Methods in Continuous Culture:

Mutation Accumulation, Evolution, and Aging

Anja Rebecca Ollodart

A dissertation submitted in partial
fulfilment of the requirements
for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:
Maitreya Dunham, Chair
Alexey Merz
Benjamin Kerr

Program Authorized to Offer Degree: Molecular and Cellular Biology

University of Washington

Abstract

Methods in Continuous Culture: Mutation Accumulation, Evolution, and Aging

Anja Rebecca Ollodart

Chair of Supervisory Committee:

Maitreya Dunham

Genome Sciences Department

There are many ways to culture a single celled organism, each way uniquely capable of answering a different biological question. Continuous culture, a form of growing organisms in which fresh medium is continually added to dilute out the doubling microbe, has existed for decades. As such, many different methods have been developed that utilize the constant dilution. Chapter 1 includes background and history on the methods used in this thesis and what they have been used to address. Chapter 2 describes a multiplexed assay to determine mutation rate in *S. cerevisiae* using the chemostat. I first successfully benchmarked the method against Luria-Delbrück fluctuation tests using a collection of published *MSH2* variants. Msh2 is a DNA repair enzyme that is associated with Hereditary Nonpolyposis Colorectal Cancer. In Chapter

three, I describe the generation of an additional 185 clinically relevant human variants in the yeast ortholog and how I estimated their pathogenicity. I also associated the data with clinical findings. In Chapter 4, I describe the implementation of two previously published devices, the turbidostat and the Miniature-Chemostat Aging Device (MAD). The turbidostat, a device which adds in additional medium as the culture rises above a density set point, successfully determined that aneuploidy is a common suppressor of the slow growth phenotype associated with low ribosomal abundance in *S. cerevisiae*. The implementation of the MAD – a device which removes daughter cells from an aging population of *S. cerevisiae* – was used to probe the proteomic changes associated with age. Lastly, chapter 5 focuses on how all these robustly built devices can be used together or separately to answer interesting biological questions going forward.

Professional Acknowledgements

There are many people I would like to thank for their time, care, and effort in completing my PhD. I would like to thank all Dunham lab members, past and present, for always coming to practice talks, being engaged and helpful during lab meetings, and generally being supportive. I also want to give thanks for the years of comradeship from the Queitsch lab and being a second lab home. I need to thank the graduate school union UAW 4121, the Genome Sciences (GS) graduate council, the MCB steering committee, both the MCB and GS DEI committees, Women in Genome Sciences, and the Genome Sciences Association for the Inclusion of Marginalized Students for their tireless work in support of graduate students. I really appreciate the continued mentorship of my technician mentor Christoph Grundner throughout graduate school. I would like to thank Christopher Large, for generating the whole genome sequencing pipeline and patiently walking me through it. I would like to thank Bryce Taylor for his help editing manuscripts and generally being a calm and understanding presence in the lab. Thank you, Emily Mitchell, for always being game to help with all the devices I was building and Bonny Brewer for being a wonderful collaborator and mentor. I appreciate Chiann-Ling Cindy Yeh, for her help with PacBio sequencing and general coding advice. Huge thanks to Pengyao Jiang for her help with the traditional fluctuation assays and all things mutation rate. I appreciate Mario Leutert, for being my partner in crime on the MADs and reminding me that science is fun. Lastly, I would like to thank Maitreya Dunham for always being in my corner, every step of the way.

Personal Acknowledgements

In personal thanks, I appreciate my cohort of 2015 and specifically Ashley Hall, Dylan Udy, Emma Wren, Erika Hayes, Julia Berkson, Ken Jean-Baptiste, and Mollie McDonnell, for the rock-solid confidence in me and my abilities. Thank you to The Gang™ (Alex Siddons, Cait Hunstman, Chris Busby, Conrad Meyer, Elaine Chen, Ellen Fissel, Erica Olavarria, Geoff Dennis, Kellen Donohue, Megan Lacy, Sam Fout, & Valerie Needham), for always being a slack message away and keeping me grounded. Rachel Elizabeth, for just letting me be me. I would like to thank Maija & Louis Stein, Jessie, Nic, & baby Hartmann, for accepting me into your family and treating me as one of your own. To my extended family including my grandparents Elizabeth & Heinrich Matthys, Lois & Bob Keegan, my aunts Claudia Matthys and Kathy Ollodart, my uncles Urs Matthys, Olaf Smith, and Bruce Ollodart, my cousins Ben & Jenna Ollodart, and Arthur & Philippe Matthys, thank you for being a wonderful family I am so proud to be a part of. I am sincerely grateful to my parents Peter Ollodart and Susanne Matthys-Ollodart, and brother Dave Ollodart for their unconditional love and support. Lastly, I would like to thank Zach Stein, my partner, for helping and loving me every step of the way.

Table of Contents

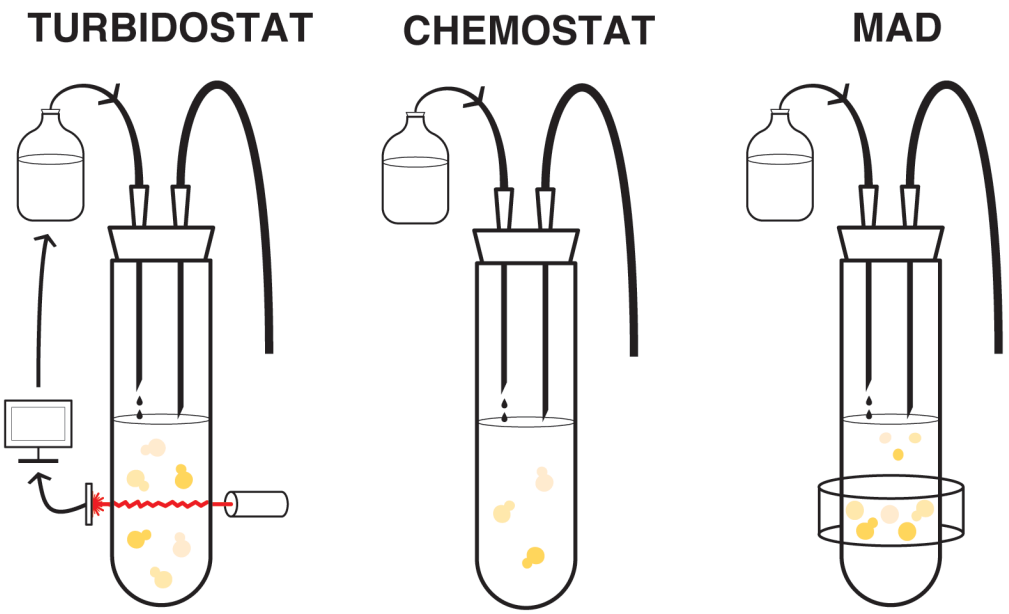
List of Figures

List of Tables

Chapter 1: Continuous culturing techniques and their uses

1.1 Different forms of culturing and the history of how they are built

There are many ways to culture microbes, *S. cerevisiae* in particular. The different ways of culturing can be used to ask different types of questions. For example, viability of a culture is best determined by plating onto solid media and counting colonies. Determining invasive behavior is also best accomplished on solid media, as evidenced by determining the invasive growth characteristics of natural variants of yeast (Hope and Dunham, 2014). Liquid culture can be used to determine growth rates or grow large amounts of cells for genetic analysis. Many different liquid culturing techniques exist. Batch culture involves using different kinds of media which allows for the rapid expansion of a small population cells in simple glass beaker. Chemostats use a nutrient limitation to slow down growth and maintain a steady population size throughout the experiment. Turbidostats allow for feedback on growth, in which the media added matches the doubling time. Miniature-chemostat Aging Devices (MADs) use magnetically labeled yeast to select for the mother cells, which results in an almost pure population of aged cells. The three main culturing techniques used within this thesis and their associated properties are compared in Fig 1.1. Each technique can be used in a variety of ways to answer interesting biological questions, and each has its own unique history.



	TURBIDOSTAT	CHEMOSTAT	MAD
SELECTION	Growth Rate	Environment	Age
FLOW RATE	Variable	Set	Set
GROWTH RATE CONTROL	Intrinsic to Strain Genetics	Nutrient Limitation	Intrinsic to Strain Genetics
GENERATIONS PER DAY	10-14	5-6	10-14
MEASUREMENT AUTOMATION	Yes	No	No

Figure 1.1 A comparison of continuous culture methods used in this thesis. The turbidostat, chemostat, and Miniature-chemostat Aging Device (MAD) were compared across 5 different categories. Selection refers to the limit on growth of yeast as well as what is retained in the vessel. In the MAD, age is selected for by preferentially retaining mother cells and washing out daughter cells. Variable flow rate indicates the amount of medium added is controlled by the density of the culture, versus remaining static throughout the experiment. Growth rate control indicates what the limiting factor on cell division is. Generations refers to how many buds the yeast produces in 24 hours. Measurement automation is whether the optical density must be taken manually or is supplied by the continuous culture device itself.

In a fantastic review by Gabriele Gramelsberger, there is a layout of the very early continuous culture techniques used in labs:

> “Thus, the potential of continuous cultivation for theoretical studies of cell growth was recognized and continuous culture techniques turned into important research tools for microbiology from the 1940s on. The prototypical continuous culture techniques, built in the 1940s and 1950s to grow cells at their maximum rate in order to study them quantitatively, were called ‘‘automatic syringe mechanism’’ (Sims and Jordan, 1942, 1941), ‘‘turbidostat’’ (Myers and Clark, 1944), ‘‘chemostat’’ (Novick and Szilard, 1950a), ‘bactogen’’ (Monod, 1950), and ‘‘microbial auxanometer’’ (Anderson, 1953) respectively.” (Gramelsberger, 2018)

Gramelsbergger goes on to state in this review how these continuous culture devices were useful to determining the mathematical basis of many biological processes, including enzyme and growth kinetics. One of the reasons that continuous culture can be used to quantify processes is because it is so much simpler to accurately count generations and keep the time between them constant. In liquid batch culture, generation tracking is determined by density measurements, flo-cytometry, microscopy, or plating. This does require however frequent observation, and it difficult to keep the generational time constant due to different phases of growth caused by nutrient conditions. By using continuous culture, generations can easily be deduced by tracking the influx of media into the system when it is at steady state. This precise control and tracking capabilities are why many scientists turn to continuous culture to ask scientific questions regarding rates of enzyme kinetics, growth, and evolution.

1.2 Chemostats

Chemostats, as the name implies, keep the chemical environment static. The dilution rate remains constant throughout the experiment, allowing the density to rise or fall when there is a change in the fitness of the cultured organism. This first chemostat was for the culturing of *E.coli*, developed by Novick and Szilard (Novick and Szilard, 1950b) and the same time as Monod, who named it the bactogen (Monod, 1950). Chemostats can have many different chemical environments, with a great many being nutrient limitation (Gresham and Dunham, 2014). Experiments in the chemostat have a set nutrient limitation that specifies the growth rate of microbial organisms, making it more logistically simple to do experiments revolving around enzyme kinetics. Obviously, you get what you select for, so all experiments in nutrient limitation must keep in mind adaptations to the culture conditions themselves if that is not the primary purpose of the experiment. Despite this, chemostats have long been in the tool kit of researchers attempting to understand dynamics of microbial growth.

1.2.1 – Picking your chemostat conditions

Chemostats have been used to answer a wide variety of questions, from flocculation to genome organization to hybrid dynamics. In my thesis I want to focus on studying mutation rate. The reason to use a chemostat to do this is summarized in this perspective by (Fox, 1998), one of the early researchers of determining mutation rate within the chemostat:

> "To the best of my knowledge, steady state devices have been the only systems that provide an opportunity to measure mutation rates directly. In lieu of direct

> measurement, the Luria and Delbrück (1943) approach, made readily applicable by Lea and Coulson (1949), has been and continues to be the method of choice for inferring mutation rates." (Fox, 1998)

Novick and Szilard (Novick and Szilard, 1950b) first developed a way to determine mutation rate in the chemostat based on accumulation of – what they thought – as a neutral resistance to the T5 phage in a tryptophan limited chemostat using *E.coli*. They had tested the T4 phage and found that it was selected against in the chemostat and as a result the One of the main conclusions of this paper was that mutation accumulation was a factor of time, not number of replications. They were surprised by this finding:

> "This result is not one that could have been foreseen. If mutants arose, for instance, as the result of some error in the process of gene duplication, then one would hardly expect the probability of a mutation occurring per cell division to be inversely proportionate to the rate of growth. If the processes of mutation could be considered as a monomolecular reaction-as had been once suggested by Delbruck and Timofeeff-Ressovsky then, of course, the rate of mutation per unit time should be constant." (Novick and Szilard, 1950b)

They go on to say that if it is indeed a monomolecular reaction- aka a reaction in which a single molecule reacts at a constant speed, it would be characterized as having a constant 10^8 times lower than any other monomolecular reaction, and thus further experiments in other limitations and with other readouts were needed. They did these additional experiments in a second manuscript focusing on nitrogen and phosphorus limitation and did not see the same effect of growth rate on mutation rate (Novick and

Szilard, 1951). Further work showed that this correlation with time and not generations was only a factor in tryptophan-limited chemostats (Fox, 1955; Kubitschek and Bendigkeit, 1964), though to my knowledge I don't think they figured out exactly why. All of this to say, it's important to pick resistance marker and media conditions correctly, to accurately determine the effects of a genetic or environmental perturbance on mutation rate. With this in mind, I looked to studies on yeast mutation rates within the chemostat. Work by Paquin and Adams showed that resistance to canavanine accumulated linearly in a glucose-limited chemostat at the mutation rate ascertained Luria-Delbrück fluctuation assays. However, after about 50 generations, they saw a complete removal of the canavanine resistant population, and posited that this was due to a beneficial mutation arising in the canavanine sensitive population and outcompeting the canavanine-resistant population(C. Paquin and Adams, 1983). They followed this experiment out to many generations and saw that this behavior- resistance to canavanine rising at the mutation rate and then being swept out- was cyclical. They repeated this experiment using resistance to cycloheximide and showed much quicker and more frequent turnovers in the canavanine resistant population (Adams et al., 1985), indicating that once again, the medium used in these assays in is important. Based on the results of this paper, I was reasonably confident that I could determine mutation rate using the chemostat.

1.2.2 – Mutation rate determination by chemostat ideal to study mismatch repair complex

Once I settled on the experimental conditions, a range of biological questions were accessible. This method could be applied to studying the effects on mutation rate . I focused on the effects of variation in the protein coding sequence of DNA repair enzymes, specifically of Msh2. For more information on the ways that Msh2 has been looked at in different systems, please see section 2.1, and for information on the clinical relevance of Msh2 see section 3.1. A huge genetic and biochemical literature centers on proteins in the MMR pathway due to their basic science and clinical relevance. Much of pioneering work has been done in *S. cerevisiae*. Msh2 is part of the mismatch repair complex in yeast which comprises 5 conserved proteins: Mut-L Homologue 1 (Mlh1), postmeiotic segregation 1 (Pms1), Mut S Homologue (Msh)2, Msh3, and Msh6. Mut-L and Mut-S are the core mismatch repair proteins in *E.coli*, where this system was first discovered . These proteins are responsible for finding and fixing small indels and mismatches (as reviewed in (Boiteux and Jinks-Robertson, 2013; Kunkel and Erie, 2005; “The Nobel Prize in Chemistry 2015,” n.d.). Msh2 recognizes all mismatched base pairs and single-strand indels, while Msh3 is recruited for mismatched base pairs and large indels. Msh6 recognizes mismatched base pairs and small indels. The partial redundancy of Msh3 and Msh6 is likely responsible for the reduced effect of each single knock out, both *msh3Δ* and *msh6Δ* have modest increases in mutation rate. *msh3Δ msh6Δ* show similar mutations rates to a single *msh2Δ* and the triple *msh2Δ msh3Δ msh6Δ* does not appear to have a much higher mutation rate than *msh2Δ* (Johnson et

al., 1996; Marsischky et al., 1996). Once the Msh2 dimer with either Msh3 or Msh6 assembles on the lesion formed by the mismatched base or indel, Mlh1 (sometimes referred to as Psm2) and Pms1 are recruited via a binding domain on Msh2 (Mendillo et al., 2009). These proteins are responsible for nicking the DNA, and recruiting proliferating cell nuclear antigen (PCNA) and exonuclease 1 (Exo1) to remove the mismatched base pair (Kadyrov et al., 2007, 2006). In *E.coli*, it is indicated that the cell differentiates between the newly generated strand which is most likely to contain the error by its methylation state (as reviewed in ("The Nobel Prize in Chemistry 2015," n.d.). In yeast, there is evidence that the direction of nicking and repair is determined by the loading of PCNA and the free 5' prime ends of Okazaki fragments (Pavlov et al., 2003; Pluciennik et al., 2010). Lastly, polε is recruited to replace the correct bases (Bowen and Kolodner, 2017). While a lot of proteins in the mismatch repair complex also have function outside mismatch repair, such as in resolving crossover events in meiosis (Boiteux and Jinks-Robertson, 2013), the focus of my thesis was on their role in mismatch repair. The method I developed could theoretically work for any of the proteins in the MMR, however I do have a concern for the relatively weak mutator phenotypes of *msh3Δ* and *msh6Δ* single mutants, as the protocol is heavily dependent on a large canavanine resistant population to get accurate coverage of variants. As a result, I would suggest studying these proteins using a double knock out background. Thus, this project first focused on Msh2, both because of stronger mutator phenotype and its clinical relevance. While I developed this assay to look at the effects of variation in Msh2, my hope was that it would be of broader use to study the effects of any

perturbation, be it genetic or environmental, on mutation rate. This is a multi-use culturing technique which can be used on its own, or in combination with other culturing techniques to get a comprehensive view of biology.

1.3 The history and uses of turbidostats

Turbidostats, also known as bioreactors and a very similar device called a fermenter, are continuous-feedback culturing devices. They feed back on turbidity – thus the name turbidostat – the turbidity remains static. From the first instance in 1944 of a turbidostat for the photosynthetic algae Chlorella (Myers and Clark, 1944) to the turbidostat implemented during my thesis (McGeachy et al., 2019), the concept has stayed the same, even if the technology has improved. There have been many different types of build-it-yourself turbidostats constructed for different purposes. Smaller volume turbidostats are useful for determining the effects of many different conditions on a culture, as their small nature makes them amenable to multi-plexing. Some examples include the 20ml turbidostat developed by Takahashi (Takahashi et al., 2015), which includes the ability to track fluorescence as well as absorbance. Wong took this design a step further with a manifold that can mix various media sources, and heat sleeves which can clamp or ramp temperature (Wong et al., 2018). He went on to identify genes that are responsive to high temperature stress in this system. Both technical innovations allow for precise control of environmental conditions, especially useful when trying to determine gene-by-environment interactions. The Omni-stat, developed by Ekkers in the Sander Von Doorn lab, has a variable culture size: 25ml to 250ml (Ekkers et al., 2020). In addition, it has many ports which allow for additional sensors, such as pH and

oxygen, which would make this device well-suited for studying physiological changes in the culture.

My choice to implement the 200ml turbidostat designed by Anna McGeachy and Nick Ingolia was driven by a combination of ease of implementation and the ability to travel to the lab in question and build it with them. Originally built to complete a deep mutational scan on RPL28, an essential ribosomal protein which binds cycloheximide, it appeared to be the perfect device to answer both evolution questions as well as to perform screens (McGeachy et al., 2019). In addition, the output was simple and straightforward, and the Ingolia lab had shown that it could be run for many hours without leaks, crashes, or other mishaps. Lastly, the relatively large volume is useful for selection during suppressor screens, and the lab, as well as I, were focused on determining the effects of variants in mismatch repair, aneuploidy, and liver enzymes. It seemed as if building a 200ml, robust, easy to implement turbidostat would not only be useful for my thesis, but to the lab as a whole.

1.3.1 – Use of the turbidostat to study effects of rDNA perturbation

In a collaboration with Bonny Brewer, we wanted to understand how yeast responded to a severe shortage of ribosomes, caused by editing the rDNA gene array. Ribosome production is a key requirement for cell function, and disfunction in the rDNA array has implications for human health (Bahadori et al., 2018). In yeast, extra chromosomal rDNA circles, which are pieces of the repetitive rDNA locus which have recombined into a plasmid like structure, and reduced rDNA array stability are associated with a reduction in the number of buds a yeast cell can produce (Ganley et al., 2009). Genetic

manipulation of the rDNA locus has traditionally difficult. Joe Sanchez, a graduate student in the Brewer lab, used CRISPR to generate two rDNA edited strains: a full deletion of the rDNA locus, or a replacement of the AT-rich rDNA origin with GC (Sanchez et al., 2019). While edits to the rDNA array had been made before using a cassette containing a variant of rDNA which is hygromycin resistant (Chernoff et al., 1994), this strain represented the first time we were able to determine the effects of an entire deletion. The original strain had approximately 135 rDNA repeats and a doubling time of ~90minutes. Both of the edits made resulted in a reduction of the rDNA locus to 10 repeats and a doubling time of approximately 195 minutes (Sanchez et al., 2019). The rDNA edits had a direct effect on growth rate, making it simple to pick up suppressors, as they would increase the growth rate and thus the density of the culture, which is what the turbidostat tracks and feeds back on. Dr. Brewer had evolved these strains via serial batch propagation, diluting the samples such that they stayed in log phase for the duration of the experiment (for more information on results see section 4.2), so we were interested to find out whether the turbidostat could be used to replace or compliment batch culture evolutions. Serial back-dilution is a common way of evolving strains, the most famous and longest-running being the Lenski lines in *E.coli* (Blount et al., 2018). However, serial dilution provides some obstacles if the goal is even selection over time, including the greater potential for contamination to avoid freeze thaws and cyclic differences in density and nutrient availability. We wanted to evolve these rDNA edited strains to determine what, if any, suppressors arose to offset the large increase in doubling time, and if the turbidostat would be able to track the increase

in the suppressor-containing population in real time by the increase in media usage. Once proven robust enough to be run for many generations, the turbidostat developed here can be used to answer many different kinds of questions.

1.4 High throughput isolation of aged cells: current and past technologies

The last type of continuous culture device I implemented is the Miniature-chemostat Aging Device (MAD). Instead of mother and daughter cells being indiscriminately diluted out, as in the previously discussed methods, magnetically labeled cells are stuck to the side of the culture tube via a magnet. Daughter cells do not inherit this magnetic label because it is attached to amino groups of proteins embedded in the cell wall of the mother cell, and the cell wall and associated proteins are newly synthesized in the daughter cells. This asymmetric inheritance allows for the retention of mother cells by magnetic force while daughter cells are washed off, which in turn allows for a large collection of cells over time which are replicatively aged. The method was developed in *S. cerevisiae*, where a single yeast cell can produce a limited number of progeny before it senesces (Mortimer and Johnston, 1959). This is a good proxy for aging, a complex phenomenon which is made genetically tractable by using a fast-growing model organism with well-developed genetics. A review reflecting on yeast studies contribution to aging research states:

> "[…] the downstream pathways discovered to date [in yeast], such as regulation of stress responsive transcription factors (Msn2/4, Gis1, and Gcn4), reduced translation, enhanced autophagy, control of oxygen radicals, and the protective

responses they invoke have conserved effects on aging in multicellular eukaryotes." (Longo et al., 2012)

In order to further develop these aging studies, there is a great need to procure a sizeable amount of aged yeast cells. The concept of magnetically labeling a small population of yeast in order for later retrieval was first reported in 1996 (Smeal et al., 1996). Further purification by labeling with a dye that stains bud scars, and using FACS to isolate brighter cells which would have more budscars, allowed for even greater purity of old cells (Chen and Contreras, 2007). Later, the Gottschling lab (Lindstrom and Gottschling, 2009) developed the "mother enrichment program," which uses some genetic tricks to cause cell arrest only in daughter cells. These systems were all used in batch culture, and mother cells were subsequently isolated. Later, another group developed a method in which media is washed over magnetically labeled yeast held in place by magnetic force, which results in retention of mother cells and elution of unlabeled daughters (Janssens et al., 2015). This system, vs batch culture, has the advantage of maintaining a constant chemical environment. The system used in my thesis is based off of the Miniature-chemostat Aging Device, or MAD system, developed at Calico (Hendrickson et al., 2018), in which they combine the aging system developed by Janssens et al. and the mini-stat set up developed by Aaron Miller, a former grad student in the Dunham lab (Miller et al., 2013). The result is a system which can produce aged cells which have been kept in a constant environment. The original purpose of the MAD system was to generate aged cells for genomics and transcriptomics, and they were able to determine how the transcriptome changed with

age. In our collaboration with the Villén lab, we wanted to take the system a step further to do whole-scale proteomics.

