, our results provided new evidence for an interdependent relationship between inositol and SUMO. Here we summarize the findings for Ubc9p/Smt3p strains.

F04_2 has a frameshift mutation in *OPI1* (c.821_822insTC:p.F274fs:het), a transcriptional repressor for *INO1*, which encodes the rate-limiting enzyme of inositol biosynthesis (Greenberg et al., 1982). If it is a gain of function mutation, then inositol level decreases, and global sumoylation increases, which requires more frequent interaction between Ubc9p and Smt3p (and possibly but not necessarily a steady-state increase in the abundance of the Ubc9p/Smt3p complex). Otherwise, if it is a loss of function mutation, then inositol level increases and inositol excretion increases, then global sumoylation decreases, suggesting lower frequency of Smt3p/Ubc9p interaction (and possible a steady-state decrease in the abundance of the Ubc9p/Smt3p complex). Either pathway could enhance interaction between Ubc9p and Smt3p. Also *opi1Δubc9Δ* has synthetic growth defect (Felberbaum et al., 2012). Additionally, *opi1Δ* has haploproficient phenotype, which means the heterozygous diploids display an increased growth rate relative to homozygous wild-type diploids (Pir et al., 2012).

F05_2 has a point missense mutation in *SNA3* (c.A272C:p.D91A:het), which is directly upstream of *INO1*. It is reported that intermediate levels of inositol represses *INO1* and induces *SNA3*, and there is a coregulation mechanism of *SNA3* and *INO1* (Shetty et al., 2013). So disrupted *SNA3* might influence *INO1* expression. Additionally, there is frameshift deletion in

APC1 (c.1693_1695del:p.565_565del:hom), encoding the largest subunit of the Anaphase-Promoting Complex/Cyclosome (Zachariae et al., 1996), when Smt3p/SUMO and Ubc9p are required for efficient proteolysis mediated by APCC (Dieckhoff et al., 2004). There is a missense mutation in *APJ1* (c.T793C:p.C265R:het) , a chaperone with a role in SUMO-mediated protein degradation (Sahi et al., 2013). Therefore, mutations in *APC1* and *APJ1* provide alternative explanations for increased Ubc9p/Smt3p interaction.

F06_2 has variants in *OXA*1 (c.T377C:p.L126S:het) and *AFG3* (c.T1221A:p.D407E: het), both are related to mitochondrial inner membrane, and overexpression of *OXA1* can rescue *afg3Δ* (Rep et al., 1996). Previous study showed that *oxa1Δ* has increased phosphatidylinositol accumulation, which might be accompanied by decreased inositol concentration. Oxa1p also has physical interaction with Smt3p (Wohlschlegel et al., 2004). F06_2 also has a non-frameshift deletion in *SRP40* (c.160_162del:p.54_54del:het), whose product has a role in pre-ribosome assembly and transport (Ikonomova et al., 1997). Large-scale bimolecular fluorescence complementation analysis shows physical interaction between Srp40p and Smt3p (Sung et al., 2013). Another interesting mutation is a nonsense mutation in *ACS1* (c.A2134T:p.K712X:het). *ACS1* and *ACS2* encode two isoforms for Acetyl-coA synthetase (Van den Berg and Steensma, 1995), and *acs1Δacs2Δ* double mutant is not viable. There is an inositol/choline-responsive element (ICRE) in the *ACS2* upstream region as an activating sequence regulated by Ino2p and Ino4p (Hiesinger et al., 1997), which are also positive regulators of *INO1*. Thus the *ACS1* nonsense mutation represents a second potential connection in this strain between SUMO and inositol.

E04_1 has a non-frameshift deletion in *SGF73* (c.432_437del:p.144_146del:het). It encodes a SAGA complex subunit, which is necessary to recruit the histone deubiquitination module into SAGA complex. The *sgf73*Δ mutation exhibits inositol auxotrophy (McMahon et al., 2005). So this mutation might decrease inositol level *in vivo*.

In addition to a point mutation in *SGF73* (c.G436A:p.D146N:het), E05_1 also has a nonsense mutation in *RPO21* (c.T4794A:p.Y1598X:het), encoding RNA polymerase II. Partial truncations of the carboxyl terminal domain (CTD) of *RPO21* may produce inositol auxotrophy by decreasing *INO*1 expression (Archambault et al., 1996). Rpo21p also has physical interaction with Smt3p (Sung et al., 2013). Two missense mutations in *DEF1* (c.G1218T:p.Q406H:het;

c.G1209T:p.Q403H:het) might change RNAPII degradation given previous phenotypes observed for this gene (Woudstra et al., 2002). Additionally, *VPS36* harbors one missense mutation (c.A1111T:p.M371L:het), which is interesting given previous studies showing that *vps36Δ* has an increased inositol level (Hancock et al., 2006). This indicates possible relationship between *VPS3*6 and intracellular inositol level.

E06_1 is the only strain for which I did not find any mutation related to inositol. This strain has a frameshift mutation in *HHY1* (c.193_196GTATGT:frameshift:hom). One possibly interesting side note is that *hhy1Δ* mutation has the haploproficient property (Pir et al., 2012).

In summary, 5 out of 6 strains with DB-Ubc9/AD-Smt3 harbored mutations that could at least plausibly alter inositol levels (Figure 14), but there is little study on the relationship between inositol and SUMO pathway. These results motivate us to conduct further studies on this relationship, such as measuring intracellular inositol level in post-selection populations.

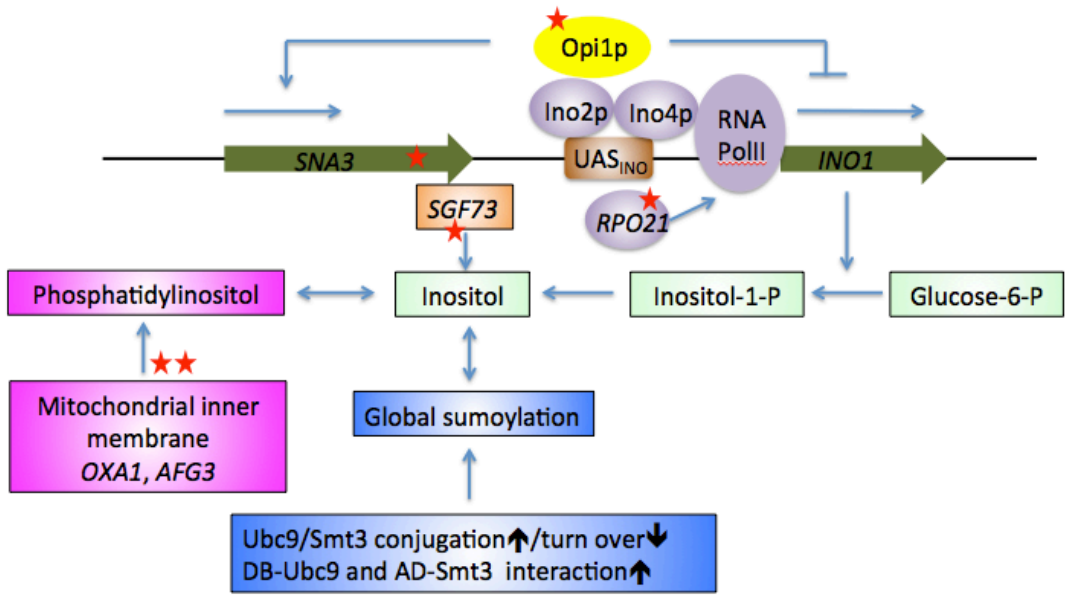


Figure 14. Relationship between SUMO and inositol. F04_2 has a frameshift mutation in *OPI1*, a transcriptional repressor for *INO1*, which encodes the rate-limiting enzyme of inositol biosynthesis. F05_2 has a point missense mutation in *SNA3*, which is upstream of *INO1*. It is reported that intermediate levels of inositol represses *INO1* and induces *SNA3*, and there is a coregulation mechanism of *SNA3* and *INO1*. F06_2 has variants in *OXA*1 and *AFG3*, both are related to mitochondrial inner membrane, and overexpression of *OXA1* could rescue *afg3*Δ. And *oxa1*Δ has increased phosphatidylinositol accumulation, which might be accompanied by decreased inositol concentration. E04_1 has a non-frameshift deletion in *SGF73*. It encodes a SAGA complex subunit, which is necessary to recruit the histone deubiquitination module into SAGA complex. Also *sgf73*Δ has inositol auxotrophy. E05_1 also has a nonsense mutation in *RPO21,* encoding RNA polymerase II. Partial truncations of the carboxyl terminal domain (CTD) of *RPO21* may produce inositol auxotrophy by decreasing *INO*1 expression.

3.2.3.2.2 DB-Lsm5p/AD-Lsm6p

Lsm5p and Lsm6p are both Lsm ("Like Sm") proteins, and are parts of two alternative seven-subunit complexes, either Lsm1p-7p or Lsm2p-8p (Mayes et al., 1999). The Lsm1p-7p complex is cytoplasmic and physically interacts with the Pat1p and Xrn1p exoribonucleases, suggesting it to have a role in decay of mRNA, while nuclear Lsm2p-8p complex associates with U6 snRNA and is involved in pre-mRNA splicing and tRNA, snoRNA, and rRNA processing (Bouveret et al., 2000). Three strains, F10_2, F11_2 and F12_2 (CEN, Type II), contain DB-*LSM*5 and AD-*LSM6* plasmids. Since Lsm5p/Lsm6p work as part of a heteroheptameric complex, I found some mutations that might influence the function of the complexes.

F10_2 carries a missense mutation in *RTT101* (c.A586G:p.T196A:het). Rtt101p is a cullin subunit of a Roc1p-dependent E3 ubiquitin ligase complex, and is involved with Mms1p in nonfunctional rRNA decay (Fujii et al., 2009). That both Rtt101p and the Lsm2p-8p complex share a function (processing rRNA) makes it more likely that the mutation in *RTT101* is causal for the apparently-increased interaction between Lsm5p and Lsm6p. Also, genetic interaction mapping showed a negative genetic interaction between *RTT101* and *LSM3* (Wilmes et al., 2008), suggesting possible parallel functions between them. The F10_2 strain also has a non-frameshift insertion in *EBP2* (c.376_377insAAG:p.G126delinsEG:het). *EBP2* has a positive genetic interaction with another component of the Lsm complexes, *LSM7* (Wilmes et al., 2008), suggesting that they are possibly involved in the same pathway. Ebp2p is indispensable for 25S rRNA maturation and 60S subunit assembly (Tsujii et al., 2000), so it has overlapping function with Lsm2p-8p.

F11_2 has a non-frameshift insertion in *ZDS1* (c.2716_2717insAACAAC: p.E906delinsEQQ:hom). Zds1p has been implicated in exporting mRNA from the nucleus to the cytoplasm (Estruch et al., 2005). We can speculate that disrupted function of Zds1p could cause mRNA buildup in the nucleus, perhaps leading to more Lsm2p-8p complex within the nucleus, with accompanying increase in measured interaction between Lsm5p and Lsm6p (as measured by the Y2H reporter within the nucleus).

F12_2 harbored mutations in *TAD3* (c.G229A:p.D77N:het), *SRP40* (c.A575G:p.E192G:het) and *SIM1* (c.378_380del:p.126_127del:het). Tad3p is a subunit of tRNA-specific adenosine-34 deaminase, which converts adenosine to inosine (Gerber and Keller,

1999). Also, *TAD3* has negative genetic interactions with both *LSM1* and *LSM7* (Wilmes et al., 2008). Srp40p plays an important role in pre-ribosome assembly or transport (Ikonomova et al., 1997), as well as being a chaperone of small nucleolar ribonucleoprotein particles (snoRNPs) (Yang et al., 2000). *SRP40* has a negative genetic interaction with *LSM7* (Wilmes et al., 2008), and Lsm5p is a multicopy suppressor of the synthetic lethality in mutants with *SRP40* null background (Yang et al., 2000). Regarding *SIM1*, we noted that the *sim1*Δ mutant has a haploproficient phenotype (Pir et al., 2012), but we noted no compelling connection to the Lsm complex.

In summary, we identified a number of mutations in papillation strains that fell within genes related to the function of Lsm complexes, making them interesting candidates for future investigation as causal variants affecting the papillation phenotype in these strains.

3.2.3.2.3 DB-Tap42p/AD-Sit4p

Tap42p is an essential protein in Tor signaling pathway, which associates with type 2A-related protein phosphatase Sit4p and the type 2A phosphatase catalytic subunit (Di Como and Arndt, 1996). Tap42p/Sit4p physical interaction is regulated by nutrient growth signals and the rapamycin-sensitive Tor signaling pathway (Di Como and Arndt, 1996). Sit4p is required for the cell to transit from late G1 in to S phase of the mitotic cycle after dissociating from inhibitor complex Tap42p/Sit4p (Sutton et al., 1991). Three strains, B04_2, B05_2 and B06_2 (2-micron, type I), carry plasmids with DB-*TAD42* and AD-*SIT4*. Because Tap42p/Sit4p complex formation is cell cycle dependent (increased interaction in the G1 phase), any mutations that alter the cell cycle distribution have the potential to increase (or decrease) the Tap42p/Sit4p interaction.

B04_2 has a non-frameshift insertion in *ZDS1* (c.2716_2717insAACAAC:p.E906delinsEQQ:hom). Besides the function of Zds1p mentioned above, it also contributes to maintaining Cdc55p in the cytoplasm to promote mitotic entry and regulating Cdc14p to induce the mitotic exit (Queralt and Uhlmann, 2008). If the mutation causes cell cycle delay, especially arrest in G1 phase, this could stabilize the Tap42p/Sit4p complex.

B05_2 has a nonsense mutation in *RPO21* (c.T4773A:p.Y1591X:het), and its product Rpo21p physically interacts with Sit4p (Jouvet et al., 2010). Cells with temperature-sensitive

mutation *rpo21* were partially impaired for cell-cycle progress at a permissive temperature, and were arrested in G1 at the cell-cycle regulatory step, START, at a restrictive temperature (Drebot et al., 1993).

Two candidate causal mutations for B06_2 were identified: *TMA23* (c.489_491del:p.163_164del:het) and *VPS36* (c.C1131A:p.N377K:het). Tma23p is nucleolar protein, possibly involved in ribosome biogenesis, and its deletion mutant could extend chronological lifespan (Burtner et al., 2011). Vps36p is component of the ESCRT-II complex (Teo et al., 2006), and its null mutant also exhibits haploproficiency (Pir et al., 2012).

3.2.3.2.4 DB-Das2p/AD-Pbn1p

We sequenced one non-papillation strain, C02_1, and three papillation strains，C04_1, C05_1 and C06_1 (CEN, Type II) with plasmids containing *DAS2* and *PBN1*. *PBN1* is an essential gene that encodes a component of glycosylphosphatidylinositol-mannosyltransferase. Pbn1p contributes to the autocatalytic post-translational processing of the protease B precursor Prb1p (Naik and Jones, 1998). *DAS2* is a non-essential gene, identified from screening mutants with increased levels of rDNA transcription (Hontz et al., 2009). However, because of unknown function of Das2 and limited study of the Das2p/Pbn1p interaction, I could not narrow down strong candidate mutations/genes for most strains.