1.4.1 – The effect of replicative age on budding yeast

The mechanism behind aging in yeast cells has been a point of much study for last several decades. The most basic question is: Is aging a matter of just random entropic breakdown, or are cells programmed to age? With the establishment of a replicative age- defined as the total number of buds a cell makes- of 25-30 divisions in commonly used budding yeast strains (Mortimer and Johnston, 1959), researchers worked to dissect the mechanisms underlying this aging. Aging factors were established as dominant, indicating that there must be a segregation of proteins between mothers and daughters for full rejuvenation of daughter cells (Müller, 1985). It was shown that there was an active barrier which prevented the flow of extra-chromosomal circles and nuclear pore complexes. Removal of this barrier results in the failure of full rejuvenation of the daughter cells (Shcheprova et al., 2008). In a review from 2014, Lippuner et al. state:

> "The decreased rejuvenation of daughters produced by very old mothers could reflect age-induced defects of the molecular machinery involved in retention or titration of the retention machinery by large amounts of the aging factors. Interestingly, while the replicative lifespan of the last bud is decreased, its first daughter and granddaughter show a gradual restoration of a normal lifespan (Kennedy et al., 1994) suggesting that one or more factors must be diluted to achieve full rejuvenation. Together, these observations have led to the general

paradigm that aging is caused by the accumulation of aging factors." (Denoth Lippuner et al., 2014)

The lack of rejuvenation of daughters from old mothers may be because of a breakdown of the barrier, or simply that so much of the protein exists that it spills over into the daughter cell. While a lot of the work to determine the mechanisms behind aging has been done on a protein-by-protein basis, recently more large-scale global studies have been completed. The Gottschling lab performed mass spectrometry-based proteomics and found ~135 proteins which they coined long lived asymmetrically retained proteins (LARPS). They found that the majority of these LARPS were fragments of longer proteins which localized to the plasma membrane and in large cytoplasmic structures, and that they may be actively causing an aging phenotype (Thayer et al., 2014). (Whether these proteins meet in isolated areas of the endoplasmic reticulum to pretend to be *E. coli* proteins is unknown. This is a Joke™) Whole proteome analysis by the Heinemann lab compared transcript abundance to protein levels of yeast at different replicative ages. In a model suggested by their data, decoupling of transcript and protein levels of biogenesis proteins drives aging, by causing disruption of important stoichiometry between protein complexes which results in metabolic changes and activation of stress responses (Janssens et al., 2015). In my thesis, in collaboration with Mario Leutert, we sought a more comprehensive understanding of age-dependent proteome changes with innovations in proteomics technology, as well as probe the kinase networks that may be changed with age by doing phospho-proteomics, which identifies proteins with added phosphor groups used in regulation. While the MAD was

generated to understand proteomic changes during aging, it can be broadly useful in determine aging phenotypes. The MAD, the turbidostat, and the chemostat multi-plexed mutation rate methods are all capable of studying a great variety of problems, in combination or on their own.

Chapter 2: Developing a multi-plexed mutation rate assay

This chapter is part of a manuscript that has been submitted. It was primarily written by me, Anja Ollodart, with edits by co-authors on the manuscript Chiann-Ling Cindy Yeh, Adam Gordon, Aaron Miller, Brian Shirts, and the corresponding author Maitreya Dunham. Thus, the tone of this chapter may differ from Chapter 1, 4, and 5 as it was crafted by multiple people.

2.1 Abstract

Despite the fundamental importance of mutation rate as a driving force in evolution and disease risk, common methods to assay mutation rate are time-consuming and tedious. Established methods such as fluctuation tests and mutation accumulation experiments are low-throughput and often require significant optimization to ensure accuracy. We established a new method to determine the mutation rate of many strains simultaneously by tracking mutation events in a chemostat continuous culture device and applying deep sequencing to link mutations to alleles of a DNA repair gene. We applied this method to assay the mutation rate of hundreds of *Saccharomyces cerevisiae* strains carrying mutations in the gene encoding Msh2, a DNA repair enzyme in the mismatch repair pathway (MMR). We first benchmarked our method against Luria-Delbrück fluctuation tests using a collection of published *MSH2* variants. Our

pooled screen successfully identified previously-characterized non-functional alleles as high mutators.

2.2 Introduction

Mutation rate is the timer for many different error-prone processes: how many cycles of PCR before the polymerase makes a mistake, how long before the bacterial infection becomes resistant to existing medications, or how quickly will DNA damage result in the uncontrolled growth of a cancerous tumor.

However, existing methods for measuring mutation rate are tedious and not scalable for the challenge of functionally testing hundreds or thousands of Variants of Uncertain Significance (VUS). Methods to study mutation rate all have their benefits and detriments (Foster, 2006). In microbial systems, Luria-Delbrück fluctuation tests and mutation accumulation lines are two of the most used (Luria and Delbrück, 1943; Lynch et al., 2008). All of these methods require initiating populations with single clones to determine the effect of a mutation, which limits the ability to multiplex experiments.

In addition to ease of use, yeast is a good model system to study effects on MMR because much of the sequence and function of the pathway is conserved between humans and yeast (Boiteux and Jinks-Robertson, 2013). Many discoveries about MMR originate in studies of *S. cerevisiae,* as the MMR complex is more highly conserved with its human orthologs than that of *E.coli* (Strand et al., 1995, 1993). In addition to general biology of the MMR complex, yeast has been used to determine the mutation rate of MMR alleles using traditional fluctuation assays, mutation accumulation lines, qualitative patch assays, and fluorescence-based assays (Demogines et al., 2008; Drotschmann et

al., 1999a; Gammie et al., 2007; Lang et al., 2013; Lang and Murray, 2008; Martinez and Kolodner, 2010; Shor et al., 2019). These assays all test the full functionality of Msh2, however none of them allow for pooled assessment of many alleles simultaneously. Although medium-throughput assays exist that take advantage of automated liquid-handling systems (Gou et al., 2019), these still require considerable effort if studying the effect on mutation rate of many different alleles and require keeping clonal populations.

In an attempt to improve on the scaling problems of other methods for measuring mutation rate, we have developed a chemostat-based assay that utilizes pools of variants to assay hundreds of alleles simultaneously. The chemostat is a continuous culture device which maintains a constant population size over time. The first use of the chemostat to determine mutation rate was over half a century ago (Fox, 1955, p. 1; Kubitschek and Bendigkeit, 1964; Novick and Szilard, 1950b; C. E. Paquin and Adams, 1983). The continuous dilution in the chemostat means, on average, an increase in the frequency of a neutral mutation is a result of *de novo* mutation, as opposed to other methods, where such an increase could be explained by both *de novo* mutation and exponential growth and thus requires many replicates to differentiate between the two. If the neutral mutation is also selectable, such as with some types of drug resistance (e.g. canavanine resistance in yeast), the mutation rate can be calculated by simply tracking the frequency of resistance over time. Combining this very old technique with next-generation sequencing allows for the pooling and therefore high-throughput study of alleles. We applied this new method to study mutation rate differences caused by

variants in *MSH2*, a gene that is associated with HNPCC. Msh2 is a part of the MMR complex, which in combination with Msh3, Msh6, Mlh1, and Pms2 binds and fixes small mismatches and indels (Boiteux and Jinks-Robertson, 2013). Msh2 is an integral part of the recognition complex (Edelbrock et al., 2013). We completed a proof of principle with previously published variants of *MSH2* and found that the pooled assay recapitulated the results of traditional Luria-Delbruck fluctuation tests, qualitative patch assays, and yeast two-hybrid assays (Gammie et al., 2007). These data taken together show that our new multiplexed mutation rate assay is an accurate and scalable assay to study the mutation rate of many strains in a pooled format.

2.3 Results

2.3.1 – A new assay to determine mutation rate

The chemostat is a continuous culture device which matches the growth rate of an organism to the dilution rate, stabilizing population size and environmental conditions throughout an experiment. Many ways to study mutation rate take advantage of drugs for which WT organisms are sensitive but a loss-of-function mutation causes resistance, which makes it straightforward to track mutation frequency (Boeke et al., 1987; Whelan et al., 1979). However, to determine rate, one must know the number of generations that have elapsed since the mutational event, which is difficult in batch culture. Luria-Delbrück fluctuation assays and mutation accumulation lines use different tactics to determine the generational time. In continuous culture, since the population size stays stable, an increase in resistance, isn't due to an increase in a lineage, as long as certain underlying assumptions are met, as described in the following section. In the

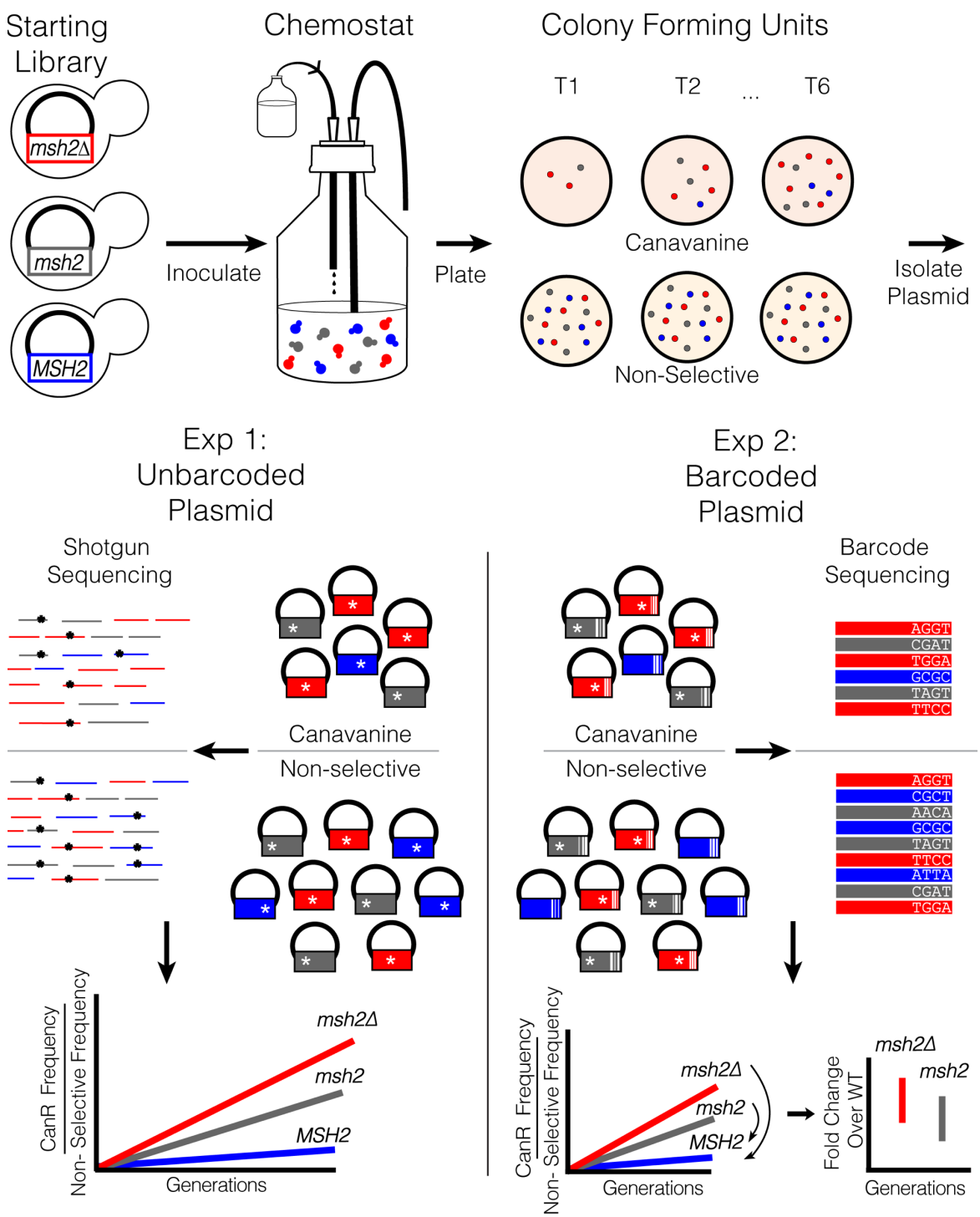


Fig 1. A schematic outlining the multiplexed mutation rate method. A pool of alleles is inoculated into a chemostat. Samples are plated onto non-selective media or media containing canavanine to select for LOF mutations in *CAN1*. Plasmid is isolated from the canavanine-selected plates as well as from the non-selective pool. The assay can handle both unbarcoded and barcoded plasmids, using shotgun or barcode sequencing respectively. In both cases the frequency of the allele on selective canavanine media is divided by presence in the total pool and tracked over time to generate the mutation rate. With barcoded plasmids, barcoded WT can be used to determine the fold change of variants against an internal control.

assay we have developed, outlined in Fig 1, we can track many lineages in a pooled manner to determine all of their mutation rates at once. We do this by keeping track of *de novo* mutational events on selective media containing a drug, while controlling for any changes in the overall population size of a given lineage by monitoring growth on non-selective media. Next-generation sequencing of the plasmid recovered from both the non-selective and selective media allows us to track the various lineages over time. This assay is amenable to both barcoded and unbarcoded libraries. With unbarcoded libraries, we use shotgun sequencing of the allele isolated from the plasmid, using the mutation within the gene itself as a way to track the variant over the course of the experiment. In barcoded libraries, the barcode and variant are first linked using long read sequencing, after which amplicon sequencing of just the barcode reveals the frequency of each variant at each time point. Our method is amenable to both types of analysis to make it more generalizable. In both cases, the increase in resistance frequency over time for all lineages can be calculated, giving us their mutation rates. Our first application of this method utilizes *Saccharomyces cerevisiae* and focuses on variants of Msh2, a clinically relevant DNA repair enzyme. However, this assay is amenable to the study of any microbial strains which can be cultured within the chemostat and any molecular pathway that yields a neutral selectable mutation as a phenotypic readout.

2.3.2 – *Msh2Δ* confers minimal fitness cost in glucose limitation

If we are to multiplex this assay, null-like *msh2* variants should not have a large fitness effect, otherwise we cannot determine the difference between a *de novo* mutation and

expansion or contraction of a resistant lineage. In a head-to-head competition between WT and *msh2Δ*, we found a 1.097% fitness defect of *msh2Δ* over the short course of our experiment (Figure 2.3.2). This means we will likely slightly underestimate mutation rate of high mutators. However, by correcting our canavanine resistant frequency by the relative population frequency of each variant, we can mitigate the effects of both the strain manipulation and resistance to canavanine.

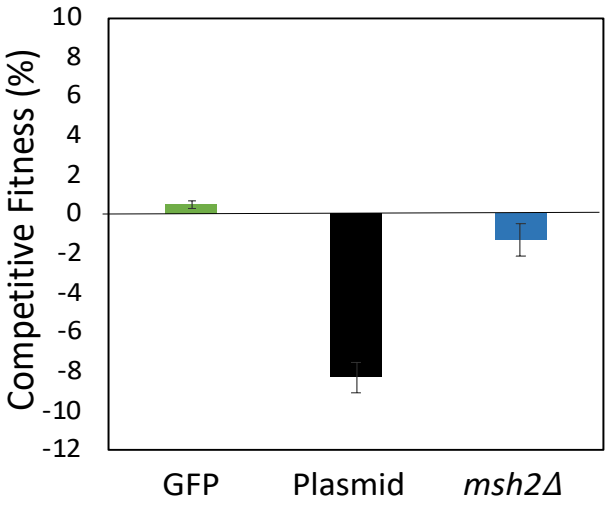


Figure 2.3.2 Determining fitness effect of *msh2Δ*. Competitive fitness is in comparison to the wt FY4 strain used as the background for all experiments. GFP was used as the read out for comparison in the head-to-head competition and serves as a negative control as it doesn't exhibit a fitness effect. Maintenance of the plasmid is known to be costly, so it served as a positive control for a fitness effect. The competition between a *msh2Δ* and *MSH2* strain revealed a1.5%/generation fitness effect of being a *msh2Δ* mutant.

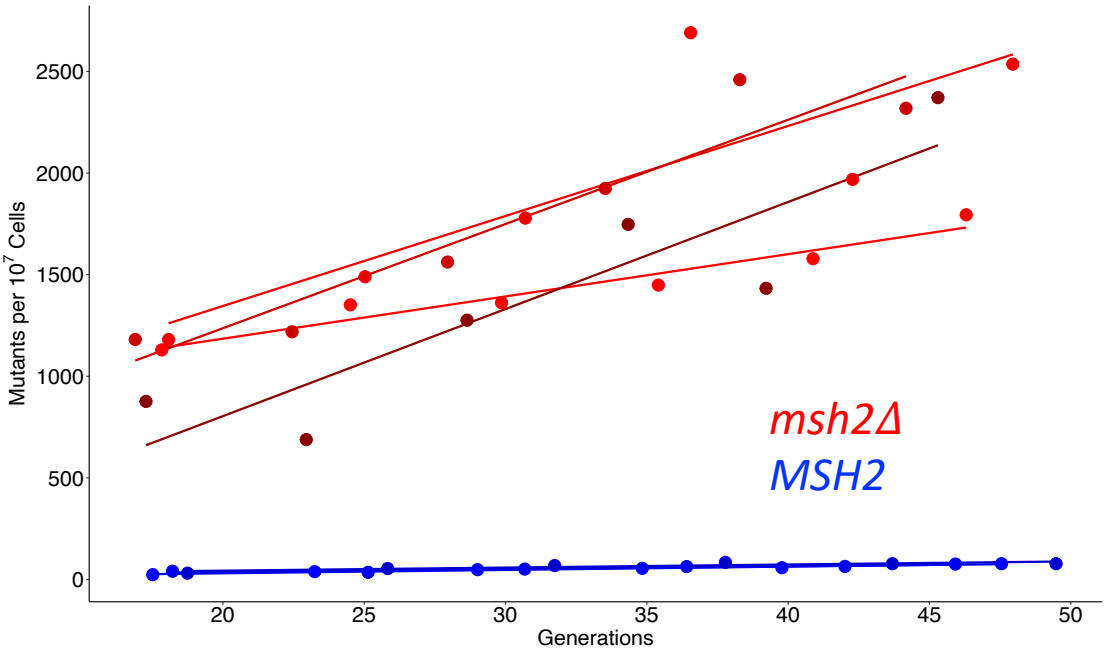


Figure 2.3.3 Determining the linear range of mutation accumulation. Chemostats were inoculated with the stated genotype and CanR / generation was tracked over the course of the experiment

2.3.3 – 200ml chemostats inoculated with control strains accumulate resistance at expected rates

To determine the timeframe over which we could observe linear accumulation of resistance, we assayed the mutation rate of *msh2Δ* strains containing either WT *MSH2* (*MSH2)* or pRS413 (*msh2Δ*) using the same conditions as future pooled experiments. Previous work has shown complementation by plasmid-borne WT *MSH2*, and that variants that abrogate activity elevate the mutation rate in yeast (Strand et al., 1993; Drotschmann et al., 1999b; Gammie et al., 2007). Chemostats were individually inoculated with *MSH2* and *msh2Δ* strains, and samples were plated every 24 hours to determine which timepoints correspond to the range for linear accumulation of Can^R mutants (Fig 2.3.3). We found that between 12 and 45 generations, resistance to canavanine accumulates at very similar rates as stated in previous literature (Table 2.3.3) (Gammie et al., 2007). The lag in linear accumulation can be explained by a requirement for cells to reach steady state, at which point they are growing and being diluted at the

Table 2.3.3
Mutation rates of control pure cultures and pools

Relevant Genotype	CanR Rate[a]	Fold Induction CAN
MSH2	$1.52 \times 10^{-7} \pm 0.07$	1
msh2Δ	$49.44 \times 10^{-7} \pm 2.6$	29
Unbarcoded pool	$21.62 \times 10^{-7} \pm 2.53$	13
Barcoded Pool in *msh2Δ*	$39.6 \times 10^{-7} \pm 2.1$	26
Barcoded Pool in *MSH2*	$2.06 \times 10^{-7} \pm 0.05$	1.3

a. Rate, mutations per generation

same rate (reviewed in (Gresham and Dunham, 2014). In these specific conditions, after 45 generations, selection on adaptive mutations is likely the cause of the non-linear increase (C. E. Paquin and Adams, 1983). From this, we determined that all future experimental timepoints must be taken between 12 to 45 generations to accurately determine the mutation rate.

2.3.4 – Mutation rate assessment in 20ml chemostats noisy and unreliable

Once the mutation rate had been established in 200ml chemostats, we wanted to know if we could use the mini-stat array (Miller et al., 2013) to be able to generate many biological replicates at once. Individual mini-stats were inoculated with the stated genotype, and their mutation accumulation and slope of canavanine resistance accumulation are represented in Figure 2.3.4. This data represents the increase in the canavanine-resistant population which has not been normalized to the total population due to errors in the total population counts. It appears that while on the whole the data makes sense, with high and low mutators clustering as expected, there is a large variation in the values obtained within this data. In some cases, we have generated negative slope values which indicate that the canavanine-resistant population is being washed out, or that the overall population size is growing faster than the resistant population. This may be the result of the small population size being more subject to drift, a logistical hurdle of needing to sample for longer in order to achieve proper counts, or a result of difficulty processing that many samples at a given time. However, with additional work the mini-stats may be capable of determining the mutation rate of individually inoculated alleles and may present a nice validation of pool findings.

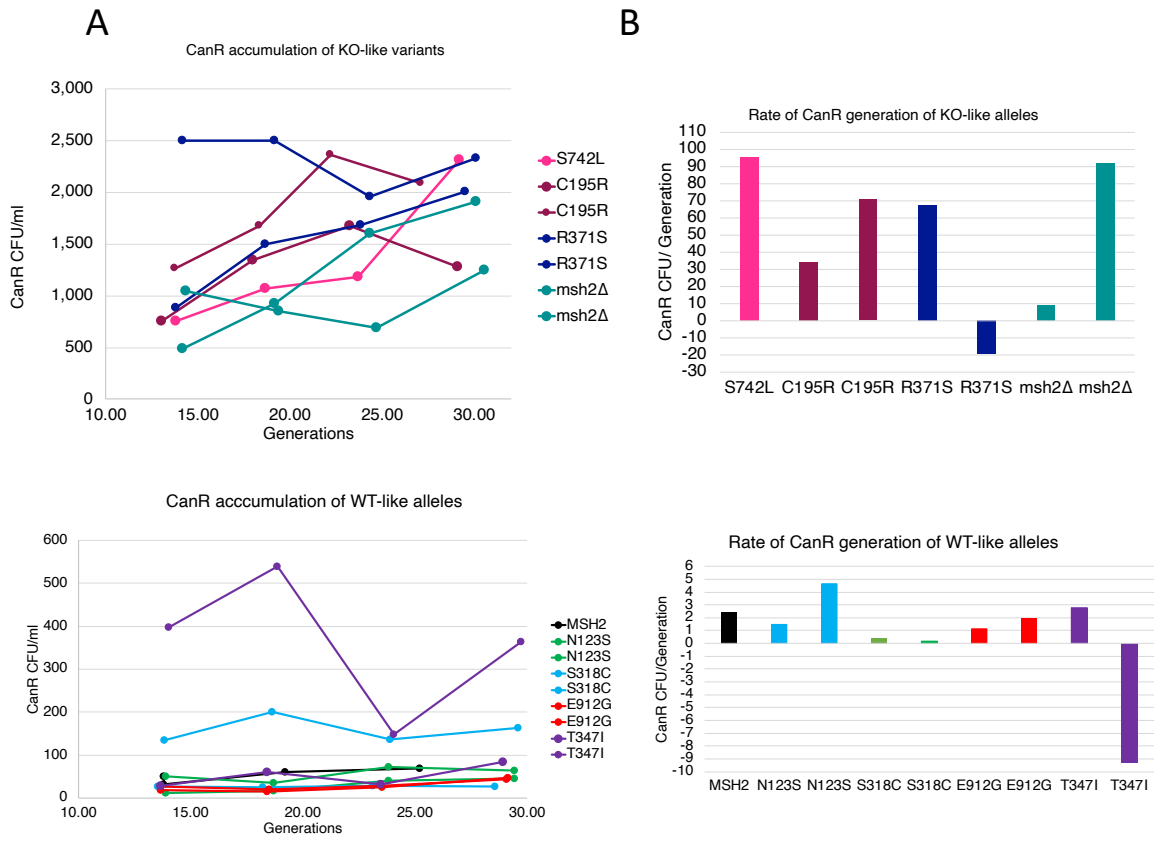


Figure 2.3.4 Mutation accumulation within the mini-stat produces data that is inconsistent but generally correct. A. The number of canavanine resistant colonies/ml were plotted against the number of generations. This is not normalized by total population. The data was split by KO-like and WT-like alleles. B. The rate of increase in canavanine resistance per generation of KO-like and WT-like alleles.