Only C05_1 has a missense mutation in *UBI4* (c.C565T:p.P189S:het). Ubi4p has a physical interaction with Das2p (Swaney et al., 2013). However, there is not enough evidence to form a strong hypothesis that that an ubiquitin mutation could increase Das2/Pbn1 interaction.

3.2.3.2.5 DB-Grr1p/AD-Skp1p

The Skp1/cullin/F-box protein (SCF) complex is one class of E3 ubiquitin ligases, which is involved in the ubiquitin-dependent proteasomal degradation. Grr1p is the F-box protein component of SCF, mediating substrate specificity (Hsiung et al., 2001). Skp1p is a conserved kinetochore protein (Stemmann and Lechner, 1996). SCF complex regulates G1/S and G2/M transitions during cell cycle progression. Intracellular Grr1p level increases with glucose, and Grr1p is also more stable (Fey and Lanker, 2007). The Grr1p/Skp1p interaction increases under high levels of glucose (Li and Johnston, 1997). G04_1, G05_1 and G06_1 contain plasmids with

GRR1 and *SKP1* (CEN, Type II). These three strains might be explained by the high copy number we inferred for DB-*GRR1*, but I also identified some other candidate mutations.

G04_1 has a missense mutation in *SHP1* (c.G1126A:p.E376K:het). Shp1p contains a UBX (ubiquitin regulatory X) domain, and the *shp1* mutant has severe impaired in growth and delay at the metaphase to anaphase transition of cell cycle progression by reducing Glc7 activity (Bohm and Buchberger, 2013). This strain also contains a missense mutation in *TOM1* (c.C4499T:p.S1500L:het). Tom1p is a hect-domain E3 ubiquitin ligase, which could control Cdc6 degradation in G1 phase, acting separately from SCF (Cdc4) (Skp1/Cdc53/F-box protein) complex and the F-box protein Dia2 (Kim et al., 2012). The functions of Shp1p and Tom1p suggest a potential relationship between them and Grr1p/Skp1p interaction.

G05_1 harbors a nonsense mutation in *UBP12* (c.C1355A:p.S452X:het). Ubp12p is an ubiquitin-specific protease and can remove ubiquitin from ubiquitinated proteins, thereby antagonizing ubiquitin ligase activity (Amerik et al., 2000), possibly including that of the SCF complex. However, it is possible to argue an alternative view: deneddylation by the COP9 signalosome (CSN) and deubiquitylation by Ubp12p both protect the components of cullin-RING ubiquitin ligases (CRLs) (and the SCF complex is a cullin-RING ubiquitin ligase) to keep physiological CRL activities (Wu et al., 2006). So Ubp12p might have a dual role in E3 ligase activity, making it difficult to decide the impact of the mutation. However, that mutation in *UBP12* might change Grr1/Skp1 interaction makes it a candidate worth further study.

G06_1 has a mutation in YLL066W-B (c.72_73del:p.24_25del:het). Although its product has unknown function, YLL066W-B overexpression could lead to a cell cycle delay or arrest (Niu et al., 2008). Although we do not know whether the Grr1p/Skp1p interaction is cell-cycle dependent, the involvement of the SCF complex in cell cycle control makes this worth considering.

3.2.3.2.6 DB-Gyl1p/AD-Rvs167p

The Rab GTPase-activating protein (GAP) Gyl1p controls polarized exocytosis at the small-bud stage in yeast, together with Gyp5p. Both Gyl1p and Gyp5p interact with Rvs167p, recruiting Rvs167p to the small-bud tip during polarized growth (Prigent et al., 2011). Rvs167p and Rvs161p function together to regulate actin cytoskeleton, endocytosis, etc. (Lombardi and

Riezman, 2001). Three papillation strains, H04_1, H05_1 and H06_1 (CEN, Type II), and one non-papillation strain H02_1 with DB-Gyl1p/AD-Rvs167p have been sequenced.

H02_1 has a missense mutation in *CHS6* (c.A1949T:p.Q650L:het). Chs6p mediates export of some cargo proteins. Synthetic lethal screens showed that *RVS161* and *RVS167* have synthetic lethal interactions with *CHS6*, suggesting that Chs6p and Rvs161p/Rvs167p have parallel functions (Friesen et al., 2006). *AVL9* has a missense mutation (c.A1918T:p.N640Y:het), and it encodes a protein involved in exocytic transport from the Golgi (Harsay and Schekman, 2007). As both Chs6p and Avl9p share some functions with Rvs161p/Rvs167p, these mutations are candidates for decreasing the interaction between Gyl1p and Rvs167p.

H04_1 has a deletion in *PTP3* (c.2169_2171del:p.723_724del:het), which encodes phosphotyrosine-specific protein phosphatase (Zhan et al., 1997). Genome-scale genetic interaction mapping revealed a negative genetic interaction between *PTP3* and *RVS161* (Costanzo et al., 2010). H04_1 also has a mutation in *NNT1* (c.A398G:p.D133G:het), encoding S-adenosylmethionine-dependent methyltransferase (Anderson et al., 2003). Nnt1p contributes to lifespan determination, and *nnt1Δ* mutation has haploproficient phenotype (Pir et al., 2012). H04_1 also has a deletion in *GID8* (c.534_536del:p.178_179del:het), encoding a subunit of GID complex. Genome-scale genetic interaction mapping showed negative genetic interactions between *GID8* and both *RVS161* and *RVS167* (Costanzo et al., 2010). This weak evidence suggests the *GID8* mutation to be a candidate for increased Gyl1p/Rvs167p interaction. But due to the potential for false positives, the genetic interactions need further conformation.

H05_1 harbors mutations in *SEC23* (c.A602C:p.K201T:het) and *TRS130* (c.A2377G:p.R793G:het). Sec23p is also a GTPase-activating protein. Sec23p/Sec24p heterodimer, one component of COPII vesicles, is involved in ER to Golgi transport (Duden, 2003). Trs130p is part of transport protein particle (TRAPP) complex II, which regulates intra-Golgi and endosome-Golgi traffic (Sacher et al., 2000). So both Sec23p and Trs130p participate in intracellular transport, which are closely related to exocytosis and endocytosis involving Gyl1p and Rsv167p.

3.2.3.2.7 DB-Sed5p/AD-Sly1p

Sly1p is essential for vesicle trafficking between the ER and Golgi (Dascher et al., 1991). Sly1p binding to the syntaxin tSNARE Sed5p stimulates Sed5p assembly into a *trans*-SNARE membrane-protein complex from *cis*-SNARE complex (Kosodo et al., 2002). But abolishing their high-affinity interaction did not disrupt their function at all, indicating that the interaction between them is dispensable for transport (Peng and Gallwitz, 2004). Three strains, H10_1, H11_1 and H12_1 (2-micron, Type I), have plasmids with *SED5* and *SLY1*.

H11_1 carries a missense mutation in *MYO2* (c.T4642C:p.Y1548H:het), encoding an essential class V myosin heavy chain. It functions as molecular motors in actin-based transport of cargos (Pruyne et al., 2004). Since Sed5p and Sly1p are required for vesicle trafficking between the ER and Golgi, if *MYO2* mutation has an impact on intracellular transport, Sed5p/Sly1p interaction might also be influenced.

H12_1 harbors a missense mutation in *FIT1* (c.T611A:p.V204D:het). Fit1p is mannoprotein that is part of cell wall, involved in its retention of siderophore-iron (Philpott et al., 2002). *FIT1* is treated as a candidate gene since there is no better candidate gene. So it is still worthy to study the relationship between *FIT1* and increased interaction between Sed5p/Sly1p.