2.3.5 – Pools of existing unbarcoded plasmid accumulate mutations at a rate between WT and deletion control strains

Mutation rate can vary even under very similar conditions, and thus multiple replicate assays must be done to obtain accurate measurements. We created a pooled assay to easily increase the number of replicates for each individual mutation rate assessment. We first established a proof of principle assay with 46 previously published alleles of *MSH2* (Gammie et al., 2007) (Fig 2.3.5). We pooled equal amounts of strains carrying each allele, inoculated aliquots of the pool into 4 independent 200ml chemostats, and from each collected 6 samples on selective and nonselective media over 50 generations, as determined from the control experiments described above. Plasmid was isolated from all samples, plasmid borne *MSH2* alleles were amplified by PCR and

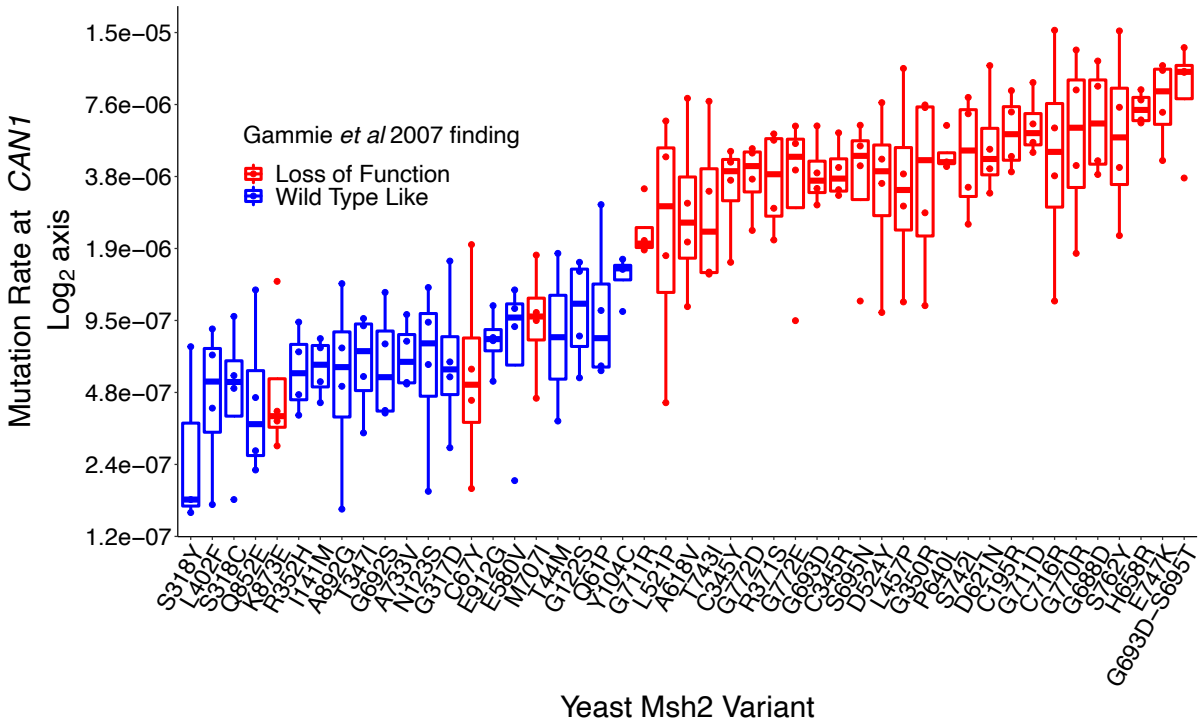


Fig 2.3.5 Calculation of mutation rates of indicated alleles using multiplexed mutation rate assessment. Mutation rate (CanR per generation) of previously studied alleles, colored by the phenotype found in Gammie et al. 2007. Mutation rates are plotted on a $\log_2$ axis and points represent measurements from separate chemostats.

subjected to shotgun short read sequencing, and the frequency of each allele from the canavanine resistant pool was normalized by the frequency in the non-selective pool and converted into colony counts (see methods).

The average mutation rate of this pool is 13X over wild type (Table 2.3.3), which corresponds with what is expected in a pool containing both loss-of-function and WT-like alleles. In Figure 2.3.5, the mutation rates of each allele as measured at *CAN1* are plotted. We compared the results to the phenotype found in previous work using qualitative patch assays, Luria-Delbrück fluctuation tests, and yeast two-hybrid assays (Gammie et al., 2007). With the exception of 3 alleles (K873E, C67Y, and M707I), alleles previously described as WT-like all grouped together, as did the loss-of-function alleles. C67Y was classified as loss-of-function due to a lack of subunit interaction and a qualitative patch assay in previous work. This lack of subunit interaction may not be reflected in our mutation rate assay and perhaps explains the lack of correspondence to the qualitative measurement. K873E and M707I both showed a LOF phenotype measured at *CAN1* but were found to be WT-like when testing for dinucleotide instability. These alleles exist at the edge of the classification between loss of function and WT-like in the compared study and could potentially explain why the results are discordant. These data show a grouping of the high mutators and low mutators, indicating that our new method can largely replicate the results of previous efforts to measure allele-specific effects on mutation rate.

2.4 Discussion

We developed a new method for high-throughput mutation rate assessment that combines a mid-20th century method to determine mutation rate with 21st century next generation sequencing. This allows for the pooling and multiplexing of mutation rate assessment-which has not been accomplished before. We were able to determine the mutation rate of over 40 variants of *MSH2*, a critical component of mismatch repair, without the use of barcodes. Though we included a high frequency of WT sequences, our analysis indicates many of these could be substituted with additional VUS to increase throughput at minimal cost to accuracy. The assay is limited by the canavanine-resistant subpopulation within a 200ml chemostat, which is dependent on the mutation rates of the lineages in a pool. One could increase the number of variants to be assayed in a single experiment by increasing the volume of the chemostat, though the logistics of expanding the volume beyond the 2L size of available commercial fermenters may be difficult. Another possible modification would be to utilize alternative marker loci that generate selectable mutations at higher rates than the WT *CAN1* sequence. In addition, the inclusion of barcoded null mutants may provide an additional control to better normalize the results to established mutation rates.

Our method should be widely applicable and can be used to answer many other questions associated with mutation rate outside of clinical variant interpretation. Accurate, multiplexed measurement of mutation rate variation could be used to screen polymerases for increased or decreased fidelity, to screen the yeast deletion collection for knock-outs that increase mutation rate, or to uncover differences in mutation rate

among natural variants of yeast. In conclusion, this method is broadly applicable to many different problems in which mutation rate is a factor and can be used to estimate the pathogenicity of clinically relevant DNA repair enzymes.

2.5 Materials and Methods

2.5.1 – Running competition experiments

Competitions were set up by individually inoculating glucose limited mini-stats with 1ml of saturated culture of each stated strain. The strains were allowed to grow up for 2 days before the pumps were turned on. After reaching steady state after ~ 10 generations, cultures were mixed half and half and GFP percentage was monitored twice a day via flow cytometry. Fitness effects were calculated by taking the slope of the natural logarithm of the GFP tagged to non-GFP tagged strains over time.

2.5.2 – Determining mutation rate in the chemostat

For individually inoculated chemostats both 200ml and 20ml, 1ml of overnight culture was inoculated into 230-245mls of glucose-limited media (Calcium Chloride 4g/L, Sodium Chloride 4g/L, Magnesium Sulfate Heptahydrate 20g/L, Potassium Phosphate Monohydrate 40g/L, Ammonium Sulfate 200g/L, Glucose 32g/L). Pools were thawed from the freezer and inoculated straight into the chemostat. Cells were allowed two days to grow to saturation and the pumps were turned on to a rate of 40mls/ hour or ~ 5 replacement volumes per day. Chemostats were allowed to reach steady state as determined by a stabilization in the CFU counts and optical density and samples were then collected starting at 15 generations. Sampling the non-selective population involved spinning down ~ 2 X 10^8 cells, as well as plating ~ 200 cells onto SC -histidine

for accurate counts. For each chemostat, to determine the number of canavanine resistance mutants, sufficient culture to reach an estimated countable number of colonies (approximately 200) was plated onto SC -arginine -serine -histidine + 60mg/L canavanine to select for loss-of-function mutations in *CAN1*. For pools, ~6 x 10^8 cells were plated, in addition to those used for counts, onto 15cm SC + canavanine plates and then allowed to grow at 30°C for 3 days at which point they were scraped for downstream analysis.

2.5.3 – Generating unbarcoded pools of plasmids

DNA was extracted from *E. coli* strains sent from Alison Gammie and transformed into DBY11069-FY4 *msh2Δ* and *his3Δ* ~ 20x coverage.

2.5.4 – Generating shotgun sequencing libraries

For both the non-selective and mutant population, cells were vortexed vigorously with acid washed beads in resuspension buffer for 3 minutes, and then put through the Mini-prep Wizard kit. They were then concentrated using the PCR clean up Wizard kit.
For unbarcoded pools, *MSH2* amplified from the plasmid vector using 15 rounds of PCR to prevent over-amplification. Then Nextera sequencing libraries were generated using the Nextera-XT kit. The average library size was 500BP, and sequenced using a Nextera 500 with paired end 150BP reads at a depth of ~30,000 reads over the length of *MSH2*. The run was conducted according to manufacturer specifications.

2.5.5 – Analysis pipeline for unbarcoded libraries

Reads were de-multiplexed using bcl2, allowing for no mismatches in the index read. Reads were then processed first by TrimGalore (FelixKrueger, 2019) to remove

adaptors, then reads were collapsed using PEAR (Zhang et al., 2014) then aligned to the *yMSH2* using Bowtie2 (Langmead and Salzberg, 2012), a SAM file was generated using Samtools (Li et al., 2009) and then the makeup at each base pair was generated using Pysamstats (Miles, 2019). Data was manipulated in EXCEL, and then data points were graphed in R using ggplot (Wickham, 2009).

2.6 Acknowledgements

Aaron W. Miller recognized that these experiments shouldn't be done in the W303 background and did the initial plasmid collection. Jolie Carlisle mini-prepped all the plasmids out of *E.coli* and also did experiments showing the pooled assay would not work in the mini-stats. Josh Cuperus gave initial guidance on which tools to use to analyze the shotgun sequencing data.

Chapter 3: Determining pathogenicity of clinically relevant variants of Msh2

This chapter is part of a manuscript that has been submitted. It was primarily written by me, Anja Ollodart, with Brian Shirts writing the primary draft of the clinical results. Edits were made by co-authors on the manuscript Chiann-Ling Cindy Yeh, Adam Gordon, Aaron Miller, Brian Shirts, and the corresponding author Maitreya Dunham. Thus the tone of this chapter may differ from Chapter 1, 4, and 5 as it was written by many different voices.

3.1 Abstract

Loss-of-function (LOF) mutations in *MSH2* are associated with hereditary non-polyposis colorectal cancer (HNPCC), an inherited disorder that increases risk for many different cancers. However, the vast majority of *MSH2* variants found in human populations have insufficient evidence to be classified as either pathogenic or benign. We created an additional 185 human variants in the yeast ortholog, including both characterized and uncharacterized alleles curated from ClinVar and other clinical testing data. In a set of alleles of known pathogenicity, our assay recapitulated ClinVar's classification; we then estimated pathogenicity for 157 variants classified as uncertain or conflicting reports of significance. This method is capable of studying the mutation rate of many microbial species and can be applied to problems ranging from the generation of high-fidelity polymerases to measuring the rate of antibiotic resistance emergence.

3.2 Introduction

An example of an application of the multi-plexed mutation rate method is germline variants in mismatch repair (MMR) pathway genes which have strong implications for human health and are responsible for the cancer risk syndrome known as hereditary non-polyposis colorectal cancer (HNPCC) (Lynch et al., 2015; Peltomäki, 2016). In patients carrying pathogenic alleles, increased surveillance can detect cancers early, improving treatment outcomes (Gupta et al., 2019). There exist a large number of variants of MMR repair genes in humans, and many are classified as variants of uncertain significance (VUS) (Starita et al., 2017). Functional data found in model organisms can be used to assess potential pathogenicity of these variants (Gordon et al., 2019; Richards et al., 2015). In addition, there are many different systems to study the effects of mutations in the MMR pathway. One method is measuring sensitivity to 6-thioguanine or N-Methyl-N'-Nitro-N-Nitrosoguanidine, which correlates with functionality of subunits of the MMR complex including Msh2 and Mlh1 (Bouvet et al., 2019; Drost et al., 2013; Houlleberghs et al., 2017; Jia et al., 2020). Genome editing methods such as CRISPR-Cas9 provide a convenient way to introduce variants into human cells, where signatures of MMR deficiency can then be tracked (Rath et al., 2019). Another alternative is cell-free systems, which allow for using the human protein in an assay that checks for ability to repair DNA (Drost et al., 2019, 2012, 2010). While this provides an easy way to see all mutations caused by errors in replication, each variant must be expressed and purified separately, and this strategy is not amenable to being pooled. Also, while these systems work well for MMR complex proteins, they do not generalize

to all sources of mutation rate variability. Computational strategies theoretically could scale to all possible sites in all proteins of interest, and have been demonstrated to be good predictors of destabilizing variants of two MMR proteins tested in human colorectal cell lines (Abildgaard et al., 2019; Nielsen et al., 2017; Stein et al., 2019). However, they still require further validation. As such, a way to screen variants to determine if they cause a heightened mutation rate, and so therefore may be pathogenic, is needed. To address some of these problems, we wanted to generate a new experimental protocol to do multiplexed, direct assessment of mutation rate that was amenable to any molecular pathways for which mutations can be a read out in an easy to culture and genetically tractable organism, *Saccharomyces cerevisiae*. Once we established a proof of principle with previously published alleles, as outlined in chapter 2, we then assayed an additional 185 *MSH2* missense variants curated from ClinVar, a public repository of clinical variant interpretations derived from diagnostic genetic testing. To do so, we recreated these variants in the homologous sites of yeast *MSH2,* barcoded them along with control WT clones, and measured their mutation rates in a pooled format. Of the 28 variants of known pathogenicity, 100% recapitulated the functional consequence implied by previous clinical interpretation. We then examined 157 VUS from ClinVar and identified 50 variants with significantly different mutation rates from WT as measured by our assay. In addition to ClinVar classifications, data were also compared to tumor sequencing in cancer patients (Shirts et al., 2018); of the 25 VUS for which tumor data was available, 64% had clinical findings that were consistent with our functional data. These data taken together show that our new multiplexed mutation rate assay is an

accurate and scalable assay to study the mutation rate of many strains in a pooled format.

3.3 Results

3.3.1 – Mapping clinically relevant Msh2 variants from ClinVar to the yeast orthologue

To determine the limit on the number of alleles that can be assayed at once, additional alleles of Msh2 were curated from the clinical database ClinVar. These human Msh2 variants were mapped onto the yeast *MSH2* gene; only sites with a residue conserved between both orthologs were considered. 28 alleles with known pathogenicity were included as well as 216 variants of uncertain significance. Alleles in the Walker A ATPase domain and the linker region, which are more highly conserved between humans and yeast, were given precedence (Gammie et al., 2007). Alleles were barcoded a median of 5 times, however variation in barcode number is high (Fig 3.3.1.1) and the barcode and variant were associated with long read Pacific Biosciences sequencing. Of the 244 alleles synthesized, 191 variants covered by 1261 barcodes were able to be assayed. In addition, 737 barcodes were associated with WT, giving a robust internal control. The 53 variants not assayed were due to low barcode coverage of the variant in the pool or other factors that caused them to fail our quality filters (Figs 3.3.1.2, 3.3.1.3). While the number of variants that can be assayed is dependent on the composition of the pool, with higher mutators being easier to assay than lower mutators, results from this pool of Msh2 variants indicates that the assay is capable of tracking approximately 2000 barcodes at once.

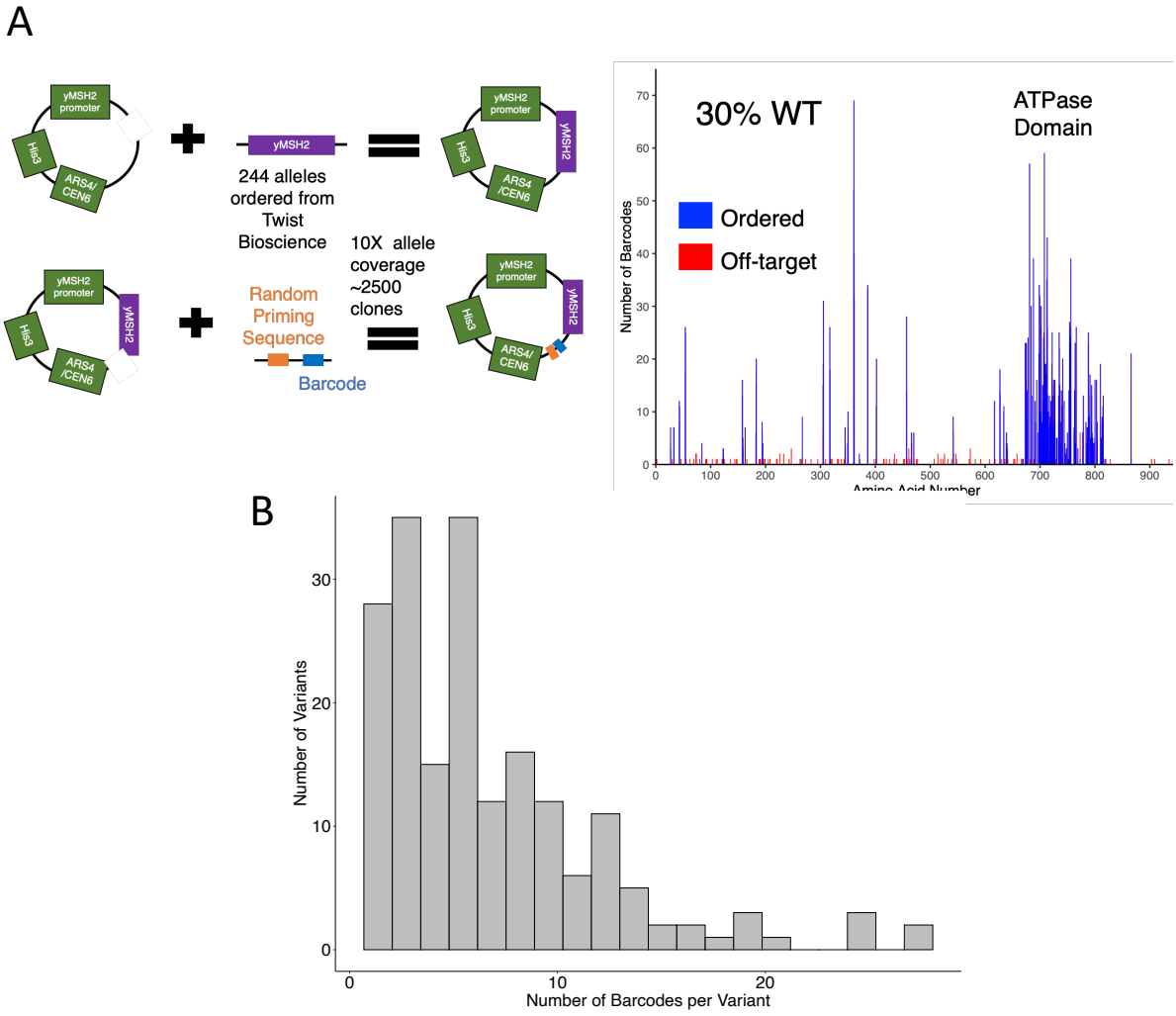
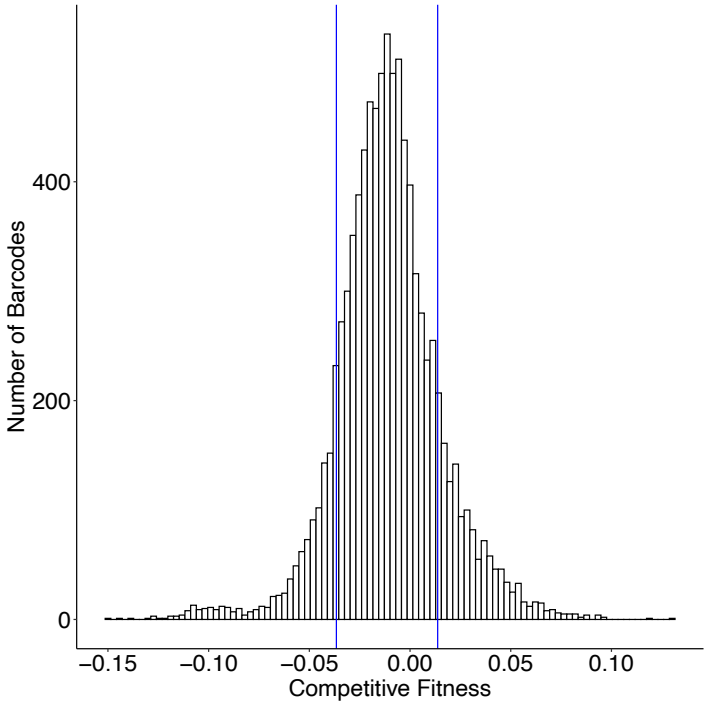


Figure 3.3.1.1 Generating barcoded variants of y*MSH2.* A. A schematic outlining the generation of new alleles, and a histogram representing the ordered vs off target variants. B. A histogram of the number of barcodes per variant.

A.



B.

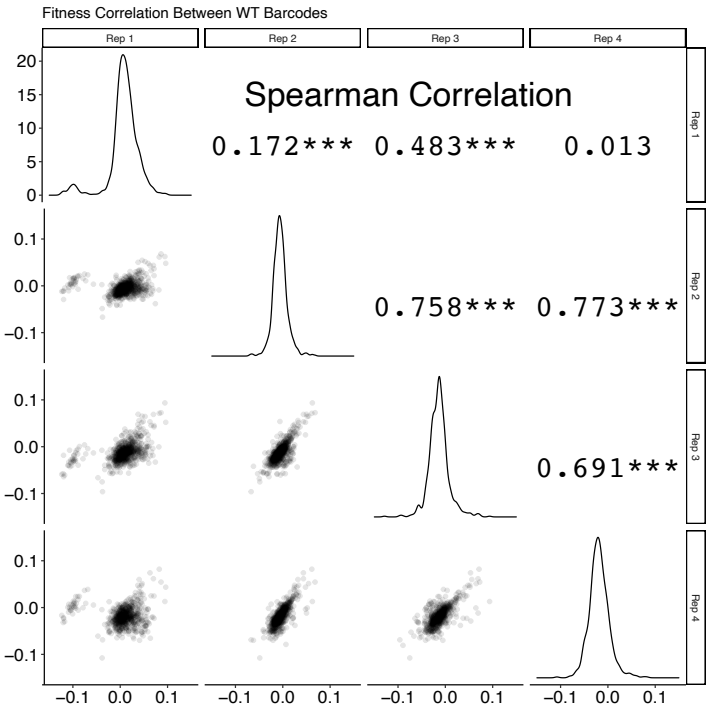
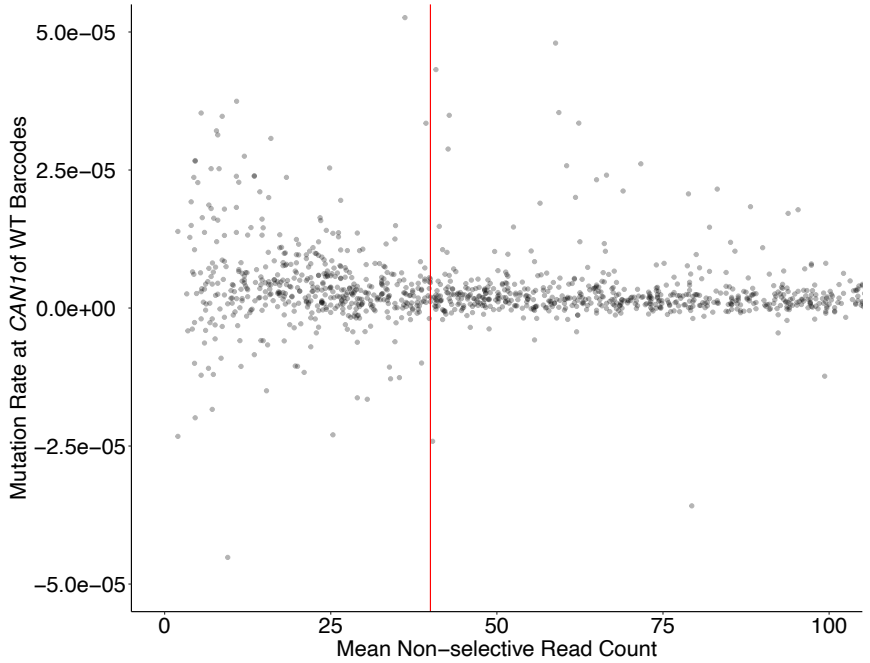


Figure 3.3.1.2 Determining filters for fitness effects. A. Blue lines indicate 1 standard deviation from the median of all barcodes. Barcodes outside 1 std deviation were not used in calculating mutation rates. B. Spearman correlation of fitness values of WT barcodes between biological replicates.