I did not find strong evidence to support mutations in some strains (such as C02_1, C04_1, C06_1, H06_1, H10_1, H12_1) as candidate mutations for stronger growth in His- medium. But even so, these mutations may represent new modifiers of these protein-protein interactions.

In summary, after computational analysis and literature review, I have built some possible hypothesis for most papillation strains (Table 7). Further validation of these hypotheses is now needed.

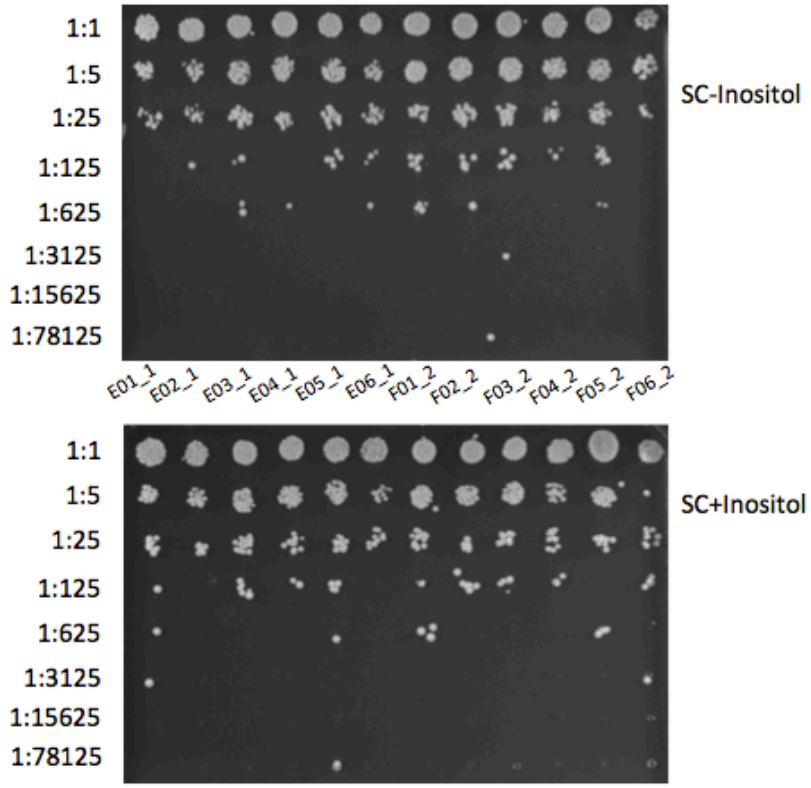
Papillation strains	F_ORF	R_ORF	CNV*	Type	Vectors	Candidate mutations
B04_2	TAP42	SIT4	high	I	pVV212 pVV213	*ZDS1*
B05_2			high	I		*RPO21*
B06_2			high	I		*TMA23, VPS36*
C02_1	DAS2	PBN1	normal	Control	CEN_DB/AD	
C04_1			normal	II		
C05_1			normal	II		*UBI4, POL32, RAD1*
C06_1			normal	II		
E04_1	UBC9	SMT3	X_high	II	CEN_DB/AD	X high copy number, *SGF73*
E05_1			normal	II		*SGF73, RPO21, DEF1, VPS36*
E06_1			X_high	II		X high copy number, *HHY1*
F04_2			high	I	pVV212 pVV213	*OPI1*
F05_2			high	I		*APC1, APJ1, SNA3-INO1*
F06_2			high	I		*AFG3, OXA1, ACS1, SRP40*
F10_2	LSM5	LSM6	normal	II	CEN_DB/AD	*RTT101, EBP2*
F11_2			normal	II		*ZDS1*
F12_2			normal	II		*TAD3, SRP40, SIM1*
G04_1	GRR1	SKP1	X_high	II	CEN_DB/AD	X high copy number, *SHP1, TOM1*
G05_1			X_high	II		X high copy number, *UBP12*
G06_1			X_high	II		X high copy number
H02_1	GYL1	RVS167	normal	Control	CEN_DB/AD	*CHS6, AVL9*
H04_1			normal	I		*PTP3, NNT1,GID8*
H05_1			normal	I		*SEC23, TRS130*
H06_1			normal	I		
H10_1	SED5	SLY1	X_high	I	pVV212 pVV213	
H11_1			X_high	I		*MYO2*
H12_1			X_high	I		*FIT1*

Table 7. Summary of papillation strains.

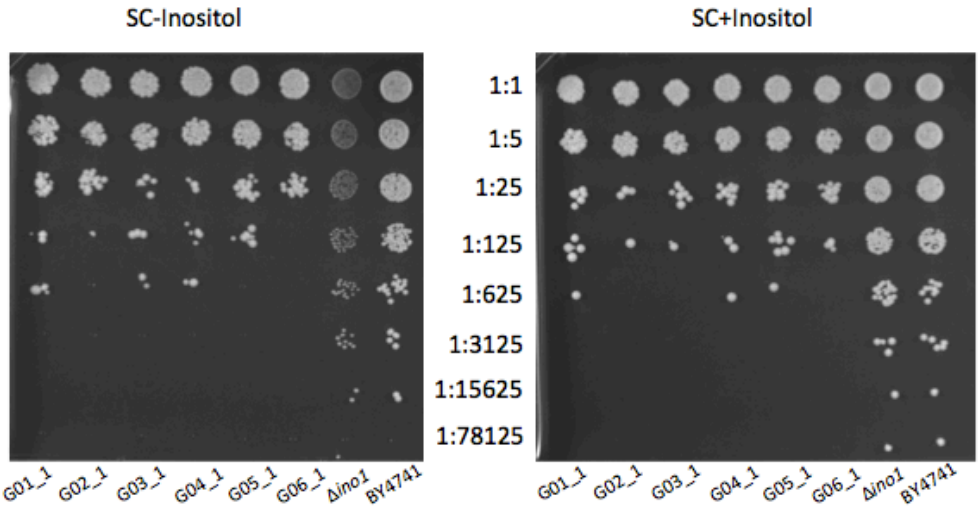
*CNV was estimated based on coverage for each strain. "Normal" means the coverage of X and Y are the same as neighboring regions, "high" means the coverage of X and Y are higher than neighboring regions, and "X_high" means the coverage of X is higher than neighboring regions while the coverage of Y remains the same.

3.2.4 Inositol auxotrophy in populations with Ubc9p/Smt3p

To test our hypothesis on the relationship between inositol and SUMO, I checked if strains exhibiting increased Ubc9p/Smt3p interaction have an inositol auxotrophy phenotype. However, there was no significant difference on growth rates when grew on +inositol or -inositol medium, either papillation strains or non-papillation strains (Figure 15). More confirmation experiments, such as measuring intracellular inositol level, will be conducted to validate this hypothesis. For example, intracellular inositol mass per 100 μg of protein can be determined by the enzyme-coupled fluorescence assay of Maslanski and Busa (Vaden et al., 2001).



A



B

Figure 15. Spotting assay for testing inositol auxotrophy.

4 Discussions

4.1 Genome sequencing is a powerful tool for locating genetic causes of mutants

The computational pipeline for processing NGS data was initially assembled by T. Yamaguchi (unpublished). I adjusted some steps according to my strains. For Ctf strains, mutations within some known Ctf genes have been identified, providing convincing evidence that they are causal mutations. SNVs and small indels that passed the stringent threshold were all confirmed by Sanger sequencing, indicating high confidence of the stringent threshold. Since I did not find any candidate SNVs for s48 and s49, I also tested some SNVs that passed a permissive threshold only (Phred score>10, read depth>=3, and >=50 of high-quality aligned reads that support the mutations), but it turned out that most are false positives. Also, three candidate mutations have been validated by experiments. Therefore, my study reconfirmed that this computational pipeline (at the stringent threshold) works well for locating genetic causes for the relevant mutant phenotype.