A.



B.

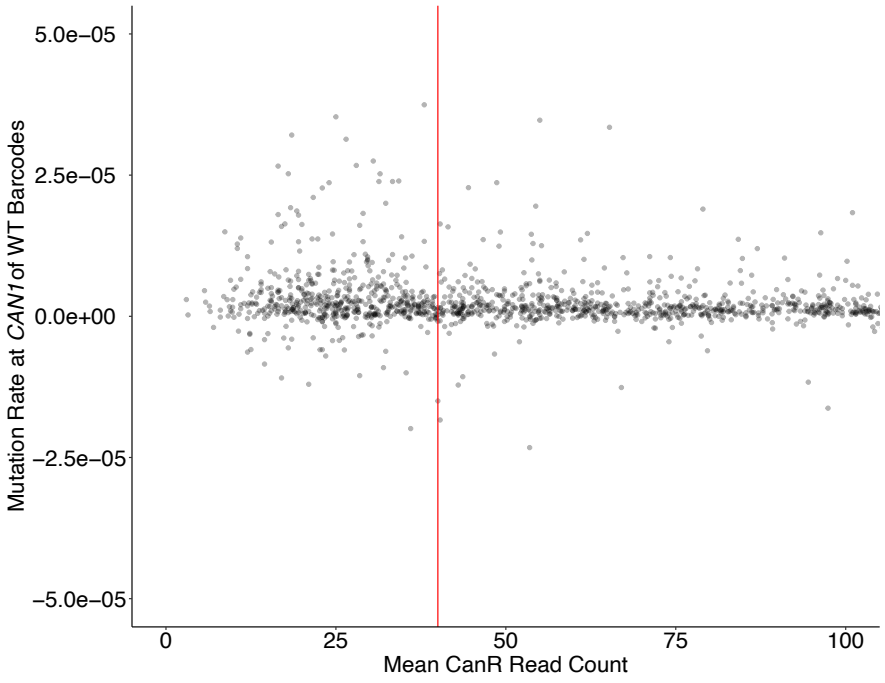


Figure 3.3.1.3 Scatter plot of WT barcode read count by mutation rate. Red line indicates cut off for minimum read number to be included in mutation rate assessment. Cut off was determined by highest reduction in standard deviation in readings.

3.3.2 – WT barcodes in deletion background show ~10 fold increase in mutation rate

The addition of barcodes, while requiring additional work and cost, can provide internal controls to the pooled mutation rate assay, as each genotype can be tracked in multiple independent lineages. The cumulative mutation rate of the barcoded pool containing novel variants of Msh2 was 3.96 X 10^{-6} Can^R/generation, a 26-fold increase over WT (Table 2.3.3). The WT barcodes in this assay showed a median mutation rate between 1.09 x10^{-6} – 3.59 x 10^{-6} – approximately 10-fold higher than the expected rate (Fig 3.3.6). This increase may be a result in the way that the pool is sequenced, as plasmids which cannot be sequenced would contribute to the number of colonies but not the number of reads. Colony PCR of 100 random colonies revealed no unbarcoded plasmid, but it might be a defect elsewhere. While the source of this global increase in mutation rate is unknown, it can be controlled for by the internal WT barcodes. All

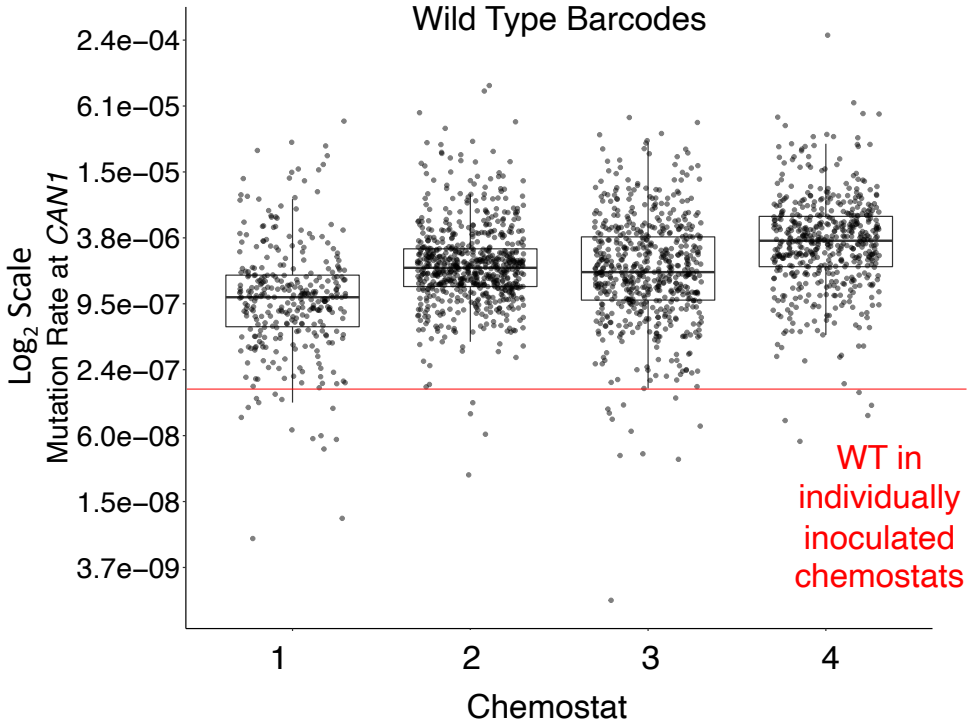


Figure 3.3.2 WT barcodes exhibit higher mutation rate than expected Calculated mutation rates (CanR per generation) of WT barcodes in four replicate experiments plotted on a Log_2 axis.

3.3.3 – Functional screen recapitulate results of alleles of known pathogenicity

To determine if the variants in the barcoded pool recapitulated clinical variant interpretations reported from diagnostic testing, we first looked at variants with known pathogenicity scores in ClinVar. The control variants used have a review status of 2 or higher indicating strong evidence for the level of pathogencity (Fig 3.3.3). Variants that shared the same ClinVar classification were found to have similar mutation rates. All 5 benign variants showed no significant difference from WT. 23 out of 23 pathogenic and likely pathogenic alleles showed a significant increase in mutation rate compared to WT. Due to the limited number of previously characterized variants in the pool, it's difficult to determine true sensitivity and specificity scores; however, these data lend confidence to our ability to bin variants into pathogenic or benign categories. Based on the results of known variants, we have created four bins: variants which do not differ significantly from WT are potentially benign (1). Those with values ranging between 1.3X -1.4X over WT are likely intermediate (2). Those which are significantly higher than WT are potentially pathogenic (3). The lowest fold change which showed a significant difference from WT was 1.4X, so we classified those which are above this threshold but do not reach significance as possibly pathogenic (4). We were heartened to see our assay recapitulates previous clinical interpretations, and that the use of control alleles allowed for the generation of bins to provide information on the VUS assayed.

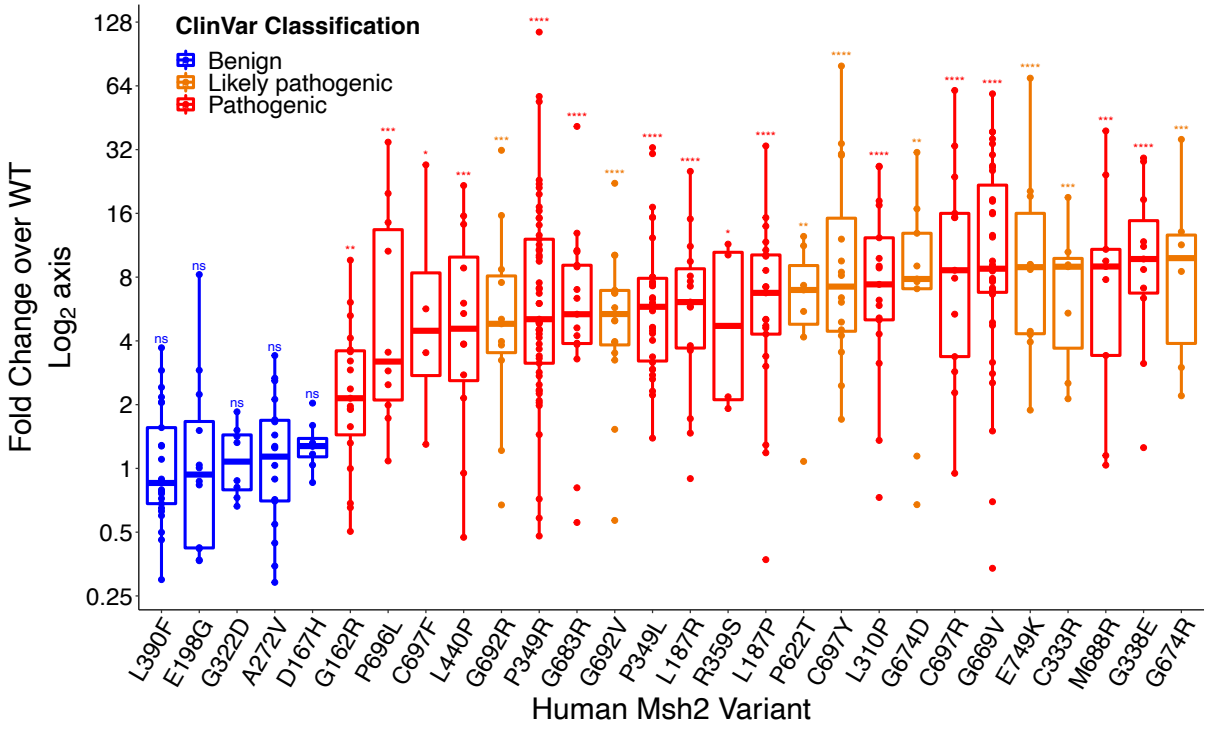


Figure 3.3.3 Fold change calculations of variants with known pathogenicity
Fold change over WT plotted on a Log_2 axis, colored by pathogenicity classification in ClinVar. Points represents individual barcoded measurements from the four replicate experiments. Significance determined by comparing variants to the WT barcodes by a Wilcoxan Rank-Sum test with the Benjamini-Hochberg correction for multiple hypothesis testing. *<0.05 ** <0.01 ***<0.001 ****<0.0001

3.3.4 – Estimating pathogenicity of variants of uncertain significance

We were able to assay 157 SNP variants of uncertain significance in this assay. 51/157 showed a significant difference in mutation rate in comparison to WT, and were classified as potentially pathogenic. The mutation rates of these 50 variants are in Table 3.3.4.1. A summary of the fold changes of all variants is found in Table 3.3.4.2. A comparison of fold change values across biological replicates is in Figure 3.3.4. We found 1 variant, K449N, had a significantly lower mutation rate than WT. The mechanism of this lowered mutation rate and whether it is a biologically relevant is unknown and could provide an interesting point of study if confirmed. The 50 VUS which showed a significant increase in mutation rate ranged from 1.39 fold over WT to 13 fold over WT. No alleles assayed had a full loss-of-function phenotype–which would be characterized by a 30-fold increase in mutation rate. This may reflect the dynamic range of the assay, or it may reflect that no true loss-of-function alleles existed in the data set. For future experiments, it may be wise to include barcoded deletion strains at a low frequency. Taken in total, this data set provides evidence of pathogenicity for an additional 157 VUS of *MSH2*. 83 will be classified as potentially benign, 7 intermediate, 17 possibly pathogenic, and 50 potentially pathogenic. While additional study would be required before these classifications could inform clinical diagnosis, these data represent a first indication of the effects of these mutations on function, and could be used as a line of evidence according to ACMG criteria (Richards et al., 2015).

Table 3.3.4.1

Mutation rates of significantly different alleles

Human Genotype	Yeast Genotype	Count[a]	Fold Induction CAN[b]	IQR[c]	Sig[d]
K449N	K466N	10	0.56	0.28	***
G761R	G780R	31	1.39	0.96	***
M672R	M691R	32	1.61	3.37	****
R711Q	R730Q	25	1.62	2.63	****
P349A	P361A	61	1.79	2.03	*
V695L	V714L	12	2.08	1.77	****
Q681E	Q700E	20	2.09	2.69	****
R359K	R371K	7	2.25	1.26	****
M672V	M691V	37	2.27	2.37	**
H783Y	H802Y	11	2.38	1.82	****
S676P	S695P	8	2.48	4.55	*
A609P	A627P	12	2.56	3.36	**
G761V	G780V	19	2.71	2.82	****
A700E	A719E	9	2.78	3.79	***
H783D	H802D	7	2.79	4.53	*
G683E	G702E	54	2.88	3.62	****
A689V	A708V	11	3.17	2.58	****
P696S	P715S	5	3.27	1.66	**
R524H	R542H	9	3.49	5.85	**
R524C	R542C	13	3.61	7.10	****
K675E	K694E	8	3.66	7.49	**
T724M	T743M	34	3.83	3.51	*
P622Q	P640Q	5	3.95	2.00	***
P622A	P640A	9	4.12	3.17	**
S676L	S695L	5	4.18	1.43	***
R524L	R542L	12	4.18	2.22	****
Y43D	Y43D	7	4.31	2.16	**
E643K	E662K	24	4.41	5.92	**
G669R	G688R	13	4.90	2.62	****
C693R	C712R	13	5.00	10.95	**
T668P	T687P	17	5.03	11.27	**
G692E	G711E	14	5.04	8.82	****
T724R	T743R	18	5.76	6.85	**
G683W	G702W	18	5.81	6.00	**
G827R	G855R	34	5.99	8.72	**

Table 3.3.4.1

(Continued)

Human Genotype	Yeast Genotype	Count[a]	Fold Induction CAN[b]	IQR	Sig[d]
Q690E	Q709E	15	6.03	8.01	*
A689P	A708P	9	6.27	5.14	***
P670R	P689R	6	6.31	6.23	****
G692W	G711W	23	6.33	6.60	*
G669D	G688D	18	6.63	11.83	*
G827E	G855E	11	6.95	7.11	****
T677R	T696R	13	7.41	4.94	****
P349S	P361S	47	7.48	8.01	****
G338V	G350V	14	7.98	15.14	****
G669C	G688C	12	8.08	5.35	**
R621Q	R639Q	4	8.29	2.07	***
N671D	N690D	13	8.61	6.56	*
R621P	R639P	13	9.28	10.78	****
L310R	L305R	14	9.50	13.27	****
C693Y	C712Y	4	9.70	3.71	****
R680P	R699P	9	13.00	17.19	****

a. Count, the number of times a variant was assayed in total
b. For stated genotype, mutation rate (mutations per cell division) was calculated and then compared to the WT mutation rate. Barcode and chemostat replicates are combined to calculate median foldchange.
c. Interquartile range of all barcode and chemostat replicates
d. Significance is calculated by a Wilcoxan Rank-sum test with the Benjamini-Hochberg correction for multiple hypothesis testing. *<.05 ** <.01 ***<.001 ****<.0001

Table 3.3.4.2
Mutation rates of all assayed alleles

Human Genotype[a]	AlleleID[b]	count[c]	Fold Induction CAN[d]	IQR[e]	Score[f]	95% CI[g]	Sig[h]	ClinVar[i]	Star Rating[j]
L390F	50080	25	0.854	0.876	-0.228	0.353	ns	B	3
E198G	96620	12	0.938	1.270	-0.093	0.776	ns	B	3
G322D	16801	8	1.100	0.647	0.137	0.385	ns	B	3
A272V	50090	18	1.142	0.985	0.191	0.478	ns	B	3
D167H	96587	8	1.275	0.254	0.351	0.261	ns	B	3
I774V	392910	10	0.888	0.543	-0.172	0.569	ns	CIP	1
H785P	181985	19	1.045	0.774	0.063	0.500	ns	CIP	1
Q374R	232572	47	1.146	1.295	0.196	0.417	ns	CIP	1
I770V	96430	19	1.338	1.219	0.420	0.524	ns	CIP	1
P349A	95987	61	1.788	2.031	0.838	0.295	***	CIP	1
G692E	259741	14	5.036	8.819	2.332	0.516	****	CIP	1
L687P	96348	4	6.132	14.400	2.616	2.820	ns	CIP	1
G692W	419395	23	6.330	6.598	2.662	0.613	****	CIP	1
G669C	473032	12	8.079	5.352	3.014	0.514	****	CIP	1
L310R	96721	14	9.505	13.271	3.249	0.743	****	CIP	1
N671K	96333	4	2.501	1.398	1.322	0.571	ns	LP	1
P349T	405779	30	4.712	7.787	2.236	0.544	****	LP	1
G692R	96353	11	4.816	4.595	2.268	0.915	***	LP	3
F694S	419396	6	4.976	4.312	2.315	1.371	***	LP	1
G692V	96355	16	5.369	3.090	2.425	0.582	****	LP	3
P622T	96280	7	6.958	4.446	2.799	0.877	**	LP	2
C697Y	181980	16	7.272	12.142	2.862	0.743	****	LP	3
G674D	96336	9	7.840	5.840	2.971	1.161	**	LP	3
E749K	96417	10	8.920	12.420	3.157	0.923	****	LP	3
C333R	96747	7	8.960	5.862	3.163	0.848	***	LP	2
G674R	96335	6	9.939	8.287	3.313	1.182	***	LP	3
M453K	96113	5	1.984	1.502	0.988	0.638	ns	P	3
G162R	96580	19	2.145	2.146	1.101	0.510	**	P	3
P696L	96356	10	3.211	11.408	1.683	1.057	***	P	3
C697F	96358	4	4.593	8.064	2.199	1.790	*	P	3
L440P	96100	12	4.628	7.568	2.210	0.928	***	P	3
P349R	95988	53	5.064	8.946	2.340	0.422	****	P	3
G683R	96343	17	5.346	5.219	2.418	0.682	****	P	3
P349L	95989	30	5.801	4.685	2.536	0.393	****	P	3
L187R	96610	15	6.099	5.088	2.609	0.653	****	P	3
R359S	96014	4	6.177	8.381	2.627	1.362	*	P	3

Human Genotype[a]	AlleleID[b]	count[c]	Fold Induction CAN[d]	IQR[e]	Score[f]	95% CI[g]	Sig[h]	ClinVar[i]	Star Rating[j]
L310P	96720	17	7.395	7.244	2.887	0.660	****	P	3
C697R	96357	13	8.622	12.631	3.108	0.925	****	P	3
G669V	96329	35	8.767	15.242	3.132	0.544	****	P	3
M688R	96349	9	8.982	7.407	3.167	1.168	***	P	3
K449N	150578	10	0.557	0.283	-0.844	0.440	**	VUS	2
I192M	561579	4	0.605	0.950	-0.724	1.482	ns	VUS	1
A739P	483395	4	0.646	0.410	-0.630	0.924	ns	VUS	1
I770T	221245	9	0.688	0.770	-0.539	0.626	ns	VUS	2
K773N	518360	21	0.716	0.590	-0.481	0.456	ns	VUS	2
D758N	393162	27	0.769	0.887	-0.379	0.437	ns	VUS	1
A781T	518449	6	0.774	3.324	-0.369	1.295	ns	VUS	2
D706E	232623	20	0.805	0.665	-0.313	0.353	ns	VUS	1
V705M	558518	33	0.809	1.323	-0.306	0.379	ns	VUS	1
I667V	232613	8	0.811	0.390	-0.303	0.941	ns	VUS	1
G761A	238858	8	0.813	0.483	-0.299	0.571	ns	VUS	1
T754S	232629	7	0.824	0.303	-0.279	0.594	ns	VUS	2
M726V	451363	36	0.834	0.812	-0.262	0.324	ns	VUS	2
I648N	232607	43	0.865	0.954	-0.209	0.319	ns	VUS	2
N671S	575673	6	0.870	1.402	-0.201	1.024	ns	VUS	2
K661R	558514	11	0.883	1.499	-0.180	0.634	ns	VUS	1
P27T	238804	6	0.886	0.652	-0.175	0.504	ns	VUS	1
A54S	181900	16	0.925	0.400	-0.113	0.342	ns	VUS	2
Q645E	96297	27	0.926	0.825	-0.111	0.373	ns	VUS	2
I774S	238859	5	0.926	0.373	-0.111	0.460	ns	VUS	1
V705L	451166	19	0.935	0.951	-0.097	0.433	ns	VUS	1
G322S	181927	29	0.952	0.720	-0.072	0.316	ns	VUS	2
I735V	171077	8	0.952	0.365	-0.071	0.540	ns	VUS	2
G669A	96328	32	0.956	0.718	-0.065	0.344	ns	VUS	NA
G162A	96581	30	0.965	0.971	-0.051	0.336	ns	VUS	2
F664L	472943	16	0.967	0.682	-0.048	0.500	ns	VUS	1
I766V	181984	4	0.968	0.604	-0.047	0.766	ns	VUS	2
L440F	472876	28	0.968	0.852	-0.047	0.335	ns	VUS	2
D646G	427155	17	0.971	0.647	-0.042	0.447	ns	VUS	2
N127T	558488	4	0.984	0.527	-0.023	0.787	ns	VUS	1
V684L	392793	8	0.987	0.793	-0.019	0.557	ns	VUS	1
I704V	180022	19	0.993	0.749	-0.010	0.334	ns	VUS	2
A763V	180027	7	1.001	0.239	0.002	0.301	ns	VUS	2
A714V	96382	12	1.009	0.908	0.013	0.789	ns	VUS	2