However, it also has some limitations. I randomly tested 8 structural variations detected by Pindel, but only one was validated by Sanger sequencing. Structural variation is more complicated than single point mutations, which may explain the lower precision. In addition, this pipeline only applies to haploid genomes, since the threshold with >=90 of high-quality aligned reads supporting the mutations filtered out all heterozygous mutations of diploid genomes. So for papillation strains, I chose a permissive threshold to filter SNVs and used another existing tool to distinguish homozygous and heterozygous mutations. However, as a result of this, the accuracy decreased. It would be better if we could add some advanced options for customizing the pipeline, which will make it work in a wider area.

4.2 Validation of Ctf genes

To validate causal roles of candidate Ctf genes, I applied tetrad dissection to check which specific mutation always tracked with Ctf phenotype after meiosis. Unexpectedly, after separating spores, most segregant colonies turned pure red, which made it hard to score Ctf phenotype and obtain 2:2 segregation. One possible explanation is that a mutation that causes mitotic failure could also result in failure in meiosis (Marston et al., 2004), However we did not

test this possibility. One solution for this is to re-plate all colonies, like s41. But if the whole colony has already lost the chromosome fragment (CF), there is no way to tell whether it has Ctf phenotype or not. Another option is to perform tetrad dissection on SC-Ura medium, and only white colonies containing CF could grow. Then all colonies can be re-plated on YPD medium to score Ctf phenotype. But for this protocol, it is difficult to get all four spores to survive and grow into colonies. Thus we did not find a perfect solution for this problem.

As an alternative experiment to validate candidate genes, plasmid transformation rescue was attempted by Philip Hieter's lab to test all non-synonymous SNVs (personal communication). There are two positive hits by this method. But due to limited time, they did not confirm the expression of the wild-type allele. So the rest of the candidate genes are still possible genetic causes for Ctf mutants.

There has been 25 years of progress in identifying Ctf genes. The Hieter Lab's Ctf screen as the classical genetic screen for chromosome instability (CIN), has expanded our understanding of mechanisms of CIN and its mutability in yeast. In the original screen, >65% of 136 randomly selected Ctf mutations were mapped to 13 genes, 12 of which were involved in sister chromatid cohesion and/or kinetochore function (Barber et al., 2008; Spencer et al., 1990; Stirling et al., 2011). The enrichment of random Ctf alleles indicates that these genes are more easily mutable to cause Ctf phenotype. Random mutations are presumed to initiate CIN phenotype of tumors, and Ctf genes are the most sensitive to random mutagenesis. Therefore, although systemic reverse screens have expanded CIN genes dramatically, it's still worthy to identify Ctf genes from random approaches. In addition, screening the systematic gene deletion collection could not find gain-of-function or neomorphic alleles but only collections of identified Ctf genes that lead to CIN by loss of function.

It is possible that human orthologs of Ctf genes are new CIN genes with relevance to tumorigenesis. There are already several successful uses of yeast CIN genes to predict human candidate CIN genes in cancer (Barber et al., 2008; Spencer et al., 1990; Stirling et al., 2011). So exploring yeast CIN genes is helpful to understand how CIN contribute to human disease, especially cancer.

4.3 Papillation strains identified genetic modifiers of specific protein-protein interactions.

Traditionally, the phenomenon of papillation has been described in an environment where wild-type cells stop growth due to the lack of nutrients, and mutants that can utilize other nutrients overcome this barrier to continue growing (Yang et al., 2011).

In our study, we used papillation phenotype to refer to increased-growth colonies isolated from selective medium in Y2H assay. Under the hypothesis that spontaneous mutations leads to at least two kinds of genotypes that contributed to this gross difference, we performed whole-genome sequencing and analysis to identify candidate genetic modifiers that could regulate PPIs. Papillation happens X/Y dependently (it is not observed for many interacting X/Y pairs), suggesting that some specific interactions are more easily modulated by spontaneous mutations. From preliminary analysis, we identified two types of candidate causal genetic variants.

Many interactions *in vivo* are transient, so that higher concentrations of the interacting proteins might enhance interaction. This is why we take high copy number of CEN_DB plasmids as an explanation for papillation strains. We saw that increased interactions between Grr1p and Skp1p or between Ubc9p and Smt3p may be explained by high concentration of one of the interactors. To explore this further, we should first test whether high copy number leads to high expression level of Grr1p or Ubc9p.

Another interesting class of mutations suggests potential genetic modifiers of these PPIs. Based on identifying non-synonymous mutations and literature review, I attempted to identify potential functional relationships between specific mutations and specific interactions. The strongest hypothesis to emerge is that inositol level could regulate Ubc9p/Smt3p interaction, since 5 out 6 papillation strains containing Ubc9p and Smt3p harbored mutations related possibly to inositol level. It is also because Ubc9p/Smt3p interaction is the best studied. Due to some unknown functions of some interacting proteins, such as Das2p, it is difficult to pick candidate causal mutations. However, it presents opportunities to elucidate their functions by identifying modifiers of the corresponding PPIs. I did not see significant inositol auxotrophy for post-selection Ubc9p/Smt3p Y2H strains. However, tests of inositol dependence may be good at identifying lowered intracellular levels of inositol but not heightened levels. Thus, we still could

not rule out the possibility that inositol level regulates Ubc9p/Smt3p interaction, but we need more confirmation experiments.

Therefore, starting from identifying genetic causes of papillation strains, the study leads us to uncover new regulators of interaction between some specific pairs of proteins. So papillation assay could be developed into a new tool for finding new modifiers of PPIs in Y2H system.

5 Conclusions and Future Directions

My project was aimed at identifying genetic causes for two distinct *Saccharomyces cerevisiae* mutants, Ctf mutants and papillation mutants. Identification was carried out using whole-genome sequencing, computational analysis, confirmation sequencing, and some experimental validation.

Ctf mutants, with increased rate of chromosome loss, have red sectoring in medium with limited adenine. The causal genes and mutations for some Ctf strains have remained unidentified for about 25 years after they were first generated, and whole-genome sequencing analysis now makes it possible to more easily locate these mutations. Based on computational analysis of whole-genome sequencing, I identified some candidate mutations for Ctf strains, including known Ctf genes from systematic screening and some possibly novel Ctf genes. We also attempted validation by tetrad analysis and plasmid transformation rescue that three known Ctf genes were responsible for three mutants, respectively. In the future, we can continue to validate the remaining candidate genes, and it is highly possible to find novel Ctf genes. If there are new Ctf genes, we could also find out if the Ctf gene is conservative from yeast to human, then explore if it could contribute to revealing some mechanisms of human diseases, such as cancer.

Papillation assay offers a tool for finding genetic modifiers of protein-protein interactions (PPIs). In our study, papillation is due to genetic changes that allow the cells to grow stronger in His- medium in yeast two-hybrid screen. As a reporter gene in our system, this stronger growth indicates a high expression level of *HIS3*, which is likely due to increased interaction between prey and bait proteins. These PPIs may be regulated by spontaneous mutations. In summary, based on locating unique non-synonymous variants by whole-genome sequencing analysis and literature review, I generated hypotheses for many of the papillation strains. For instance, the best current model for Ubc9p/Smt3p papillation strains is that intracellular inositol level regulates their interaction, thereby altering SUMO conjugation pathway *in vivo*. For this case, in the future, we will measure intracellular inositol mass to provide direct evidence that mutations we found could increase or decrease inositol level. Intracellular inositol mass per 100 μg of protein can be determined by the enzyme-couples fluorescent assay of Maslanski and Busa (Vaden et al., 2001). We could also perform Y2H in yeast with these mutations to build causal relationships between specific mutations and the corresponding increased interactions.

In conclusion, our pipeline for processing whole-genome sequencing data works well and effectively, and candidate causal genes for Ctf strains and papillation strains were identified. We need more experiments to validate our hypothesis. In addition, we will apply papillation assay to a wider field to search for genetic modifiers for more pairs of PPIs.

6 References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nature methods *7*, 248-249.

Amerik, A.Y., Li, S.J., and Hochstrasser, M. (2000). Analysis of the deubiquitinating enzymes of the yeast Saccharomyces cerevisiae. Biological chemistry *381*, 981-992.

Anderson, R.M., Bitterman, K.J., Wood, J.G., Medvedik, O., and Sinclair, D.A. (2003). Nicotinamide and PNC1 govern lifespan extension by calorie restriction in Saccharomyces cerevisiae. Nature *423*, 181-185.

Archambault, J., Jansma, D.B., and Friesen, J.D. (1996). Underproduction of the largest subunit of RNA polymerase II causes temperature sensitivity, slow growth, and inositol auxotrophy in Saccharomyces cerevisiae. Genetics *142*, 737-747.

Barber, T.D., McManus, K., Yuen, K.W., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S.*, et al.* (2008). Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America *105*, 3443-3448.

Bohm, S., and Buchberger, A. (2013). The budding yeast Cdc48(Shp1) complex promotes cell cycle progression by positive regulation of protein phosphatase 1 (Glc7). PloS one *8*, e56486.

Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Seraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. The EMBO journal *19*, 1661-1671.

Burtner, C.R., Murakami, C.J., Olsen, B., Kennedy, B.K., and Kaeberlein, M. (2011). A genomic analysis of chronological longevity factors in budding yeast. Cell cycle *10*, 1385-1396.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PloS one *7*, e46688.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S.*, et al.* (2010). The genetic landscape of a cell. Science *327*, 425-431.

Dascher, C., Ossig, R., Gallwitz, D., and Schmitt, H.D. (1991). Identification and structure of four yeast genes (SLY) that are able to suppress the functional loss of YPT1, a member of the RAS superfamily. Molecular and cellular biology *11*, 872-885.

Di Como, C.J., and Arndt, K.T. (1996). Nutrients, via the Tor proteins, stimulate the association of Tap42 with type 2A phosphatases. Genes & development *10*, 1904-1916.

Dieckhoff, P., Bolte, M., Sancak, Y., Braus, G.H., and Irniger, S. (2004). Smt3/SUMO and Ubc9 are required for efficient APC/C-mediated proteolysis in budding yeast. Molecular microbiology *51*, 1375-1387.

Drebot, M.A., Johnston, G.C., Friesen, J.D., and Singer, R.A. (1993). An impaired RNA polymerase II activity in Saccharomyces cerevisiae causes cell-cycle inhibition at START. Molecular & general genetics : MGG *241*, 327-334.

Duden, R. (2003). ER-to-Golgi transport: COP I and COP II function (Review). Molecular membrane biology *20*, 197-207.

Estruch, F., Hodge, C.A., Rodriguez-Navarro, S., and Cole, C.N. (2005). Physical and genetic interactions link the yeast protein Zds1p with mRNA nuclear export. The Journal of biological chemistry *280*, 9691-9697.

Felberbaum, R., Wilson, N.R., Cheng, D., Peng, J., and Hochstrasser, M. (2012). Desumoylation of the endoplasmic reticulum membrane VAP family protein Scs2 by Ulp1 and SUMO regulation of the inositol synthesis pathway. Molecular and cellular biology *32*, 64-75.

Fey, J.P., and Lanker, S. (2007). Delayed accumulation of the yeast G1 cyclins Cln1 and Cln2 and the F-box protein Grr1 in response to glucose. Yeast *24*, 419-429.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Friesen, H., Humphries, C., Ho, Y., Schub, O., Colwill, K., and Andrews, B. (2006). Characterization of the yeast amphiphysins Rvs161p and Rvs167p reveals roles for the Rvs heterodimer in vivo. Molecular biology of the cell *17*, 1306-1321.

Fujii, K., Kitabatake, M., Sakata, T., Miyata, A., and Ohno, M. (2009). A role for ubiquitin in the clearance of nonfunctional rRNAs. Genes & development *23*, 963-974.

Geigl, J.B., Obenauf, A.C., Schwarzbraun, T., and Speicher, M.R. (2008). Defining 'chromosomal instability'. Trends in genetics : TIG *24*, 64-69.

Gerber, A.P., and Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science *286*, 1146-1149.

Greenberg, M.L., Reiner, B., and Henry, S.A. (1982). Regulatory mutations of inositol biosynthesis in yeast: isolation of inositol-excreting mutants. Genetics *100*, 19-33.

Hancock, L.C., Behta, R.P., and Lopes, J.M. (2006). Genomic analysis of the Opi- phenotype. Genetics *173*, 621-634.

Harsay, E., and Schekman, R. (2007). Avl9p, a member of a novel protein superfamily, functions in the late secretory pathway. Molecular biology of the cell *18*, 1203-1219.

Hay, R.T. (2005). SUMO: a history of modification. Molecular cell *18*, 1-12.

Heilig, C.E., Loffler, H., Mahlknecht, U., Janssen, J.W., Ho, A.D., Jauch, A., and Kramer, A. (2010). Chromosomal instability correlates with poor outcome in patients with myelodysplastic syndromes irrespectively of the cytogenetic risk group. Journal of cellular and molecular medicine *14*, 895-902.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America *89*, 10915-10919.

Hiesinger, M., Wagner, C., and Schuller, H.J. (1997). The acetyl-CoA synthetase gene ACS2 of the yeast Saccharomyces cerevisiae is coregulated with structural genes of fatty acid biosynthesis by the transcriptional activators Ino2p and Ino4p. FEBS letters *415*, 16-20.

Ho, C.H., Magtanong, L., Barker, S.L., Gresham, D., Nishimura, S., Natarajan, P., Koh, J.L., Porter, J., Gray, C.A., Andersen, R.J.*, et al.* (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nature biotechnology *27*, 369-377.

Hontz, R.D., Niederer, R.O., Johnson, J.M., and Smith, J.S. (2009). Genetic identification of factors that modulate ribosomal DNA transcription in Saccharomyces cerevisiae. Genetics *182*, 105-119.

Hsiung, Y.G., Chang, H.C., Pellequer, J.L., La Valle, R., Lanker, S., and Wittenberg, C. (2001). F-box protein Grr1 interacts with phosphorylated targets via the cationic surface of its leucine-rich repeat. Molecular and cellular biology *21*, 2506-2520.

Ikonomova, R., Sommer, T., and Kepes, F. (1997). The Srp40 protein plays a dose-sensitive role in preribosome assembly or transport and depends on its carboxy-terminal domain for proper localization to the yeast nucleoskeleton. DNA and cell biology *16*, 1161-1173.

Jouvet, N., Poschmann, J., Douville, J., Bulet, L., and Ramotar, D. (2010). Rrd1 isomerizes RNA polymerase II in response to rapamycin. BMC molecular biology *11*, 92.

Kim, D.H., Zhang, W., and Koepp, D.M. (2012). The Hect domain E3 ligase Tom1 and the F-box protein Dia2 control Cdc6 degradation in G1 phase. The Journal of biological chemistry *287*, 44212-44220.

Kosodo, Y., Noda, Y., Adachi, H., and Yoda, K. (2002). Binding of Sly1 to Sed5 enhances formation of the yeast early Golgi SNARE complex. Journal of cell science *115*, 3683-3691.