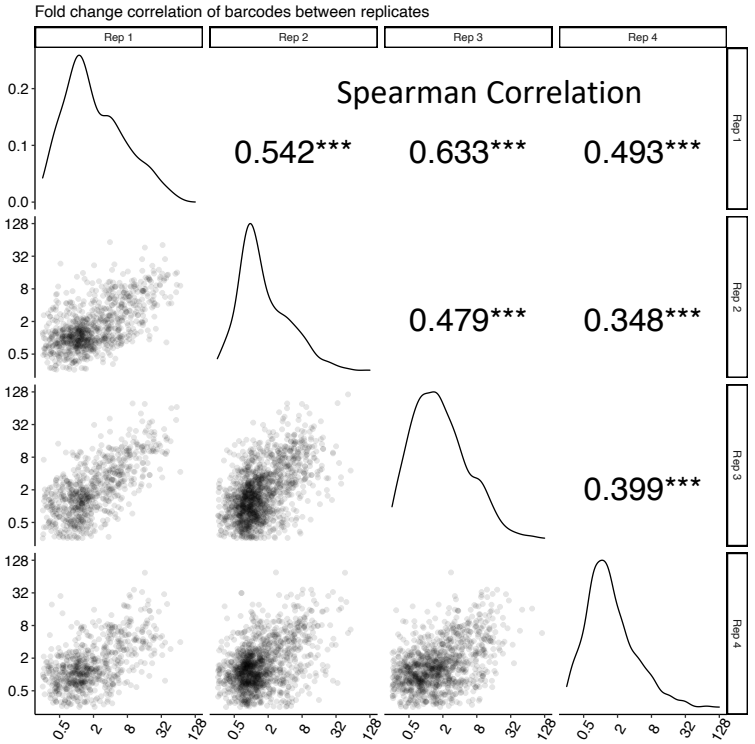
Human Genotype[a]	AlleleID[b]	count[c]	Fold Induction CAN[d]	IQR[e]	Score[f]	95% CI[g]	Sig[h]	ClinVar[i]	Star Rating[j]
L390V	482375	29	1.024	1.685	0.034	0.525	ns	VUS	1
D654E	232610	53	1.037	1.101	0.052	0.293	ns	VUS	1
Y678C	482393	64	1.040	1.023	0.057	0.271	ns	VUS	2
I708T	473041	4	1.048	1.801	0.068	1.478	ns	VUS	2
L736F	232628	52	1.049	1.063	0.069	0.355	ns	VUS	1
W764G	244394	20	1.063	0.638	0.088	0.331	ns	VUS	2
Q374H	133083	31	1.064	0.855	0.090	0.288	ns	VUS	2
A733T	212200	8	1.076	1.098	0.106	0.930	ns	VUS	2
A609S	232602	23	1.076	1.754	0.106	0.578	ns	VUS	2
T732I	483367	13	1.080	0.602	0.111	0.600	ns	VUS	1
E701A	232621	10	1.081	1.006	0.113	0.971	ns	VUS	2
M726I	152055	29	1.091	1.186	0.126	0.429	ns	VUS	2
R680G	181976	19	1.093	1.919	0.128	0.585	ns	VUS	2
Y678H	232617	25	1.094	1.036	0.130	0.388	ns	VUS	2
E768V	451108	6	1.095	1.136	0.131	1.287	ns	VUS	2
E658G	180019	71	1.097	1.108	0.133	0.275	ns	VUS	2
A54V	472743	20	1.098	0.578	0.135	0.372	ns	VUS	1
I708V	392902	9	1.117	2.411	0.159	0.971	ns	VUS	2
P616R	133093	11	1.125	0.559	0.171	0.524	ns	VUS	2
T724A	244392	35	1.131	1.061	0.178	0.335	ns	VUS	2
G683V	212198	68	1.137	1.006	0.186	0.284	ns	VUS	2
L310V	232559	47	1.140	2.005	0.189	0.346	ns	VUS	1
V655I	133094	18	1.153	0.909	0.205	0.380	ns	VUS	3
Y43F	405752	10	1.158	1.000	0.212	0.709	ns	VUS	2
T732S	180025	8	1.173	0.519	0.230	0.751	ns	VUS	1
I766M	405812	10	1.175	1.465	0.233	0.801	ns	VUS	2
H785Y	451109	20	1.180	0.978	0.239	0.393	ns	VUS	1
L187F	472789	14	1.186	1.311	0.246	0.518	ns	VUS	2
M726L	558522	36	1.190	1.127	0.251	0.389	ns	VUS	1
S734C	451371	33	1.196	0.743	0.258	0.403	ns	VUS	2
I685V	389527	6	1.203	1.278	0.266	0.876	ns	VUS	2
W764R	518334	13	1.204	1.712	0.268	0.685	ns	VUS	1
D758Y	232630	40	1.209	0.898	0.274	0.329	ns	VUS	1
A54T	232513	24	1.224	0.843	0.292	0.430	ns	VUS	2
P670L	96330	8	1.225	0.891	0.293	0.649	ns	VUS	2
I774T	393164	7	1.258	0.894	0.331	0.405	ns	VUS	2
L599S	226299	16	1.261	0.932	0.335	0.387	ns	VUS	2
W764C	96425	6	1.264	0.984	0.338	0.735	ns	VUS	2

Human Genotype[a]	AlleleID[b]	count[c]	Fold Induction CAN[d]	IQR[e]	Score[f]	95% CI[g]	Sig[h]	ClinVar[i]	Star Rating[j]
G322V	152230	24	1.298	0.528	0.377	0.506	ns	VUS	2
Q681H	180021	5	1.304	0.704	0.383	0.689	ns	VUS	2
Q645H	139567	29	1.304	1.293	0.383	0.398	ns	VUS	2
T754A	212204	8	1.309	1.187	0.388	1.009	ns	VUS	2
I704T	232622	11	1.317	0.853	0.397	0.578	ns	VUS	2
D646A	238851	22	1.360	1.362	0.444	0.551	ns	VUS	2
I685M	518347	8	1.378	1.259	0.462	0.798	ns	VUS	2
S723C	472863	16	1.382	1.512	0.467	0.511	ns	VUS	1
G761R	392796	31	1.386	0.965	0.471	0.380	*	VUS	1
V163A	232532	7	1.388	1.437	0.473	0.934	ns	VUS	1
L709V	392905	16	1.465	1.435	0.551	0.552	ns	VUS	1
T677A	482416	22	1.503	1.943	0.588	0.566	ns	VUS	2
R680Q	561662	27	1.509	1.713	0.594	0.453	ns	VUS	2
P670H	244390	10	1.561	1.526	0.642	0.801	ns	VUS	2
Y43H	181897	15	1.576	1.153	0.657	0.442	ns	VUS	1
M672R	181973	32	1.612	3.375	0.689	0.520	*	VUS	2
R711Q	472967	25	1.619	2.625	0.695	0.552	*	VUS	2
L187I	472788	16	1.632	3.751	0.706	0.844	ns	VUS	1
H785R	185986	15	1.679	1.110	0.747	0.506	ns	VUS	2
K720E	232625	6	1.685	0.774	0.753	0.704	ns	VUS	2
P27R	558125	7	1.863	1.284	0.898	0.533	ns	VUS	1
A765V	472865	4	1.915	0.938	0.938	0.578	ns	VUS	2
V695L	551836	12	2.080	1.769	1.056	0.571	*	VUS	2
Q681E	180020	20	2.087	2.686	1.061	0.483	**	VUS	2
P670S	561652	8	2.178	2.964	1.123	0.926	ns	VUS	2
R359K	472774	7	2.251	1.255	1.170	0.407	*	VUS	2
M672V	432518	37	2.270	2.372	1.182	0.412	****	VUS	1
P616S	152379	6	2.314	1.657	1.210	0.997	ns	VUS	2
H783Y	518363	11	2.382	1.823	1.252	1.207	**	VUS	1
S676P	96339	8	2.480	4.549	1.310	0.839	*	VUS	NA
T33P	96743	5	2.524	2.497	1.336	0.807	ns	VUS	2
A609P	427153	12	2.561	3.362	1.357	0.725	**	VUS	1
G761V	232631	19	2.706	2.820	1.436	0.561	**	VUS	2
A700E	232619	9	2.780	3.788	1.475	1.162	**	VUS	1
H783D	451390	7	2.787	4.526	1.479	0.806	**	VUS	1
G683E	482418	54	2.879	3.619	1.526	0.353	****	VUS	2
A689V	393145	11	3.169	2.578	1.664	0.576	**	VUS	1
P696S	472955	5	3.271	1.662	1.710	0.981	*	VUS	2

Human Genotype[a]	AlleleID[b]	count[c]	Fold Induction CAN[d]	IQR[e]	Score[f]	95% CI[g]	Sig[h]	ClinVar[i]	Star Rating[j]
R524C	405792	13	3.607	7.102	1.851	0.895	***	VUS	2
K675E	393142	8	3.659	7.490	1.872	1.039	***	VUS	1
T724M	180024	34	3.831	3.515	1.938	0.420	****	VUS	2
P622Q	416960	5	3.953	2.002	1.983	0.526	**	VUS	1
P622A	181966	9	4.122	3.168	2.043	0.608	***	VUS	1
S676L	427156	5	4.183	1.428	2.065	0.703	**	VUS	1
R524L	96176	12	4.184	2.216	2.065	0.853	***	VUS	2
Y43D	551834	7	4.312	2.160	2.108	0.474	***	VUS	2
E643K	171076	24	4.406	5.920	2.139	0.455	****	VUS	2
N671Y	96332	6	4.584	11.528	2.197	1.803	ns	VUS	NA
G669R	96319	13	4.896	2.617	2.291	0.571	****	VUS	3
C693R	561664	13	4.997	10.955	2.321	0.798	****	VUS	2
T668P	405811	17	5.027	11.271	2.330	0.853	****	VUS	1
Q690E	96351	15	6.026	8.006	2.591	0.639	****	VUS	NA
A689P	419393	9	6.274	5.139	2.649	0.896	**	VUS	1
P670R	451379	6	6.308	6.228	2.657	0.803	**	VUS	2
G669D	96327	18	6.626	11.831	2.728	0.675	****	VUS	1
T677R	232616	13	7.405	4.942	2.889	0.490	****	VUS	2
P349S	482373	47	7.479	8.013	2.903	0.427	****	VUS	2
G338V	560675	14	7.977	15.141	2.996	0.760	****	VUS	1
R621Q	451338	4	8.287	2.074	3.051	0.465	**	VUS	2
N671D	232615	13	8.612	6.563	3.106	0.529	****	VUS	1
R621P	473027	13	9.280	10.779	3.214	0.736	****	VUS	2
R680P	472861	9	13.003	17.187	3.701	1.153	***	VUS	1
G827R	NA	34	5.992	8.716	2.583	0.506	****	VUS	NA
G827E	NA	11	6.945	7.109	2.796	0.631	****	VUS	NA
C693Y	380277	4	9.701	3.708	3.278	0.895	**	VUS	1

a. The yeast genotype mapped back to the human variant
b. Allele ID for Clinvar
c. Count, the number of times a variant was assayed in total
d. For stated genotype, variant mutation rate (mutations per cell division) was compared WT. Barcode and chemostat replicates were combined to calculate the median.
e. Interquartile range of all barcode and chemostat replicates
f. For stated genotype, variant mutation rate (mutations per cell division) was compared WT. Barcode and chemostat replicates were combined to calculate the median and the score was Log_2 transformed.
g. 95% confidence interval calculated on log_2 transformed fold change values
h. Significance is calculated by a Wilcoxian Ranksum test with the Benjamini-Hochberg correction for multiple hypothesis testing.
*< 0.05 ** < 0.01 ***< 0.001 ****< 0.0001
i. Initial clinical classification as stated in ClinVar. B = Benign, CIP = Conflicting Interpretations of pathogenicity, LP = Likely Pathogenic, P = Pathogenic, VUS = Variants of uncertain significance
j. ClinVar star rating 1= Single submitter - criteria provided, 2 = criteria provided, multiple submitters, no conflicts, 3 = reviewed by expert panel, NA = Variant not in Clinvar

A.



B.

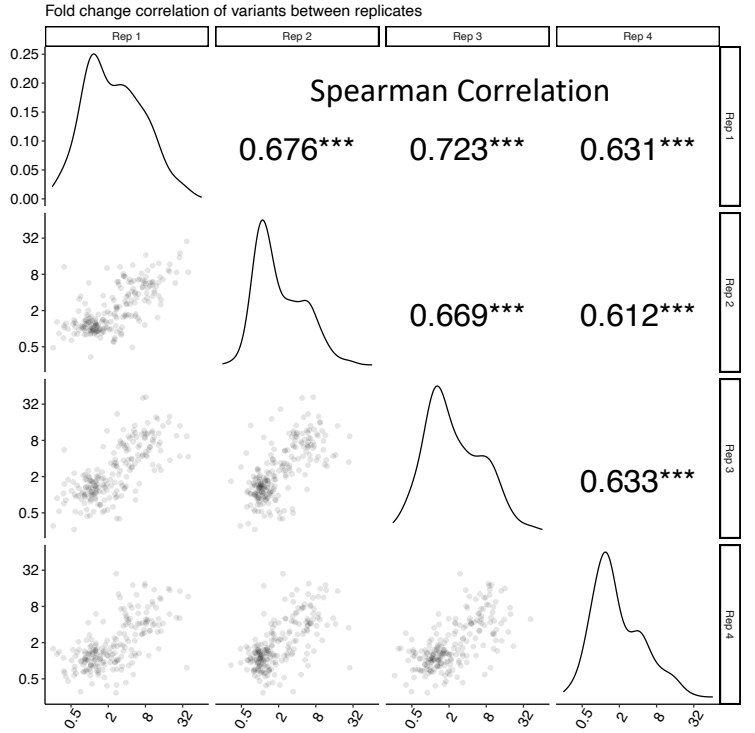


Figure 3.3.4 Comparing results across biological replicates A. Comparison of fold change calculations by barcode across all 4 replicates. B. Comparison of fold change across 4 replicates by variant, correlation is spearman correlation.

3.3.5 – Associating variant data with clinical and tumor sequencing phenotypes

In order to more accurately compare clinical data with the outputs of this screen, the fold change calculations were converted to scores. Table 3.3.5 contains information on variants which have clinical findings associated with them. Clinical summaries were gathered from data provided to the University of Washington Genetics and Solid Tumors Laboratory. An assessment of whether the clinical information is consistent or inconsistent with functional scores was provided by a board-certified molecular pathologist with expertise in this area (BHS). Clinical evidence was considered consistent with functional data when both suggested the variant was pathogenic or benign regardless of the strength or significance of the data. There are several types of information on *MSH2* that can be gathered from patients, families, and tumors (Thompson et al., 2013; Rañola et al., 2018; Shirts et al., 2018; Li et al., 2020). Personal and family history of colorectal or endometrial cancer provide weak evidence of pathogenicity while personal and family history lacking HNPCC associated cancers provide weak evidence against pathogenicity (Li et al., 2020).Tumor characteristics of microsatellite instability (MSI-H) and loss of Msh2 and Msh6 on immunohistochemistry staining provide moderate evidence supporting pathogenicity (Li et al., 2020; Thompson et al., 2013). Presence of alternative pathogenic variants in *MSH2* or other genes that explain these tumor or other tumor characteristics provides evidence against pathogenicity, while a second somatic pathogenic variant at heterozygous frequency or loss of heterozygosity in tumor provides moderate and strong evidence supporting pathogenicity, respectively (Shirts et al., 2018).

Table 3.3.5

Summary of variants with clinical or tumor data

Human Genotype	Score[a]	95% CI[b]	Sig[c]	ClinVar[d]	Clinical Information[e]	CST[f]
P27L	-1.06	0.87	ns	VUS	Heterozygous germline. MSI-H colon cancer with loss of PMS2 by IHC, under age 30 at diagnosis. Also has heterozygous VUS in *MLH1*.	Y
K449N	-0.84	0.44	**	VUS	Heterozygous germline. Ovarian cancer. No family history of cancer.	Y
A54S	-0.11	0.34	ns	VUS	Heterozygous germline. No personal history of cancer. Family history early onset colorectal cancer and breast cancer.	Y
A733T	0.11	0.93	ns	VUS	Heterozygous germline. Breast cancer. Family history of colon, ovarian, and brain cancers.	Y
P616R	0.17	0.52	ns	VUS	Heterozygous germline. MSI-H endometrial tumor with loss of MSH2 and MSH6 by IHC. Two other clearly pathogenic loss-of-function *MSH2* mutations identified in the tumor make this less likely to be pathogenic	Y
Q374R	0.20	0.42	ns	B	Heterozygous germline. MSI-H endometrial tumor with loss of MLH1-MSH2, MSH6, and PMS2 by IHC. The tumor had 2 other clearly pathogenic somatic mutations in *MSH2*.	Y
V655I	0.21	0.38	ns	VUS	Heterozygous germline. Colorectal cancer under age 30. No family history of cancer.	N
W764R	0.27	0.68	ns	VUS	Heterozygous germline. MSI-H colon tumor with loss of MSH2 and MSH6 by IHC. Evidence of LOH for *MSH2* variant in tumor.	N
L599S	0.33	0.39	ns	VUS	Heterozygous germline. Seen in patient with breast cancer and family history of breast and colorectal cancer.	Y
V78I	0.36	0.70	ns	VUS	Heterozygous germline. MSS ovarian tumor with other pathogenic variants with no LOH for *MSH2* variant in tumor.	Y
I770V	0.42	0.52	ns	VUS	Heterozygous germline. Colorectal cancer diagnosed under age 30. Seen with heterozygous germline VUS in *APC* gene.	-
G761R	0.47	0.38	*	VUS	Suspected germline variant in prostate tumor. IHC for MSH2, MSH6 equivocal. MSI equivocal. Apparent *MSH2* LOH in tumor.	-
H785R	0.75	0.51	ns	VUS	Heterozygous germline. Colorectal cancer diagnosed over age 70.	-
P349A	0.84	0.30	***	VUS	Homozygous germline. MSS colorectal cancer. The tumor also had *POLE* mutation and ultramutator phenotype.	-
Q681E	1.06	0.48	**	VUS	Heterozygous germline. MSI-H colorectal cancer diagnosed over age 50. Tumor with loss of MLH1 and PMS2 explained by double somatic *MLH1* mutation in tumor.	N
A609P	1.36	0.73	**	VUS	Heterozygous germline. MSI-H colon cancer with loss of MSH2 and MSH6 by IHC. Under age 50 at diagnosis. Tumor had one somatic pathogenic *MSH2* mutation seen at heterozygous frequency in the tumor.	Y
P696S	1.71	0.98	*	VUS	Heterozygous germline. Personal history of phenochromocytoma, family history of renal and brain cancer.	N
S676L	2.06	0.70	**	VUS	Heterozygous germline. MSI-H colon cancer diagnosed under age 50. Tumor had loss of MSH6 by IHC. Seen with *MSH2* p.G827R somatic mutation listed below.	Y

Table 3.3.5

(Continued)

Human Genotype	Score[a]	95% CI[b]	Sig[c]	ClinVar[d]	Clinical Information[e]	CST[f]
C693R	2.32	0.80	****	LP	Somatic mutation in tumor. MSI-H colon tumor with loss of MSH2 by IHC. This was seen with another heterozygous pathogenic mutation in *MSH2* (1760-1G>A).	Y
G692V	2.42	0.58	****	LP	Somatic mutation in tumor. MSI-H neuroendocrine tumor with loss of MSH2 and MSH6 by IHC. Seen in a tumor with a germline likely pathogenic variant in *MSH2* (p.L30R) that had loss of heterozygosity.	-
G827R	2.58	0.51	****	VUS	Somatic mutation in tumor. MSI-H colon cancer diagnosed under age 50. Tumor had loss of MSH6 by IHC. Seen with the germline variant p.S676L listed above.	Y
G692W	2.66	0.61	****	VUS	Heterozygous germline. MSI-H endometrial tumor with loss of MSH2 and MSH6 by IHC. Cancer diagnosed over age 50. Variant reported by another laboratory to segregate with HNPCC in one family. A variant at the same position (p.G692R, NM_000179.2:c.2074 G to C) is classified as likely pathogenic (class 4) by the InSiGHT consortium.	Y
G827E	2.80	0.63	****	VUS	Heterozygous germline. Pancreatic cancer diagnosed over age 80. Loss of *MSH2* and MSH6 by IHC.	Y
N671D	3.11	0.53	****	VUS	Heterozygous germline. Personal history of colon polyps and family history of colon, uterine, and other cancers.	Y
L310R	3.25	0.74	****	P	Heterozygous germline. Seen in a family with multiple MSI-H colon cancers that had loss of MSH2. Co-segregation likelihood ratio for pathogenicity in the family was 44:1. See (Tsai et al., 2019) for complete pedigree information.	Y

a. For stated genotype, variant mutation rate (mutations per generation) was compared to the WT mutation rate. The score represents Log_2(median fold change).
b. 95% confidence interval on Log_2(fold change).
c. Significance is calculated by a Wilcoxan Rank-Sum test with the Benjamini-Hochberg correction for multiple hypothesis testing. *< 0.05 ** < 0.01 ***< 0.001 ****< 0.0001
d. Initial clinical classification as stated in ClinVar.
e. Clinical information collected from UW Laboratory Medicine. MSI-H = Microsatellite instability- high IHC = Immunohistochemistry
f. CST = Consistent, tumor or clinical data is consistent with functional score. Y = Yes, N = No, - = undetermined

Formal strategies for combining each of these types of data with functional data are outside the scope of this effort. Rather we provide relevant clinical summaries with an overall assessment of whether the clinical information is consistent or inconsistent with functional scores. Of the 25 variants identified 16 (64%) had clinical data that was

consistent with functional data, 4 (16%) had clinical data that was inconsistent with functional data, and 5 (20%) had clinical data that was equivocal or fell in the indeterminate functional score range. Some discordance is expected given that functional data is only one component of the ACMG guidelines for clinical variant interpretation; this discordance is consistent with results of past studies seeking to use other clinical criteria to classify variants (Li et al., 2020; Shirts et al., 2018; Thompson et al., 2013).

3.3.6 – WT barcodes assayed in a WT background show heightened mutation rate

To determine whether the heightened mutation rate of the WT barcodes in the deletion background is a technical or real artifact, we repeated the experiment in the WT background. The mutation rate of WT barcodes in the WT BG was 6.70 - 7.28 x 10^{-7} CanR per generation (Fig 3.3.6). This is approximately 5 times over the WT value associated with individually inoculated chemostats, but 2 fold less than the WT barcodes in the deletion background. This is somewhat odd, since the cumulative mutation rate of the pool was only 1.3x over WT in the WT background (Table 2.3.3), indicating that the WT barcodes in the pool are on average higher than the median mutation rate of the pool. It would indicate that there is something both about the background and the pooled format that is leading to the heightened mutation rate of WT barcodes. We hypothesize that it may be variants which contribute to the mutation rate but aren't able to be sequenced, but we don't know. Thankfully, the presence of WT barcodes allows us to normalize for whatever reason exists for this error.

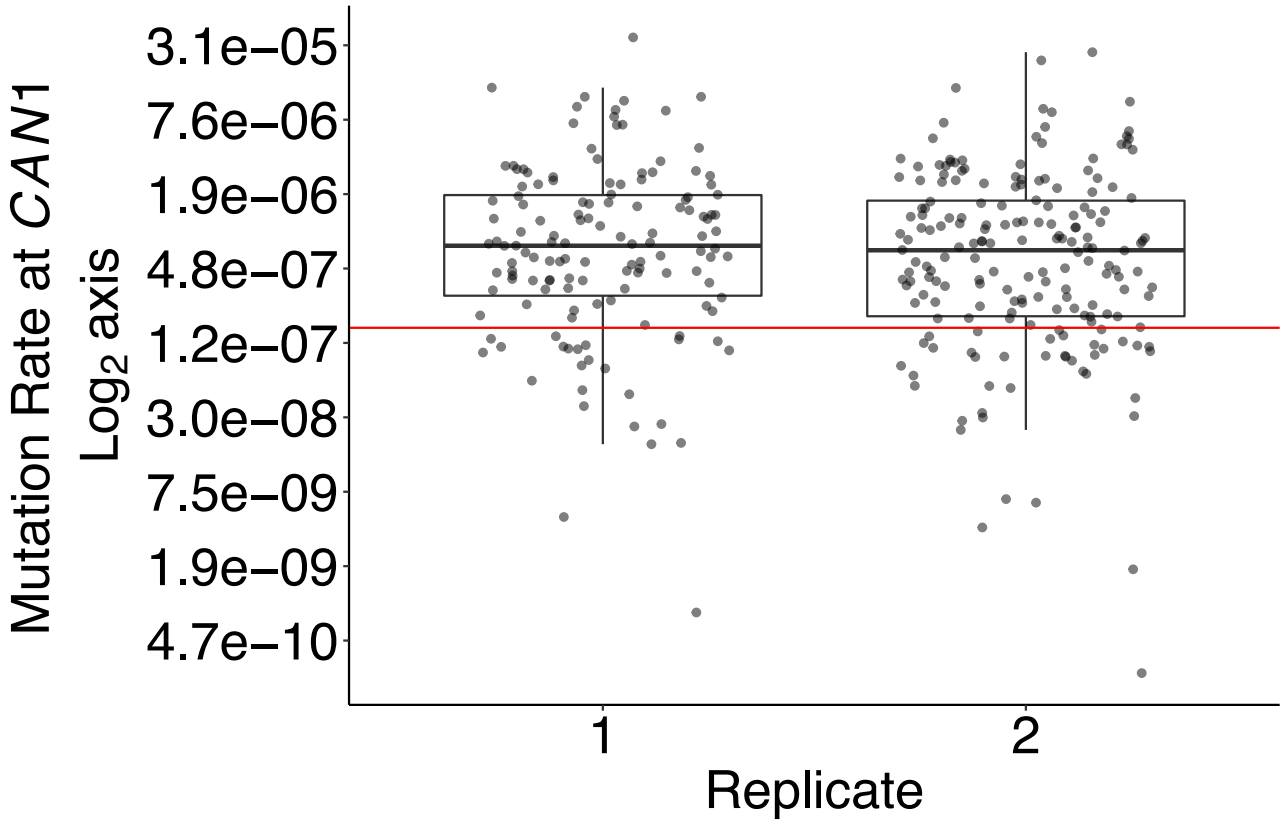


Figure 3.3.6 WT barcodes in WT BG exhibit higher mutation rate than expected
Calculated mutation rates (CanR per generation) of WT barcodes in two replicate experiments plotted on a Log_2 axis. Red line represents mutation rate of WT in individually inoculated chemostats.

3.3.7 – Data from variants in WT background is inconclusive

Fold change calculations for all variants which passed the same quality filters used for variants in the deletion background are represented in table 3.3.7. Significance for all values were calculated and no variants in the WT background showed a significant difference from WT. While there were several variants above the 1.4X threshold established in previous sections, I think the relatively few replicates and large spread of values prevented any of the variants to pass a significance threshold. The Can^R population in the WT background is much smaller, which makes it difficult to comprehensively assay all variants. In addition, there was only 1 allele of known dominance within the assayed alleles (Drotschmann et al., 1999b). Variant G688A has a calculated fold change increase of 2.92 and an inter quartile range of 6.55, but it was only assayed 4 times. While this known dominant allele did show an increased fold change, it's not possible with the current data to determine if this is a true increase. K694E, the variant with the highest fold change increase in the pooled assay, has not been tested explicitly, but it has been shown that K694A is a dominant mutation (Drotschmann et al., 1999b). This variant would be interesting to follow up on. In order to determine if there are dominant alleles of Msh2, I would use a much smaller library of variants and complete more replicates. The future smaller library of variants could focus on the variants which showed an increase in fold change in this assay. In addition, traditional assays could be performed to follow up on these results. While assaying these variants in the WT background works in theory, in practice the data is currently unreliable.

Table 3.3.7
Summary of fold change calculations of Msh2 variants in the WT background

Yeast Genotype[a]	Count[b]	Fold Induction CAN[c]	IQR[d]	Sig[e]
W783R	5	0.10	0.08	ns
G688V	6	0.23	1.71	ns
L728V	4	0.25	0.35	ns
G711E	4	0.25	1.28	ns
V724M	4	0.27	0.40	ns
T743I	8	0.29	0.09	ns
Q709E	6	0.29	0.56	ns
M745V	7	0.31	0.92	ns
R542C	4	0.32	0.15	ns
R699G	6	0.34	0.77	ns
G702R	4	0.35	2.34	ns
G702E	8	0.41	0.28	ns
I667N	6	0.42	1.05	ns
K466N	4	0.43	0.69	ns
W783G	4	0.47	1.38	ns
M691R	9	0.48	0.70	ns
E677G	8	0.49	0.67	ns
D777Y	11	0.51	3.70	ns
I670V	10	0.52	0.52	ns
T743M	5	0.55	1.47	ns
T743R	7	0.62	2.36	ns
P361T	5	0.64	1.43	ns
G688D	4	0.67	1.51	ns
S753C	4	0.71	0.28	ns
G702V	4	0.75	1.05	ns
Q386R	9	0.75	1.77	ns
H804Y	5	0.75	6.52	ns
Q664H	4	0.76	0.76	ns
R730Q	6	0.80	0.94	ns
M691V	13	0.85	1.70	ns
E662K	5	0.85	0.91	ns
D673E	7	0.85	0.94	ns
N690K	4	0.86	0.75	ns
P361R	19	0.87	1.32	ns
A752T	4	0.88	1.59	ns
Q386H	7	0.88	0.49	ns
G855R	5	0.89	0.46	ns
Y697H	4	0.94	1.27	ns

Yeast Genotype[a]	Count[b]	Fold Induction CAN[c]	IQR[d]	Sig[e]
Y697C	13	0.96	1.89	ns
P361A	8	0.96	1.59	ns
L402V	6	0.97	0.82	ns
N690D	4	0.99	4.27	ns
A627S	5	1.09	5.94	ns
L305P	4	1.13	2.13	ns
E768K	10	1.15	0.70	ns
L755F	8	1.26	2.11	ns
G317S	6	1.35	1.92	ns
G702W	8	1.35	0.40	ns
M745L	7	1.41	0.99	ns
P361L	6	1.47	2.76	ns
G780V	4	1.49	0.60	ns
A54S	7	1.53	1.84	ns
M745I	8	1.55	2.20	ns
S742C	4	1.63	0.49	ns
F803L	4	1.74	3.33	ns
T696R	5	1.77	1.42	ns
T773A	4	1.95	2.53	ns
L457F	5	2.10	1.40	ns
D777N	4	2.21	3.63	ns
S695N	4	2.34	5.42	ns
P634R	4	2.39	1.56	ns
C712R	4	2.47	2.52	ns
L305V	4	2.61	2.84	ns
A627V	4	2.63	1.38	ns
P361S	5	2.66	1.88	ns
C716R	4	2.72	3.15	ns
G158R	5	2.81	5.98	ns
G711V	4	2.89	1.21	ns
G688A	4	2.92	6.55	ns
F799V	4	3.15	3.16	ns
C716Y	4	3.32	5.74	ns
L183P	4	3.37	7.31	ns
D665G	5	3.68	4.25	ns
K694E	4	6.28	5.72	ns

a. The yeast genotype
b. Count, the number of times a variant was assayed in total
c. For stated genotype, variant mutation rate (mutations per cell division) was compared WT. Barcode and chemostat replicates were combined to calculate the median.
d. Interquartile range of all barcode and chemostat replicates
h. Significance is calculated by a Wilcoxian Ranksum test with the Benjamini-Hochberg correction for multiple hypothesis testing. *< 0.05 ** < 0.01 ***< 0.001 ****< 0.0001

3.4 Discussion

In this application of the multi-plexed mutation rate method, we were able to determine the mutation rate of over 2,000 lineages representing ~200 variants of *MSH2*, a critical component of mismatch repair. Though we included a high frequency of WT sequences, our analysis indicates many of these could be substituted with additional VUS to increase throughput at minimal cost to accuracy.

In this work, we estimated the pathogenicity of 157 variants of uncertain or conflicting significance derived from clinical testing. These results provide a key piece of information to testing labs seeking to assign pathogenicity to variants for which little other evidence is available. We have found that our results largely match clinical data obtained from tumors (Shirts et al., 2018). This data in combination with a recent deep mutation scan on Msh2 in a human cell line (Jia et al., 2020) will allow for accurate reclassification of uncertain and conflicting variants in this gene. This will provide physicians better guidance on who to screen for HNPCC and how often.

The current set of alleles could also be tested in different genetic backgrounds, such as in a *san1Δ* background (Arlow et al., 2013) which could give additional information on the stability of Msh2 variants and could determine the mechanism behind an increased mutation rate for some variants. It could also be done in the background containing deletions or change of function mutations in other proteins in the MMR complex. Both variant library and background are completely changeable in this system.

This assay works with any protein which affects mutation rate, and so therefore can assess mutation rate variation in other proteins in the mismatch repair pathway, as

well as in other DNA repair pathways. Assaying variants in the sequence context of the native human cDNA could be possible for genes that complement the orthologous yeast gene knockout (Kachroo et al., 2015; Vogelsang et al., 2009); however our initial attempts to recapitulate the complementation of *mlh1* and *pms1* with human *MLH1* and *PMS2,* other critical MMR genes associated with HNPCC, were unsuccessful. This, however, does not mean that assaying mutation rate of human alleles of DNA repair enzymes is not possible, and in fact would be a very interesting line of study.

In conclusion, in addition to basic science application to this multi-plexed mutation rate assay, it successfully provides evidence for clinical evaluation of variants within the MMR complex.

3.5 Materials and Methods

3.5.1 – Generating barcoded variants

Variants in *hMSH2* found in ClinVar were mapped to the *yMSH2*. Variants containing a homologue within the yeast allele were then ordered as gene products from Twist Biosciences. The gene products were ligated into the pRS413 vector containing the yMSH2 promoter and terminator and transformed into DH5α cells at 30X coverage. DNA was extracted using the Mira-Prep protocol (Pronobis et al., 2016) and then digested with SacI to linearize. A barcode along with randomized sequence were inserted to the linearized vector using Gibson, and then transformed into DH5α cells. Transformants were collected such that there was 5X barcode coverage for each allele. DNA was extracted again using the Mira-prep protocol, and digested with SacI to linearize any unbarcoded alleles and transformed once more to take advantage of *E.*

coli's inability to be transformed by linear DNA that does not have homology overhangs. Colony PCR was done and 0% of clones contained no barcode and ~6% contained 2-3 barcodes. DNA was once again extracted with the Mira-Prep protocol, and then transformed into DBY11069-FY4 *msh2Δ* and *his3Δ flo1Δ* FY4. The *flo1Δ* is present to reduce the prevalence of flocculants (Hope et al., 2017). Transformants were collected such that there was 20X coverage of each barcode. These were then pooled for future experiments.

3.5.2 – Running pools in the chemostat

See section 2.5.2

3.5.3 – Generating amplicon libraries

PCR amplification of the amplicon containing the barcode was done using 15-22 cycles of PCR using custom Nextera Primers (https://figshare.com/s/abd345e21796439ed121 supplemental table 3). Sufficient amplification was determined by qPCR. The amplicon was purified using duel sided Sepharose bead cleanup to isolate the 250bp amplicon. Samples were then pooled at equal molar ratios and run on the NextSeq 550 using paired end reads of both 75 or 150 cycles using custom read and index primers (https://figshare.com/s/abd345e21796439ed121 supplemental table 3) at a read depth of ~1.7 million reads for the mutant population and 0.5 million reads for the non-selective population. Number of reads roughly corresponded with the number of colonies collected for the mutant pool, and 100X coverage of the known number of barcodes for the non-selective population.

3.5.4 – Data pipeline for barcoded libraries

Reads were shortened to the barcode length using a custom python script, fed into PEAR to combine forward and reverse reads, then fed into ENRICH (Fowler et al., 2011) to count barcodes. These counts were fed into a custom R script ((https://figshare.com/s/abd345e21796439ed121 supp file #1) which manipulated data and plotted using ggplot2.

3.5.5 – PacBio Analysis

Plasmid fragments containing the barcode and variant were isolated from *E. coli* using the Wizard mini-prep kit and amplified using PCR with Kapa-HiFi and cleaned up by digesting with Dpn1 and purifying with Ampure beads. Fragments were prepared for PacBio sequencing using the SMRTbell™ Template Prep Kit 1.0 (Pacific Biosciences) and sent to University of Washington PacBio Sequencing Services for sequencing and Sequel II circular consensus sequence (CCS) analysis(Wenger et al., 2019). BAM files of CCS reads were aligned to the plasmid reference using BWA/0.7.13 mem (Li, 2013). Reads that were aligned to the reference sequence were piped to a new BAM file with Samtools/1.9 (Li et al., 2009) and analyzed with cigar strings to validate alignments. Barcodes were then extracted and two barcode-variant maps were generated. One file contains all the barcode-variant reads, and the other contains the highest quality read for each unique barcode. Errors found in these files were corrected using a multiple sequence alignment (MUSCLE 3.8.31) (Edgar, 2004) of reads sharing the same barcode. Final reads were derived from consensus sequences from these alignments.

Ambiguous sequences were fixed by aligning sequences to the highest quality reads using the Needleman-Wunsch algorithm (EMBOSS 6.4.0) (Rice et al., 2000).

3.5.6 – Clinical comparisons

Clinical comparisons were made using retrospective data gathered from clinical laboratory databases for testing performed as part of standard clinical care between 2014 and 2019. This retrospective analysis was done under University of Washington IRB 00007284.

3.6 Acknowledgements

The mapping of the human mutations was done by Adam Gordon, the clinical information was generated and interpreted by Brian Shirts. Chiann-Ling Cindy Yeh handled the generation of the PacBio libraries and generating the variant barcode map using an algorithm developed by herself, Soyeon Showman, and Clara Amorosi.

Chapter 4: Implementing alternative culturing devices

4.1 Abstract

I wanted to build a functional, easy to set up a flexible and robust continuous culture device in the lab that would be suitable for many different kinds of experiments. I began with the Takahashi mini- turbidostat (Takahashi et al., 2015), but quickly ran into intractable issues with accuracy and robustness of the device. I therefore switched to using a larger volume turbidostat that was designed by Anna McGeachy in the Ingolia lab(McGeachy et al., 2019). This device had a “paint-by-numbers” system in which the components have labels which correspond to their position on the circuit board, as well as an easily installable firmware that ran on arduinos (“Arduino - Home,” n.d.). The first use of this turbidostat was to evolve edited rDNA strains. We found that after evolving for ~80 generations there is a re-expansion of the rDNA locus to its normal size (Chapter 2). This expansion of the rDNA locus coincided with an increase in growth rate. Once whole genome sequencing was done on these samples, we found that aneuploidy was a common mechanism of genetic suppression of the slow growth phenotype found in the edited rDNA strains.

In addition to implementing the turbidostat from the Ingolia lab, in collaboration with Mario Leutert from the Villén lab, we sought to implement the MAD. This system enriches old mother cells by coating a starting population in paramagnetic beads which

leads them to stick to a magnet whereas the daughters are washed off. Our goal is to do proteomics on aged cells, to analyze changes in the proteome as cells age.

4.2 Introduction

Different continuous culture devices are useful for generating different populations of cells. This chapter describes my work toward implementing a turbidostat, based on work by the Ingolia lab (McGeachy et al., 2019) and the Mini-stat Aging Device (MAD) system designed by researchers at Calico (Hendrickson et al., 2018), which we call the MAD. For both of these projects, I collaborated with other labs to answer interesting research questions.

First, I established the turbidostat- a device that maintains a constant density in a culture by matching the growth rate of cells to the influx of media, keeping the population size constant throughout the experiment. For more information on the history of turbidostat devices, and the different kinds, see section 1.3. The turbidostat developed by the Ingolia lab has many useful features. (i) It uses standard glass bottles, with standard size stirrers, on a standard stir plate, with easy-to-source caps originally used for high performance liquid chromatography machines. Almost the entirety of the culture vessel is off-the-shelf. (ii) The electronic components are simple to assemble using and the firmware, which is the code which runs the system, was easy to install. (iii) The output of the turbidostat was made with downstream computation in mind. (iv) The lab which designed this turbidostat kindly hosted me to build one in lab, which greatly increased the likelihood for success. All of these factors contributed to the selection of the Ingolia lab's turbidostat.

Once the turbidostat was established, we wanted to determine if the turbidostat would select for the same suppressors to edited genomes as in batch culture experiments. We collaborated with the Brewer lab to probe how yeast respond to a lack of ribosomes caused by removal of the rDNA array locus. For more background on the strains used and rDNA biology, see section 1.3.1. The edited rDNA strains had a doubling time of 195 minutes – over twice the normal WT doubling time of yeast (~90 minutes). This made selection for of genetic mitigation of the rDNA edits simple and straightforward- simply look for a decrease in doubling time. Dr. Brewer had used these strains previously to do a batch culture evolution, by serial batch propagation. She found a chromosome 2 aneuploidy was responsible for the increase in growth rate, and confirmed that it was responsible by analyzing tetrads of a cross between WT yeast and evolved strain (Sanchez et al., 2019). Thus, I was interested to determine if I would also see chromosome aneuploidies develop as suppressors of ribosome depletion in the turbidostat.

In addition to the turbidostat, I worked with Mario Leutert in the Villén lab, using the MADs to age and purify *S. cerevisiae*. For more information on the history of isolating aged cells by magnetic labeling see section 1.4 The original MADs were built to generate enough aged cells for genomic and transcriptomic applications (Hendrickson et al., 2018). That device had several innovations over previous systems, the largest being the continuous removal of daughter cells throughout the culturing process. This system combined the mini-stat device(Miller et al., 2013) with a previous technology which held magnetically labeled yeast to a column which was washed

periodically with media (Janssens et al., 2015). This device makes the retrieval process of old mother cells much easier since at the point of removal, they are already enriched. However, we wanted to generate significantly more material to allow for robust proteomic analyses, while still establishing a consistent environment for the aging cells. For more information on known changes to the proteome with age, see section 1.4.1. Mario Leutert in the Villén lab had developed a rapid multiplexed way to probe kinase networks using phospho-proteomics(Leutert et al., 2019), which can be used to probe the kinase networks of aged cells. The combination of these proteomic assays could provide a more comprehensive view of the mechanisms behind aging in yeast.

4.3 Results

4.3.1 Turbidostat kept population size stable for 300 hours and responded appropriately to temperature changes

I first wanted to test out whether the turbidostat could keep density static, be robust against leaks various other failures, and be responsive to differences in growth rates. I did this test with the *msh2Δ* strain, in an attempt to see if I could do the experiments described in chapters 2 and 3 in the turbidostat. The attempt to track canavanine resistance over time in the turbidostat was unsuccessful, and it seemed that perhaps the WT strain was used instead of the deletion strain. Nevertheless, this first experiment proved that the turbidostat could be run successfully for over 300 hours (13 days) as seen in Fig 4.1. The turbidostat temperature was dropped to 22^{C} at 50 hours and returned to 30^{C} at 65 hours. We see that the doubling time increased from the standard 1.5 hours to 2.25 hours during the time the incubator was at 22^{C}, as would be

expected. This first run in the turbidostat showed that there did appear to be a suppressor that arose after approximately 100 hours in the turbidostat. As this was a test run, no sequencing was completed to confirm this. It was an exciting first pass to indicate that the turbidostat could be run robustly for 300 hours and responded dynamically to changes in temperature and potential evolution.

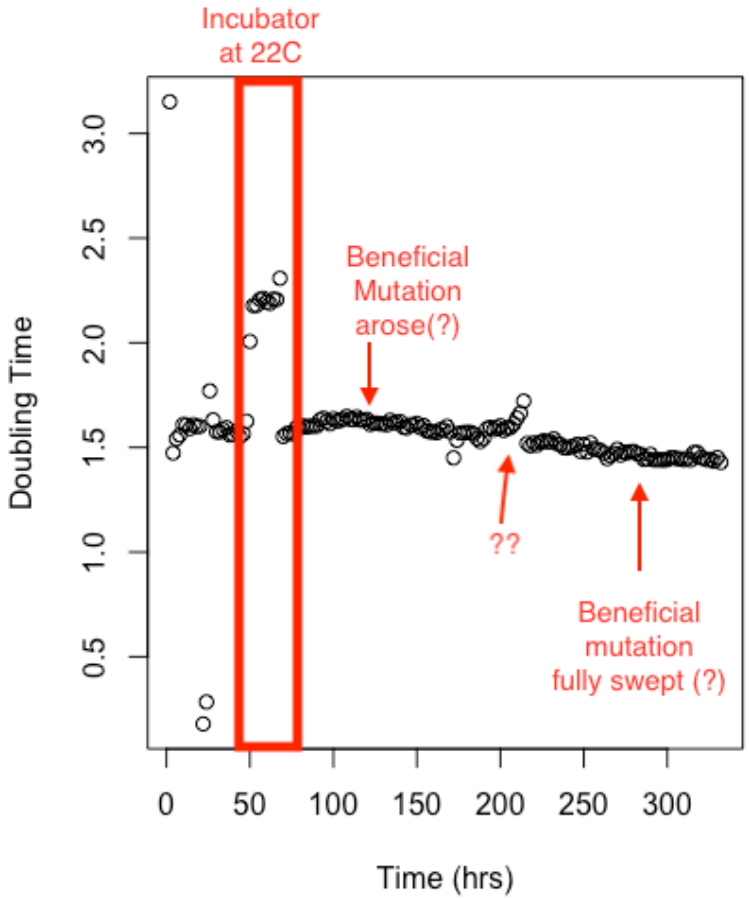


Figure 4.3.1 Turbidostat robustly determines doubling time for greater than 300 hours and responds to temperature fluctuations Influent was tracked over the stated time and used to calculate doubling time. Red box corresponds to time at 22^{C}

4.3.2 Turbidostat capable of accurately picking up sweeping events

Next, I asked if the turbidostat would be sensitive enough to track a faster growing strain out-growing a slower-growing strain. I started by inoculating a turbidostat with a slow growing strain, for which I chose the rDNA deficient strain (Sanchez et al., 2019). This strain would subsequently be used in the suppressor screen described below. Once that strain had reached steady state, characterized by a stabilized density and growth rate, a GFP marked wt strain was added. The GFP marked wt strain was tracked over time using flow cytometry and correlated with the doubling time at the point of sampling. As seen in Fig 4.2, the growth rate of the mixed population accurately reflected the composition of the two strains within it. In addition, the increase in growth rate correlated well with the increased proportion of the faster growing strain. Taken together, these data indicate that the turbidostat can accurately track the growth dynamics of a faster growing strain out competing a slower growing strain using media usage as a proxy for growth rate. This gives confidence to the turbidostat's ability to detect suppressors of slow growth phenotypes during evolution experiments, which are outlined in successive sections.

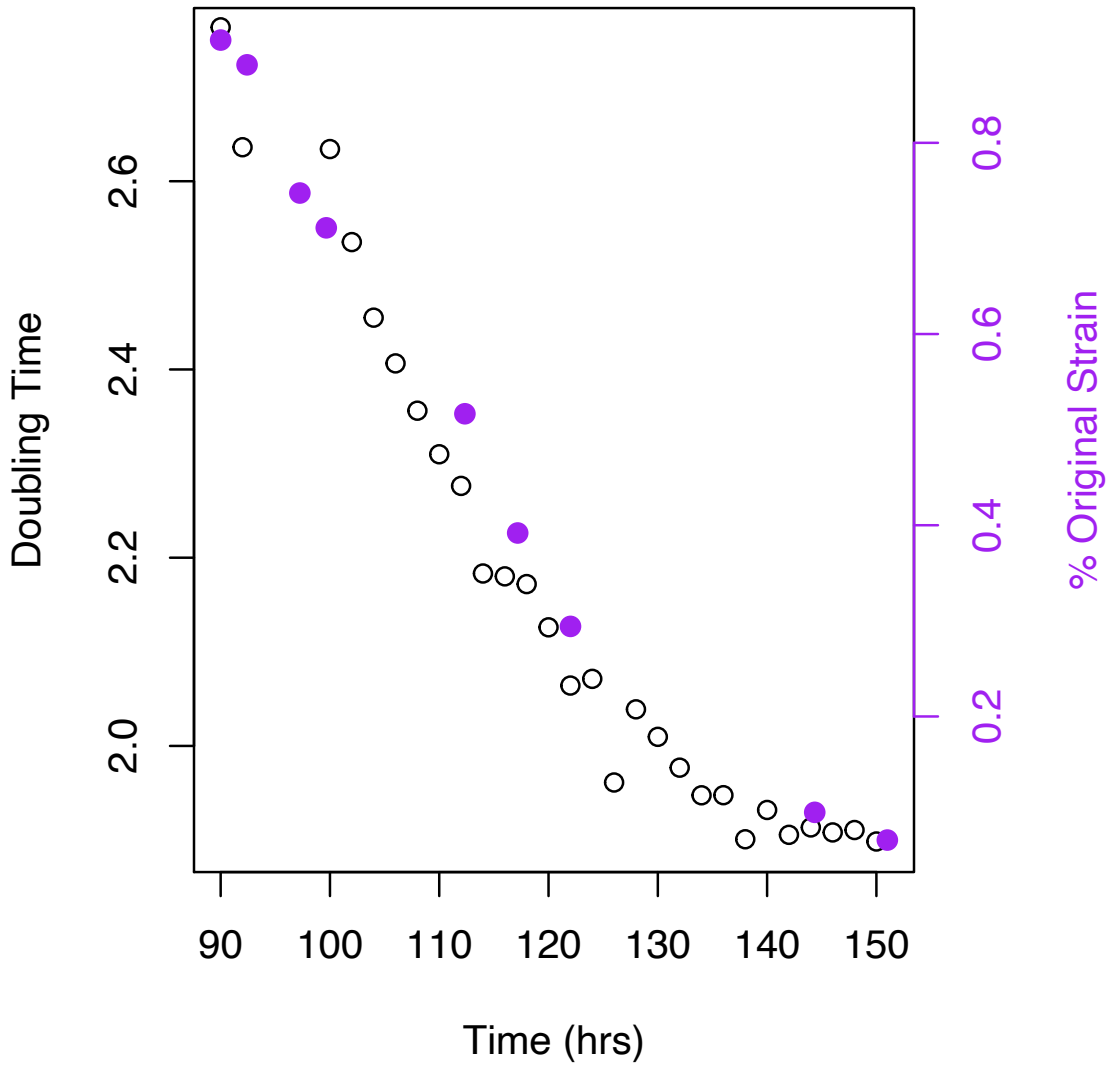


Figure 4.3.2 Addition of WT- GFP strain to isogenic population of slow growing yeast allows for accurate tracking of a sweeping event. Doubling time was calculated by tracking influent over time and is represented in black circles. The percentage of the original strain in purple was determined by percent GFP at each selected time point. The left and right axis were aligned such that 100% corresponds to the doubling time of the slow strain and 0% corresponds to the doubling time of the inoculated strain.