Lejeune, J., Gautier, M., and Turpin, R. (1959). [Study of somatic chromosomes from 9 mongoloid children]. Comptes rendus hebdomadaires des seances de l'Academie des sciences *248*, 1721-1722.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. Nature *386*, 623-627.

Li, F.N., and Johnston, M. (1997). Grr1 of Saccharomyces cerevisiae is connected to the ubiquitin proteolysis machinery through Skp1: coupling glucose sensing to gene expression and the cell cycle. The EMBO journal *16*, 5629-5638.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lombardi, R., and Riezman, H. (2001). Rvs161p and Rvs167p, the two yeast amphiphysin homologs, function together in vivo. The Journal of biological chemistry *276*, 6016-6022.

Marston, A.L., Tham, W.H., Shah, H., and Amon, A. (2004). A genome-wide screen identifies genes required for centromeric cohesion. Science *303*, 1367-1370.

Mayes, A.E., Verdone, L., Legrain, P., and Beggs, J.D. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. The EMBO journal *18*, 4321-4331.

McMahon, S.J., Pray-Grant, M.G., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Polyglutamine-expanded spinocerebellar ataxia-7 protein disrupts normal SAGA and SLIK histone acetyltransferase activity. Proceedings of the National Academy of Sciences of the United States of America *102*, 8478-8482.

Measday, V., Baetz, K., Guzzo, J., Yuen, K., Kwok, T., Sheikh, B., Ding, H., Ueta, R., Hoac, T., Cheng, B.*, et al.* (2005). Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. Proceedings of the National Academy of Sciences of the United States of America *102*, 13956-13961.

Naik, R.R., and Jones, E.W. (1998). The PBN1 gene of Saccharomyces cerevisiae: an essential gene that is required for the post-translational processing of the protease B precursor. Genetics *149*, 1277-1292.

Niu, W., Li, Z., Zhan, W., Iyer, V.R., and Marcotte, E.M. (2008). Mechanisms of cell cycle control revealed by a systematic and quantitative overexpression screen in S. cerevisiae. PLoS genetics *4*, e1000120.

Pak, T.R., and Roth, F.P. (2013). ChromoZoom: a flexible, fluid, web-based genome browser. Bioinformatics *29*, 384-386.

Peng, R., and Gallwitz, D. (2004). Multiple SNARE interactions of an SM protein: Sed5p/Sly1p binding is dispensable for transport. The EMBO journal *23*, 3939-3949.

Philpott, C.C., Protchenko, O., Kim, Y.W., Boretsky, Y., and Shakoury-Elizeh, M. (2002). The response to iron deprivation in Saccharomyces cerevisiae: expression of siderophore-based systems of iron uptake. Biochemical Society transactions *30*, 698-702.

Pir, P., Gutteridge, A., Wu, J., Rash, B., Kell, D.B., Zhang, N., and Oliver, S.G. (2012). The genetic control of growth rate: a systems biology study in yeast. BMC systems biology *6*, 4.

Prigent, M., Boy-Marcotte, E., Chesneau, L., Gibson, K., Dupre-Crochet, S., Tisserand, H., Verbavatz, J.M., and Cuif, M.H. (2011). The RabGAP proteins Gyp5p and Gyl1p recruit the BAR domain protein Rvs167p for polarized exocytosis. Traffic *12*, 1084-1097.

Pruyne, D., Legesse-Miller, A., Gao, L., Dong, Y., and Bretscher, A. (2004). Mechanisms of polarized growth and organelle segregation in yeast. Annual review of cell and developmental biology *20*, 559-591.

Queralt, E., and Uhlmann, F. (2008). Separase cooperates with Zds1 and Zds2 to activate Cdc14 phosphatase in early anaphase. The Journal of cell biology *182*, 873-883.

Rep, M., Nooy, J., Guelin, E., and Grivell, L.A. (1996). Three genes for mitochondrial proteins suppress null-mutations in both Afg3 and Rca1 when over-expressed. Current genetics *30*, 206-211.

Sacher, M., Barrowman, J., Schieltz, D., Yates, J.R., 3rd, and Ferro-Novick, S. (2000). Identification and characterization of five new subunits of TRAPP. European journal of cell biology *79*, 71-80.

Sahi, C., Kominek, J., Ziegelhoffer, T., Yu, H.Y., Baranowski, M., Marszalek, J., and Craig, E.A. (2013). Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol. Molecular biology and evolution *30*, 985-998.

Shetty, A., Swaminathan, A., and Lopes, J.M. (2013). Transcription regulation of a yeast gene from a downstream location. Journal of molecular biology *425*, 457-465.

Spencer, F., Gerring, S.L., Connelly, C., and Hieter, P. (1990). Mitotic chromosome transmission fidelity mutants in Saccharomyces cerevisiae. Genetics *124*, 237-249.

Stemmann, O., and Lechner, J. (1996). The Saccharomyces cerevisiae kinetochore contains a cyclin-CDK complexing homologue, as identified by in vitro reconstitution. The EMBO journal *15*, 3611-3620.

Stirling, P.C., Bloom, M.S., Solanki-Patil, T., Smith, S., Sipahimalani, P., Li, Z., Kofoed, M., Ben-Aroya, S., Myung, K., and Hieter, P. (2011). The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. PLoS genetics *7*, e1002057.

Stirling, P.C., Crisp, M.J., Basrai, M.A., Tucker, C.M., Dunham, M.J., Spencer, F.A., and Hieter, P. (2012). Mutability and mutational spectrum of chromosome transmission fidelity genes. Chromosoma *121*, 263-275.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719-724.

Sung, M.K., Lim, G., Yi, D.G., Chang, Y.J., Yang, E.B., Lee, K., and Huh, W.K. (2013). Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. Genome research *23*, 736-746.

Sutton, A., Immanuel, D., and Arndt, K.T. (1991). The SIT4 protein phosphatase functions in late G1 for progression into S phase. Molecular and cellular biology *11*, 2133-2148.

Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., and Villen, J. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nature methods *10*, 676-682.

Teo, H., Gill, D.J., Sun, J., Perisic, O., Veprintsev, D.B., Vallis, Y., Emr, S.D., and Williams, R.L. (2006). ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. Cell *125*, 99-111.

Thompson, S.L., Bakhoum, S.F., and Compton, D.A. (2010). Mechanisms of chromosomal instability. Current biology : CB *20*, R285-295.

Tong, A.H., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods in molecular biology *313*, 171-192.

Tsujii, R., Miyoshi, K., Tsuno, A., Matsui, Y., Toh-e, A., Miyakawa, T., and Mizuta, K. (2000). Ebp2p, yeast homologue of a human protein that interacts with Epstein-Barr virus nuclear antigen 1, is required for pre-rRNA processing and ribosomal subunit assembly. Genes to cells : devoted to molecular & cellular mechanisms *5*, 543-553.

Vaden, D.L., Ding, D., Peterson, B., and Greenberg, M.L. (2001). Lithium and valproate decrease inositol mass and increase expression of the yeast INO1 and INO2 genes for inositol biosynthesis. The Journal of biological chemistry *276*, 15466-15471.

Van den Berg, M.A., and Steensma, H.Y. (1995). ACS2, a Saccharomyces cerevisiae gene encoding acetyl-coenzyme A synthetase, essential for growth on glucose. European journal of biochemistry / FEBS *231*, 704-713.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research *38*, e164.

Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S.*, et al.* (2008). A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. Molecular cell *32*, 735-746.

Wohlschlegel, J.A., Johnson, E.S., Reed, S.I., and Yates, J.R., 3rd (2004). Global analysis of protein sumoylation in Saccharomyces cerevisiae. The Journal of biological chemistry *279*, 45662-45668.