4.3.3 – Increase in rDNA size corresponds with decrease in doubling time

Once I was convinced that the turbidostat was robust, I embarked on a collaboration with Bonny Brewer's group to determine suppressors of rDNA edits. We first wanted to see if we could track the decrease in doubling time to the increase in size of the rDNA locus. We had seen in previous batch culture experiments that in order for cells to regain their growth they had to expand the rDNA locus (Sanchez et al., 2019). In the first attempt of this experiment we ran 4 turbidostats: 2 replicates with the rDNA origin completely deleted, and 2 with the rDNA origin in which all AT sequences are replaced with GC, both of which contain a rDNA copy number of ~10. We saw that for the deletion strains more time elapsed before a decrease in doubling time was seen. Samples were taken and prepared for days 5-9, as marked on figure 4.3.3. In one turbidostat, rARS$^{\Delta}$-1, there was only a small decrease towards the end. The associated Contour-clamped homogeneous electric field (CHEF) gel, which procedure which can separate chromosome sized DNA, (right panel fig 4.3.3) revealed a late expansion of the rDNA locus, and it appeared that only a fraction of the population had this expanded rDNA locus. In contrast, for rARS$^{\Delta}$-2, there was a much swifter reduction in the doubling time which corresponds with the expansion of the rDNA locus appearing at earlier timepoints and nearing complete fixation towards the end of the experiment. For both rARS^{GC} replacement strains, there was a slow decrease in doubling time which matches the slow increase in rDNA locus size. The comparison of doubling time to rDNA locus size, as measured by CHEF gel, indicates that it is simply an increase in the number of repeats that results in the suppression of the slow growth phenotype associated with

initial edits to the rDNA array. It was heartening to see that the turbidostat accurately tracked the increase in rDNA locus size as it affected doubling time.

4.3.4 Chr 4 and Chr 12 aneuploidy found in evolved rDNA GC and rDNAΔ clones.

Next, we wanted to see what kind of suppressors may have allowed for the increase in rDNA locus size for each of the strains. I did whole genome sequencing on 3 samples for each of the 4 replicates for a total of 12 WGS samples. The three samples are marked on the left panel containing the growth curves in figures 4.3.4A-D. In the right panel, we see that there was aneuploidy in all samples. The chromosome 12 aneuploidy- which contains the rDNA locus - seen in $rARS^{\Delta}$-2 does rise in frequency over the time course, however it's difficult to accurately determine if it correlates with the drop in doubling time. For $rARS^{\Delta}$-1 and $rARS^{GC}$- 3 & 4, we see an increase in the population frequency of chromosome 4 (fig 4.3.4B-D). In Fig 4.3.4C, we see that while the middle sample shows a marked decrease in doubling time, the presence of the chromosome 4 aneuploidy is not particularly evident. It is only at the end of the time course that the aneuploidy is at a frequency of 1.24 in the population. Despite the fact that doubling time seems to have leveled off in this turbidostat, the chromosome 4 aneuploidy has not swept – as would be evidenced by an allele frequency of 2. In conclusion, the aneuploidies found in these strains may have affected the ability of the rDNA locus to expand, however, it appears that the growth rate is more closely linked to the expansion of the rDNA locus and not to the relative frequency of the suppressor mutation.

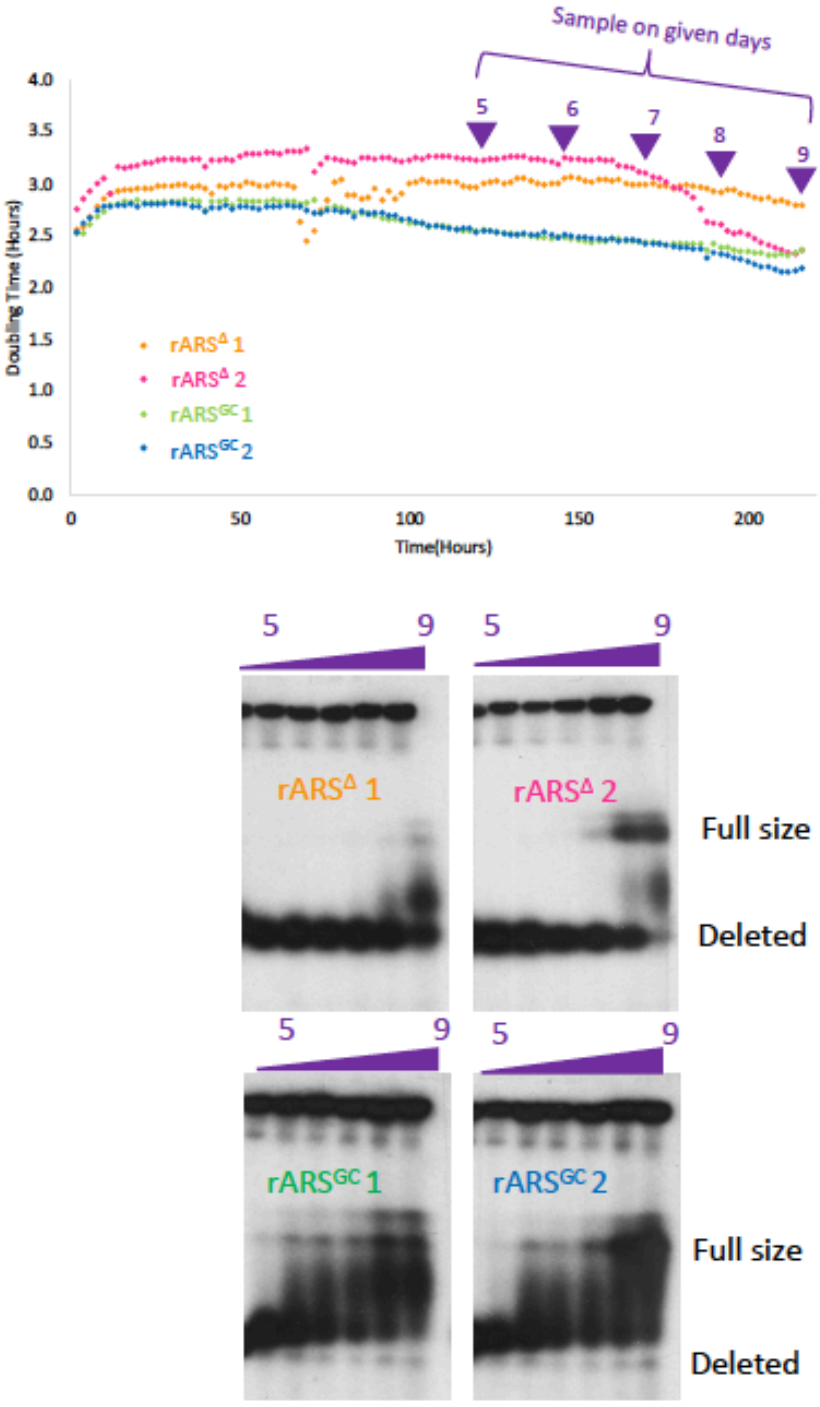


Figure 4.3.3 Increase in rDNA size tracks with decrease in doubling time in edited rDNA strains. Top panel represents doubling time calculated from influent reads and the timepoints which were sampled. Bottom panel shows CHEF gels for samples indicated in the top panel. The Southern blots were probed with the NTS2 probe and reveal changes in Chr XII/rDNA over the course of the continuous growth.

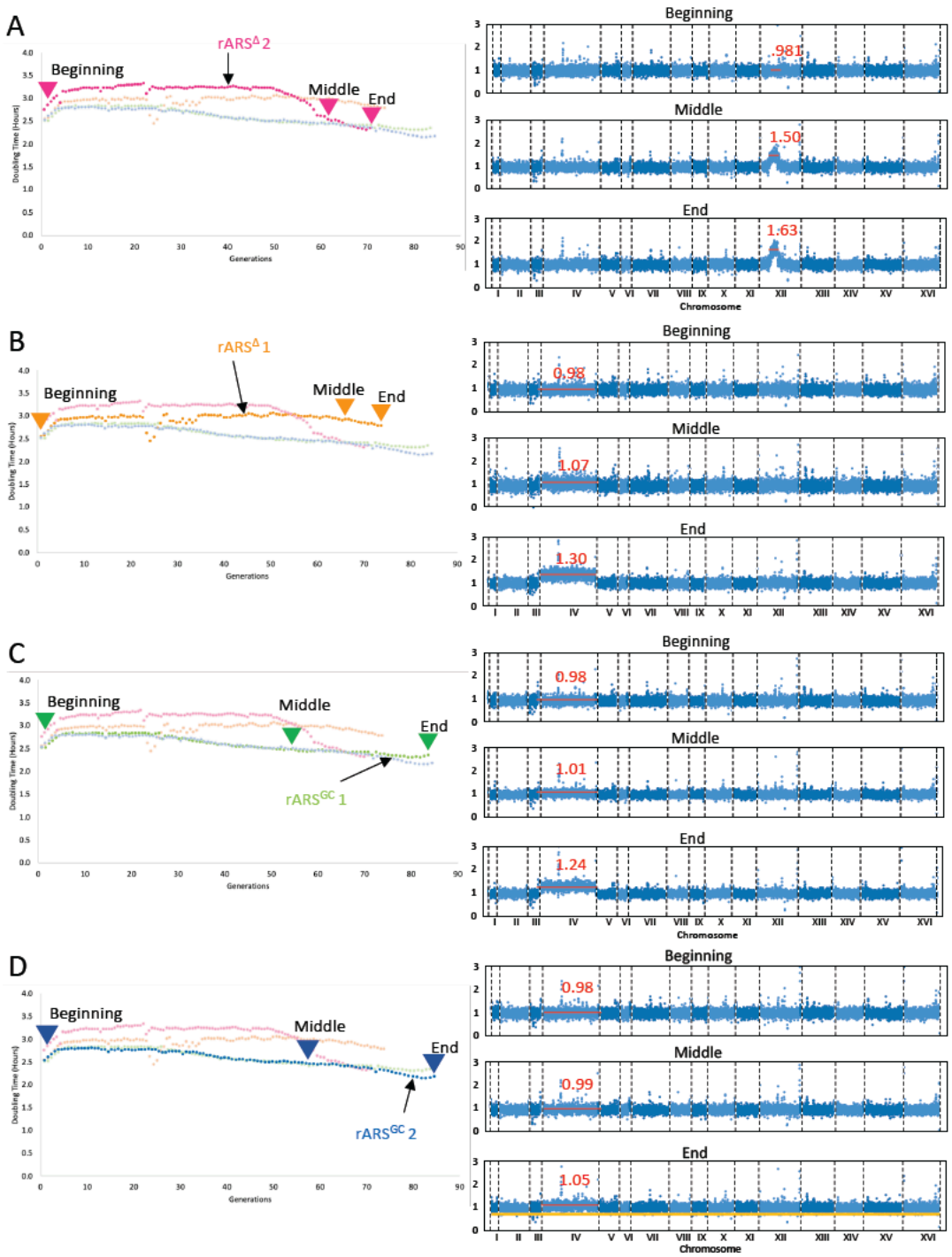


Figure 4.3.4 Chromosome 4 and 12 aneuploidies seen as suppressors to genetic manipulation of the rDNA origin. Left panels represent doubling times of stated strains over the course of the evolution. Right panels represent WGS of stated samples with the median copy number of the aneuploidy in red.

4.3.5 Additional replicates reveal aneuploidy as a common suppressor of rDNA repeat perturbation

In addition to the 4 replicates described in section 4.3.3, I repeated the experiment for a total of 5 replicates each of $rARS^{GC}$ and 3 replicates of $rARS^{\Delta}$. For the second trial of the experiment, I only completed whole genome sequencing at the end point. All samples had evidence of aneuploidy, as determined by comparing the aneuploid chromosome to the other chromosomes using a Wilcoxon Rank Sum test (Fig 4.3.5.2). Interestingly, none of the aneuploidies reached a frequency of 2 in the population, indicating a full sweep. This may be because the experiment was cut short before a full sweep could occur. However, $rARS^{GC}$-2, in red in Fig 4.3.5.1, appeared to have leveled off in doubling time by 180 hours in the turbidostat. It appears that the rDNA locus -still in red- had mostly expanded (fig 4.3.5.1C), which would explain the leveling off. However, the final frequency of chromosome 12 in the population (Fig 4.3.5.2) was only 1.09. This may indicate that the aneuploidy and the rDNA expansion are independent events. It may also indicate that that chr12 aneuploidy, which contains the rDNA array, was used for the initial expansion. Once the initial expansion occurred, the rDNA locus was recombined onto 1 chromosome, and the total chromosome 12 aneuploidy was lost. Clone sequencing of these populations as they evolved would shed light onto the dynamics associated with these sweeping events. While the exact mechanism is unknown, as summarized in Fig 4.3.5.2, aneuploidy existed in all populations evolved in the turbidostat and so therefore is likely involved in the suppression of rDNA edits.

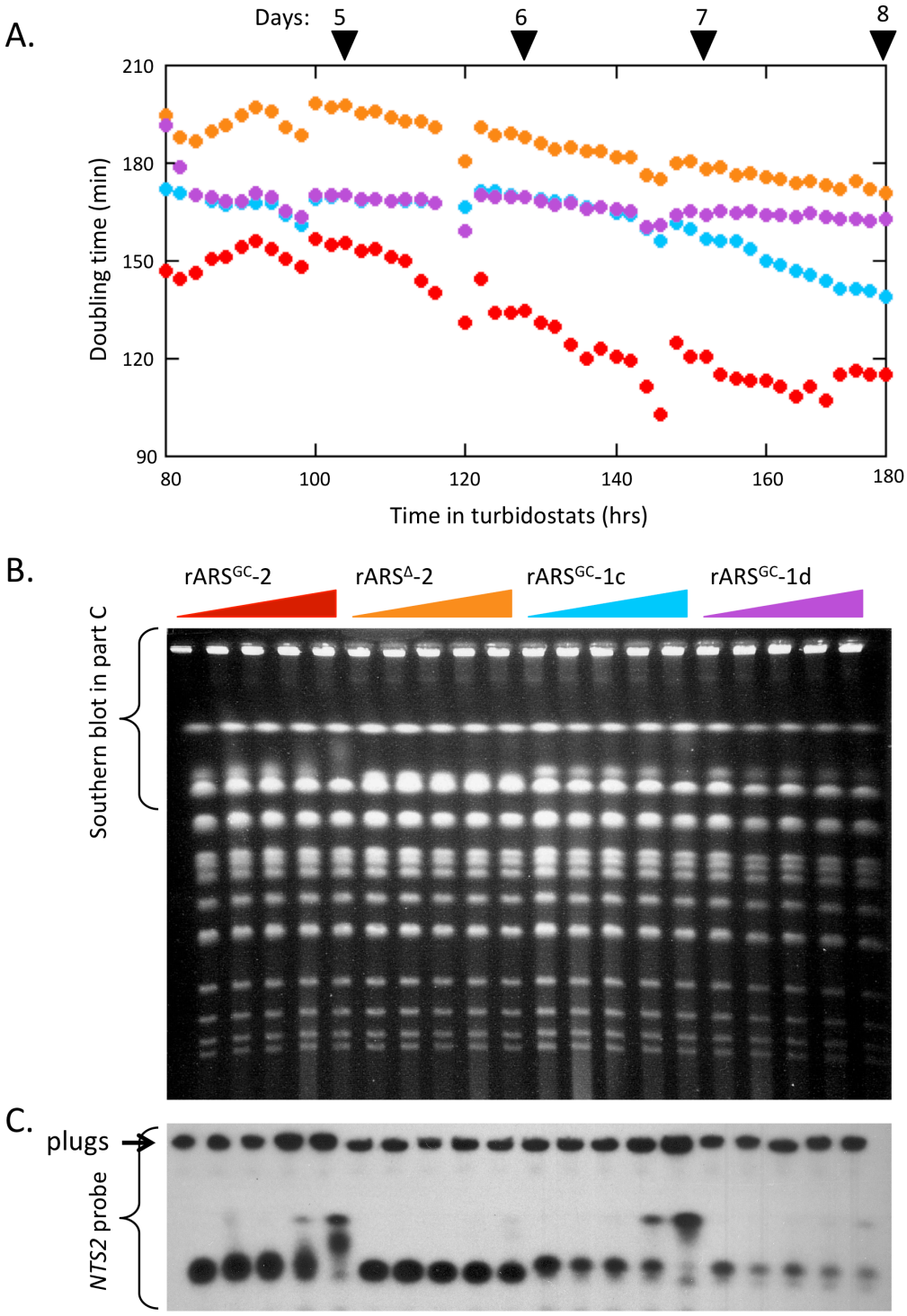


Figure 4.3.5.1 Additional evolutions of deleted and edited rDNA origin strains reveal correlation between doubling time and rDNA locus size. A. Doubling time calculations from influent reads for the strains stated in panel B. B. CHEF gels to represent all chromosomes. C. CHEF gel probed with the NTS2 probe to show changes in rDNA copy number. Figure published in Sanchez *et al* 2019.

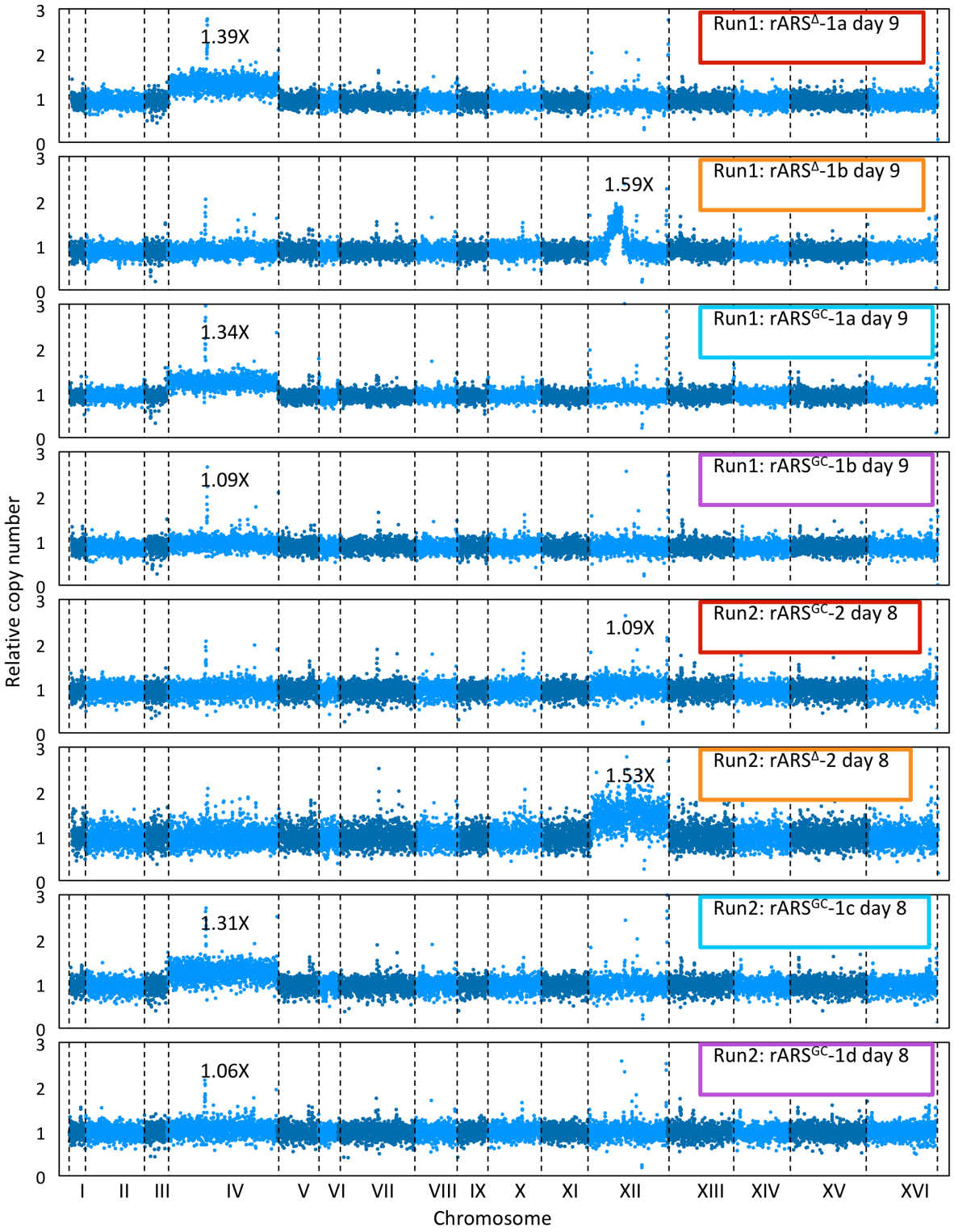


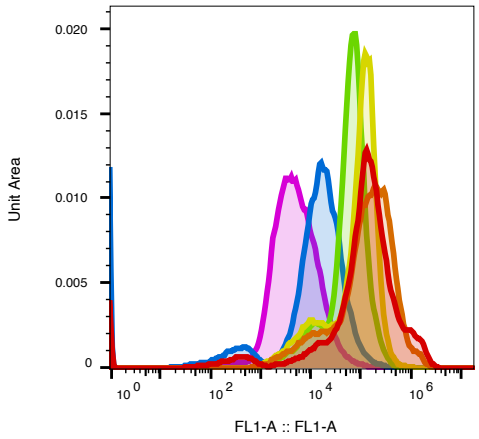
Figure 4.3.5.2 Summary of all copy number changes seen in the evolved rDNA deletion and edited strains. WGS sequencing of the stated populations was completed and the relative copy number plotted. Numbers above aneuploid chromosome represent the median copy number of the copy number variant. Figure published in Sanchez *et al* 2019.

4.4. Results : Implementing the MAD System

4.4.1 – Cellular density does not affect MAD system capability of collecting aged cells

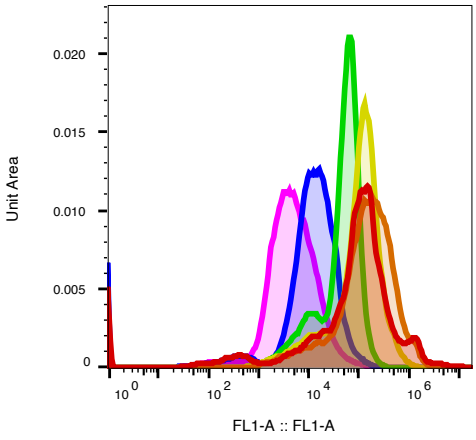
In a first experiment we wanted to determine what cell density can be used in the magnetic stats to obtain pure populations of aged cells. We inoculated two MADs with 3.75x 10^8 cells (25OD) and two with 7.5 x 10^8 cells (50OD). We sampled at 3,24,48,72, and 96 hours. I then stained the cells with wheat germ agglutinin-alexa 488 to stain the bud scars. I ran flow cytometry on the samples, with the expectation that cells that have more bud scars should be more florescent. I also stained cells that were isolated from washes to purify the aged cells, and theoretically should contain daughter cells with few bud scars that were not magnetically labeled. As seen in Figure 4.4.1.1, I see good separation of the populations based on fluorescence up to 48 hours. It appears that the 72- and 96-hour timepoint did not produce populations with significantly more bud scars than at 48 hours. Previous literature on the MAD system ended the experiments at 50 hours (Hendrickson et al., 2018), so it appears that 48 hours is the longest the MAD system can be run. By 96 hours, there were few aged cells left, as the harvesting of cells from earlier timepoints depleted the amount left. From these experiments, we determined that it was easier, more reproducible, and would produce more material if the MADs were harvested in their entirety at different timepoints, versus samples taken continuously over a time course. We also concluded that the max density tested did not affect the age of the culture, so the MAD could successfully age large amounts of cells.