Woudstra, E.C., Gilbert, C., Fellows, J., Jansen, L., Brouwer, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2002). A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. Nature *415*, 929-933.

Wu, J.T., Chan, Y.R., and Chien, C.T. (2006). Protection of cullin-RING E3 ligases by CSN-UBP12. Trends in cell biology *16*, 362-369.

Yang, H., Sikavi, C., Tran, K., McGillivray, S.M., Nizet, V., Yung, M., Chang, A., and Miller, J.H. (2011). Papillation in Bacillus anthracis colonies: a tool for finding new mutators. Molecular microbiology *79*, 1276-1293.

Yang, Y., Isaac, C., Wang, C., Dragon, F., Pogacic, V., and Meier, U.T. (2000). Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. Molecular biology of the cell *11*, 567-577.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865-2871.

Yuen, K.W., Warren, C.D., Chen, O., Kwok, T., Hieter, P., and Spencer, F.A. (2007). Systematic genome instability screens in yeast and their potential relevance to cancer. Proceedings of the National Academy of Sciences of the United States of America *104*, 3925-3930.

Zachariae, W., Shin, T.H., Galova, M., Obermaier, B., and Nasmyth, K. (1996). Identification of subunits of the anaphase-promoting complex of Saccharomyces cerevisiae. Science *274*, 1201-1204.

Zhan, X.L., Deschenes, R.J., and Guan, K.L. (1997). Differential regulation of FUS3 MAP kinase by tyrosine-specific phosphatases PTP2/PTP3 and dual-specificity phosphatase MSG5 in Saccharomyces cerevisiae. Genes & development *11*, 1690-1702.

7 Appendices

7.1 Primers for Sanger sequencing of Ctf mutants

Based on analyzing gene functions and deleterious impact of the mutations, I picked 29 mutations to assess by Sanger sequencing. Among them, 17 mutations were validated. Eight mutations were proved to be false positive (7 false positive structural variations came from Pindel detection and one SNV was from the permissive threshold in the pipeline), 2 mutations failed to be amplified because of repeat sequence regions, and 2 mutations were ruled out because they were the same as each other.

Primer ID	Sequence	Length	Region	Product length	Strain	Status
BIM1_F	CAAATCAGGGACGAAGCAAT	20	chrV:188259+188720	462bp	s11	Validated
BIM1_R	TGATGACCTCTTTGCAGTGG	20				
CHL4_F	GGGAGATGGGAAACCTTACG	20	chrIV:965451+965952	502bp	S20	Validated
CHL4_R	CCGTTCAAAGTTAGCACTAGCA	22				
INN1_F	TATGAGGCCGATACCACCAT	20	chrXIV:346783+347308	526bp	S20	Validated
INN1_R	AGGTAAGGGCGGTTGGTTAT	20				
CTF4_F	TTGAGAAGGGCAAGAAGTGA	20	chrXVI:799190+799619	430bp	S41	Validated
CTF4_R	TTGTGTGTTTCATCGTCCAA	20				
CDC14_F	TTTCAGCAGCACCAACAAAG	20	chrVI:209258+209847	590bp	s56	Validated
CDC14_R	AAATGCCAATAAAGCCGTTG	20				
SPC105_F	CAACTGGATGATTTGAGAGTGG	22	chrVII:337130+337565	436bp	s62	Validated
SPC105_R	CAGAGAGGACTTCCCGTGAC	20				
SIC1_F	AGGGAAGAGCAGGAACCATT	20	chrXII:287126+287603	478bp	s121	Validated
SIC1_R	TTCGTTTCTCTACCACCTCTCC	22				
MCM5_F	GATGGGTGGTTCCAAGAAGA	20	chrXII:692730+693180	451bp	s164	Validated
MCM5_R	GAAATCAATGTTGTCACCAGGA	22				
RAD6_F	ATAAGCCACCGCATGTCAA	19	chrVII:394179+394636	458bp	s152	FP
RAD6_R	TGGAGGAATAGAAAGCAAACG	21				
RCK2_F	TCCACGAGTACGTTGGCATA	20	chrXII:633797+634364	568bp	s48	FP
RCK2_R	TTGTTATTTGTACCGCTTGACC	22				
CRZ1_F	GAACCTTTCCTCTGGCATCC	20	chrXIV:579801+580232	432bp	s48	FP
CRZ1_R	TGAGACATCGCAGGTGAAGA	20				
PBA1_F	CCTTACTGGGACCGACAATC	20	chrXII:548053+548569	517bp	s49	FP
PBA1_R	CCTCAGGACTTGGATGTTCC	20				
YLR255C_F	AGCGAGCCTTTAAGCACAGA	20	chrXII:645602+646009	408bp	s55	FP
YLR255C_R	GCCGTTGCCTGGTATTCTAA	20				
MEC3_F	CTCAGGGTGGAGCCAGTTTA	20	chrXII:712987+713410	424bp	s55	Failed
MEC3_R	GCAGAGTTGCAGAGTTGCAG	20				
YOL014W_F	TTATCACCAAATGCGACCAC	20	chrXV:299683+300173	491bp	s55	Common

YOL014W_R	TGGACTGATTATGCCCTTGA	20				
RAD5_F	CCCTTGAAGTACGGCTCTCA	20	chrXII:205564+205993	430bp	s64	Validated
RAD5_R	AGCGGCTTTCTAGCTCTTCA	20				
ADE8_F	AATATCGGCACAGATTGCAG	20	chrIV:1288495+1288944	450bp	s139	Validated
ADE8_R	AACGCTCCTAGCCTCAAGGT	20				
TEX1_F	GGAGCAATGGTGGAAGTCAT	20	chrXIV:170667+171218	552bp	s139	Validated
TEX1_R	TATTGTGGTTGTTGCCTCCA	20				
BMH1_F	TCCGTGCTAGACTCCCACTT	20	chrV:545920+546379	460bp	s140	FP
BMH1_R	GCTGCTGTTGCTGATGTTGT	20				
YML090W_F	ACGACCGGATGAGGATGTAG	20	chrXIII:90562+91150	589bp	s140	Validated
YML090W_R	TCCCTCGAATCATACACACG	20				
SGF73_F	TCAATCCGAATGCACAACTC	20	chrVII:377781+378183	403bp	s49	FP
SGF73_R	TTTGGTGTGTTCGCAGATTC	20				
TRP5_F	TTGCGGACTTCAAGGACTCT	20	chrVII:447677+448176	500bp	s55	Validated
TRP5_R	CAAGGCTGGTGCTAATGGTT	20				
SAN1_F	GCTGCACGTTGAACACCTAC	20	chrIV:742997+743542	546bp	s55	FP
SAN1_R	TCCACAAGACCGTTTGAACA	20				
FIS1_F	GTGCGTGCGATTCATTCTTA	20	chrIX:241253+241673	421bp	s64	Validated
FIS1_R	CCCACCGCTACCATACAGTC	20				
UBR2_F	CACCATCACCATCATCATCA	20	chrXII:188358+188824	467bp	s152	Validated
UBR2_R	CAGCAGCGAAATCTCTCGTA	20				
CCW12_F	AGCAGCAGAGGTGGTGTTCT	20	chrXII:369793+370248	456bp	s152	Validated
CCW12_R	CCTTTGCCTCCGTTCAAGTA	20				
MCM7_F	TGTGCTCGTCGTTACAGGAA	20	chrII:626402+626795	394bp	s152	Validated
MCM7_R	ATCTTACCAGCGTCCCGTTA	20				

Table 8. Primers used for validating Ctf mutants.