50 OD



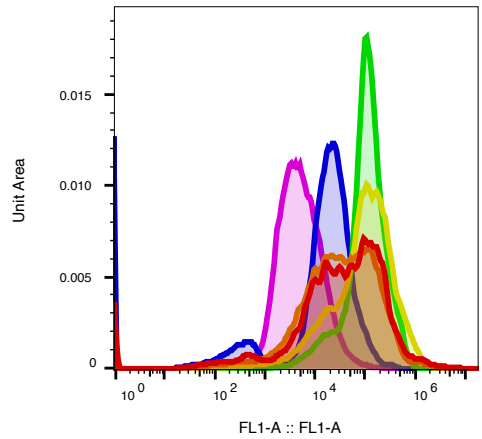
	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	50	A	96	Aged	83679
	50	A	72	Aged	113079
	50	A	48	Aged	67144
	50	A	24	Aged	51390
	50	A	3	Aged	7677
	Inc	-	0	Aged	4528

50 OD



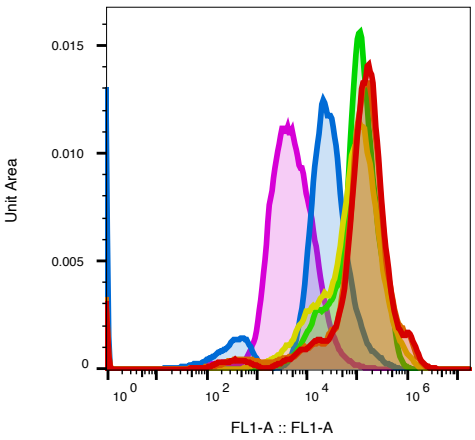
	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	50	B	96	Aged	58533
	50	B	72	Aged	131361
	50	B	48	Aged	84777
	50	B	24	Aged	39784
	50	B	3	Aged	9187
	Inc	-	0	Aged	4528

25 OD



	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	25	A	96	Aged	30067
	25	A	72	Aged	29167
	25	A	48	Aged	81347
	25	A	24	Aged	98940
	25	A	3	Aged	8459
	Inc	-	0	Aged	4528

25 OD



	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	25	B	96	Aged	94555
	25	B	72	Aged	86668
	25	B	48	Aged	74857
	25	B	24	Aged	89894
	25	B	3	Aged	10243
	Inc	-	0	Aged	4528

Figure 4.4.1.1 Flow cytometry of cells with dyed bud scars reveal MAD system produces pure populations of aged yeast regardless of input density 4 replicate plots representing flow cytometry data from populations aged for the stated amount of time. OD 25 corresponds to 3.75x 10^8 cells, OD 50 corresponds to 7.5x 10^8 cells , Fl1 represents the florescence of WGA-Alexaflor 488, a proxy for bud scar count.

I next used microscopy to examine the 0,3,24,48, and 72 hour time points in order to determine if the flow cytometry results correlated with manual bud scar counts. Microscopy images can be seen in Figure 4.4.1.2. Mario Leutert counted the bud scars, with a summary in Figure 4.4.1.3, and found that the bud scar count closely correlated with total cellular fluorescence up to 48 hours, as be expected(Janssens et al., 2015). Next, we wanted to check the daughter populations, which were washed from the aged cells and should, in theory, have only undergone 0-1 replications. Flow cytometry data on those populations are shown in Figure 4.4.1.5 with comparison to aged populations in Figure 4.4.1.5. We see that while the magnetically coated populations are more enriched for mother cells, the daughter populations also appear to have aged cells in them. This may be a result of washes on magnets that were too weak for the task, meaning that we lost magnetically coated mother cells. Alternatively, it could indicate that daughter cells were not being washed off during the course of the experiment. However, taken in total, this first proof of principle experiment indicates that we can scale the MADs system up for proteomics.

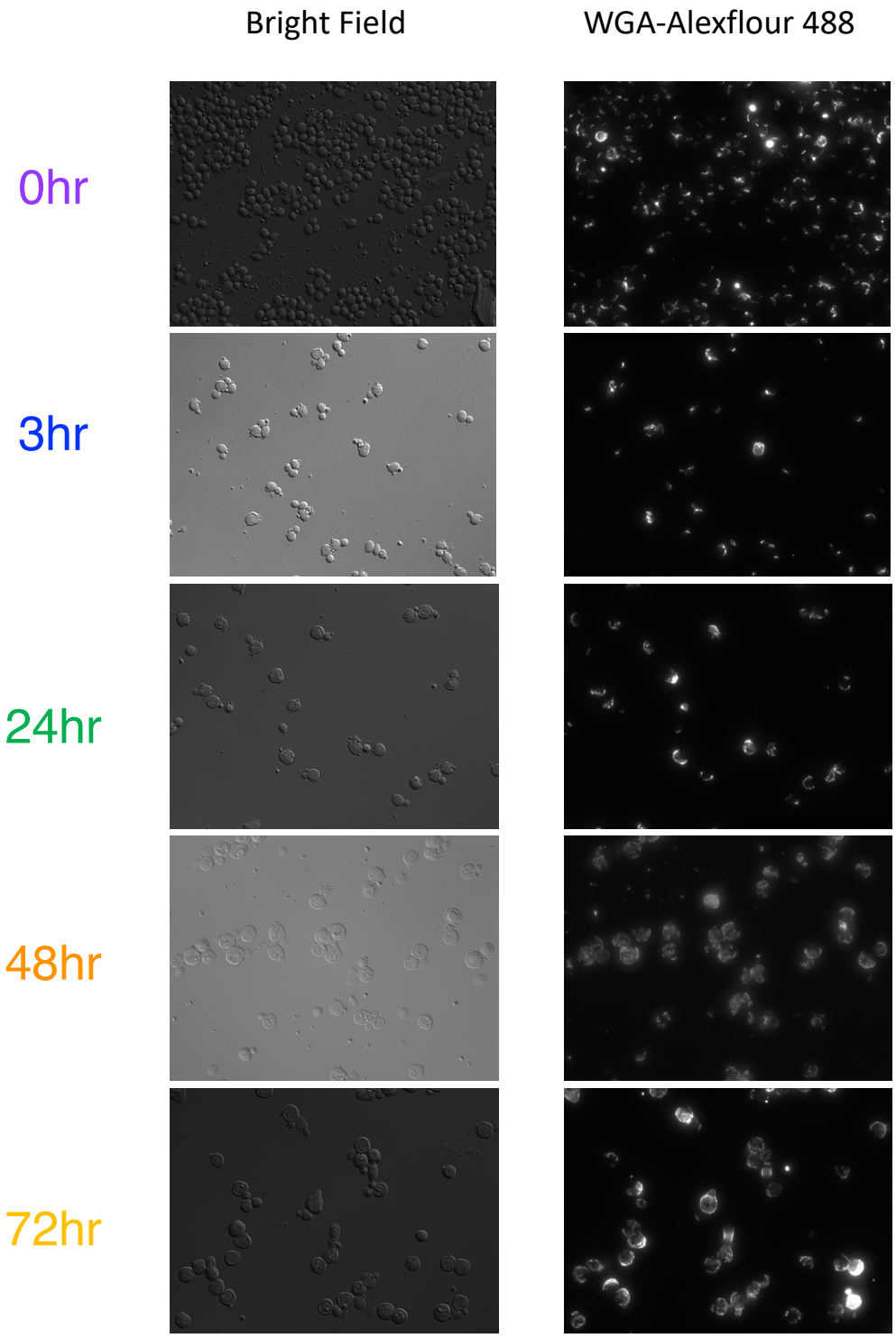


Figure 4.4.1.2 Microscopy of MAD system aged yeast cells shows cells aged in the MAD system have many budscars Left panel is bright field images, right panel is GFP florescence of WGA-Alexaflor 488 stained cells to reveal bud scars. Rows are the time spent in the MAD.

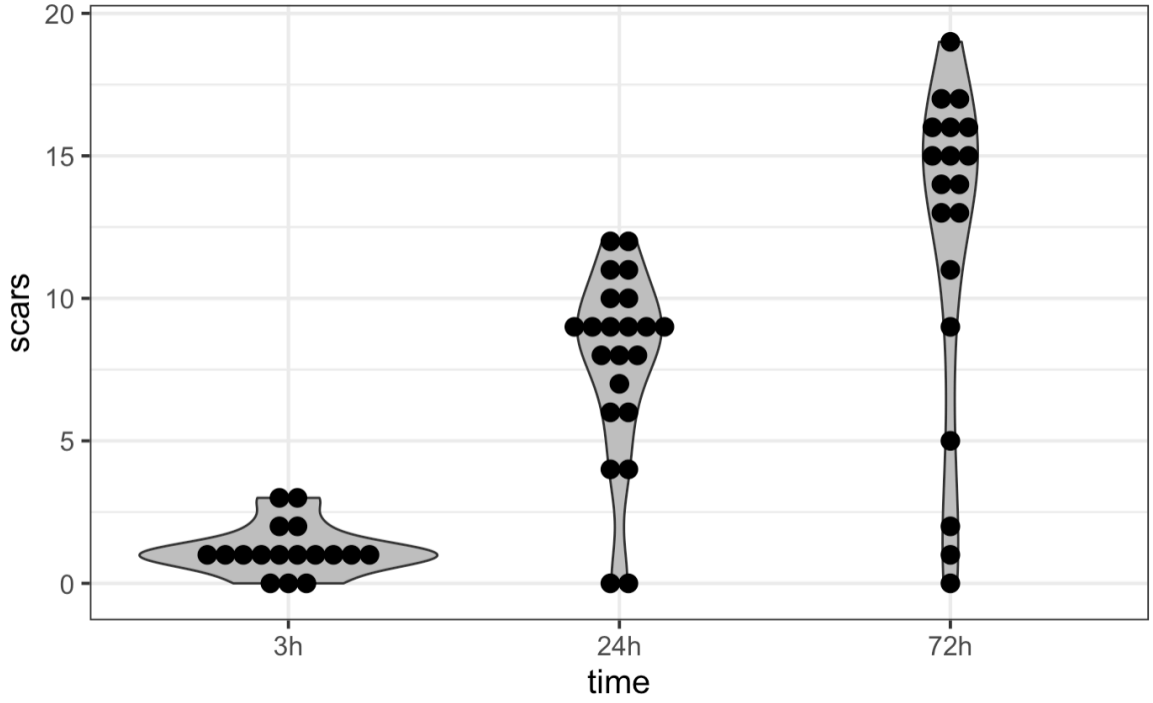
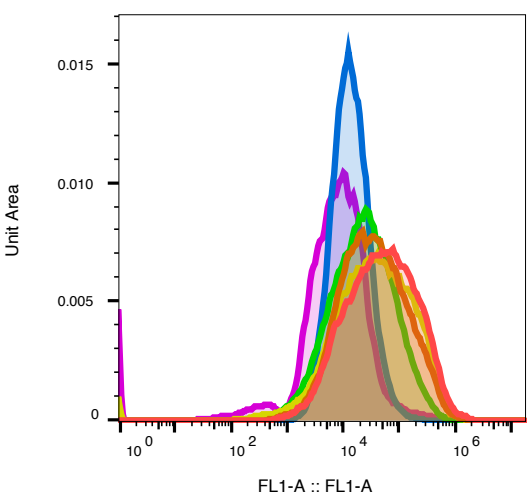
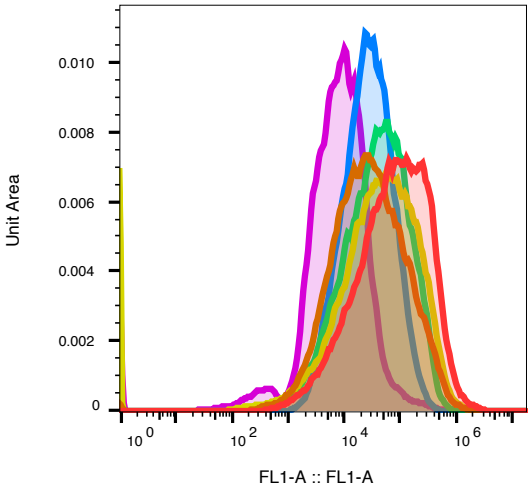


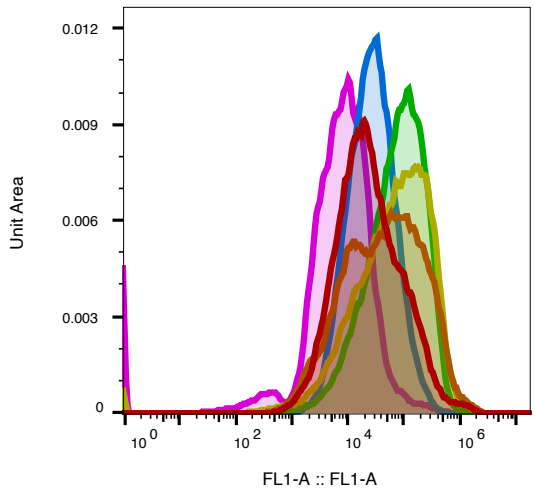
Figure 4.4.1.3 Manual counts of bud scars recapitulate bulk flow cytometry data Bud scars were counted from images in Figure 4.4.1.2 for each of the stated timepoints. Data collected by Mario Leutert.



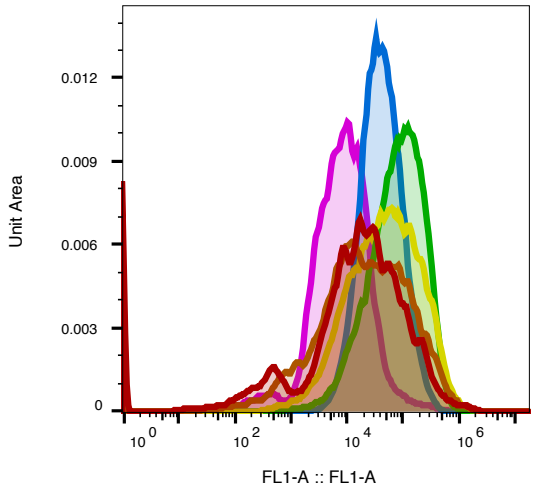
	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	50	B	96	Sup	46037
	50	B	72	Sup	33652
	50	B	48	Sup	33227
	50	B	24	Sup	21249
	50	B	3	Sup	13275
	0	Inoculum	0	Sup	6397



	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	50	A	96	Sup	69352
	50	A	72	Sup	27109
	50	A	48	Sup	23470
	50	A	24	Sup	34430
	50	A	3	Sup	25822
	0	Inoculum	0	Sup	6397

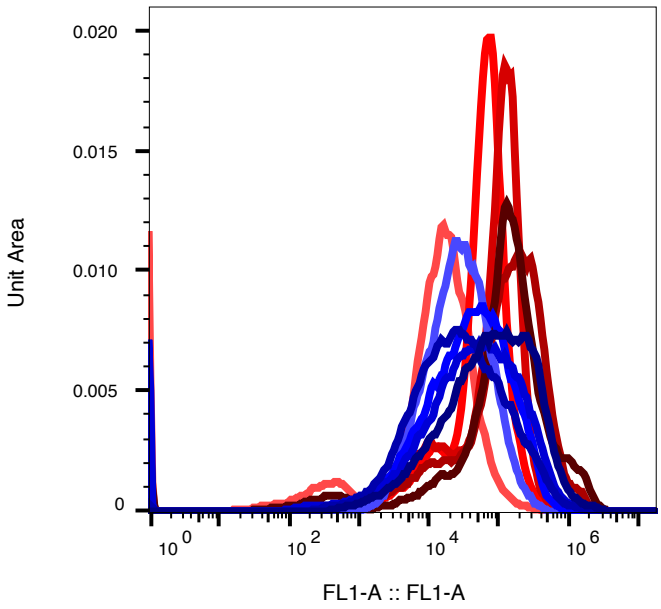


	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	25	A	96	Sup	23646
	25	A	72	Sup	41676
	25	A	48	Sup	59635
	25	A	24	Sup	70792
	25	A	3	Sup	25936
	0	Inoculum	0	Sup	6397



	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	25	B	96	Sup	10443
	25	B	72	Sup	18535
	25	B	48	Sup	44800
	25	B	24	Sup	72494
	25	B	3	Sup	39912
	0	Inoculum	0	Sup	6397

Figure 4.4.1.4 FloFlow cytometry of daughter populations from the MAD system reveal contamination with aged mother cells 4 replicate plots representing flow cytometry data from the supernatant of the purification of aged cells which contain unlabeled daughters. OD represents the density of the aging culture and FI1 represents the florescence of WGA-Alexaflor 488, a proxy for bud scar count.

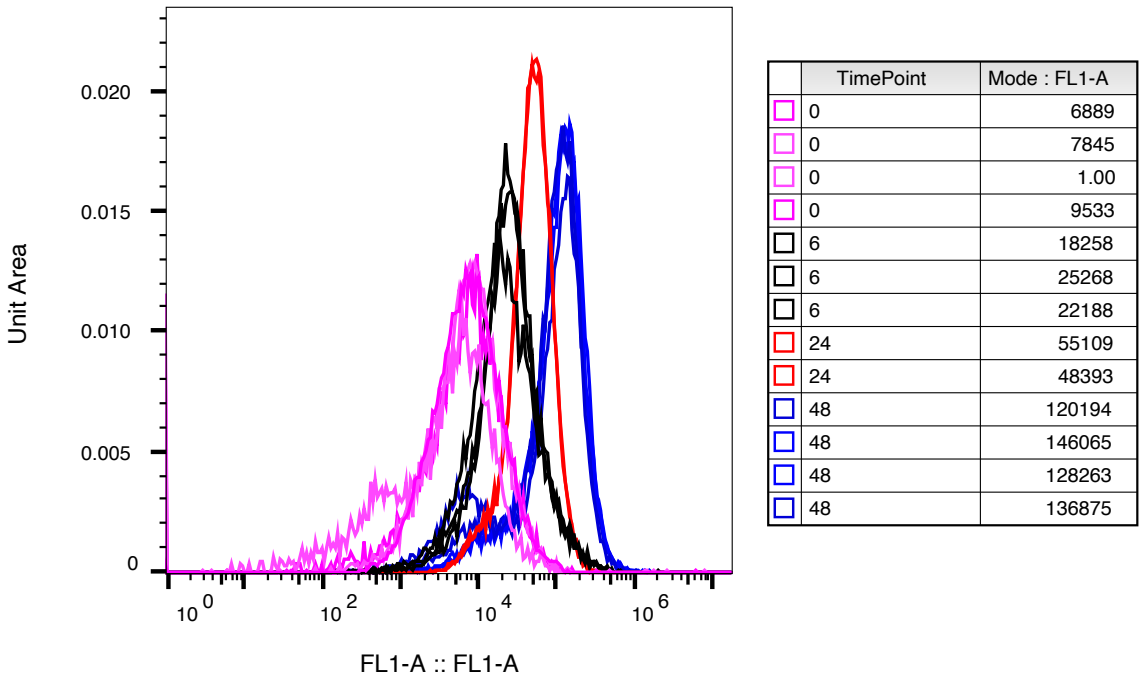


	OD	Replicate	TimePoint	Isolated	Geometric Mean : FL1-A
	50	A	96	Sup	69352
	50	A	72	Sup	27109
	50	A	48	Sup	23470
	50	A	24	Sup	34430
	50	A	3	Sup	25822
	50	A	96	Aged	83679
	50	A	72	Aged	113079
	50	A	48	Aged	67144
	50	A	24	Aged	51390
	50	A	3	Aged	7677

Figure 4.4.1.5 Comparison of aged and non-aged populations indicate daughter population from aged cells are not pure Flow cytometry data of mother (in red) and daughter cells (in blue) isolated from all timepoints in the experiment. OD represents the density of the aging culture and FI1 represents the florescence of WGA-Alexaflor 488, a proxy for bud scar count.

4.4.2 – Large scale collection of aged populations

Once we had evidence that increasing the population size of the MAD system didn't affect the age of the populations within it, we wanted to scale to a point of being able to collect enough material and replicates for quantitative proteomics and phospho-proteomics. This required the initiation of 15 MADs, as even at the higher density achieved in section 4.4.1 one stat did not provide enough material. We also increased the magnet strength, as it appeared we were losing a significant number of mother cells throughout the run. During this experiment we grew 15 MADs to 6h, 24h, and 48h) and then harvested all 15 vessels at that time. We then pooled 3,4,4, and 4 MADs together to collect sufficient material for mass spectrometry analyses. I then stained these populations with WGA-alexaflor 488 to determine the number of budscars. Flow cytometry data from these populations are in figure 4.4.2.1, where we see separation between the differently aged populations. This was verified by microscopy in figure 4.4.2.2. These data indicate that the MADs reproducibly produce aged populations with equivalent amounts of bud scars as measured by florescent intensity.



	TimePoint	Mode : FL1-A
	0	6889
	0	7845
	0	1.00
	0	9533
	6	18258
	6	25268
	6	22188
	24	55109
	24	48393
	48	120194
	48	146065
	48	128263
	48	136875

Figure 4.4.2.1 **Flow cytometry of cells with dyed bud scars reveal MAD system produces pure populations of aged yeast reproducibly** All cells were dyed with WGA-alexa-flor 488 to stain bud scars. A minimum of 10K events were collected for all populations. FI1 corresponds to florescence of bud scars. Colored by timepoint collected.

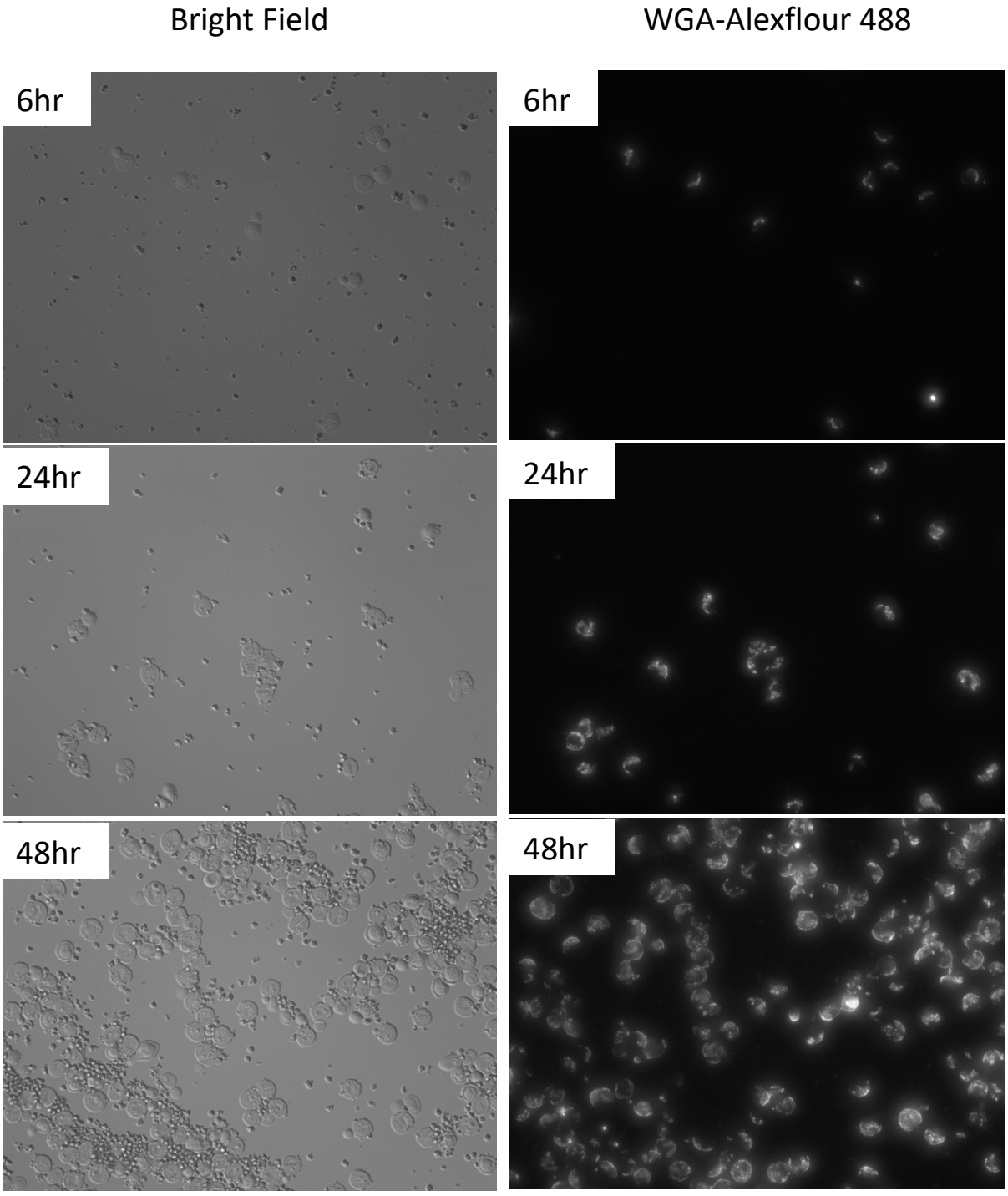


Figure 4.4.2.2 Representative microscopy images from aged populations. Left panel is bright field images, right panel is GFP florescence of WGA-Alexaflor 488 stained cells to reveal bud scars. Rows are the time spent in the MAD, and correspond to the same time stated in figure 4.4.2.1.

4.4.3 – Aged populations and their daughters have distinct proteome profiles

Once we had obtained pure populations of replicatively aged cells, we wanted to see whether on a global level there are proteomic changes with age. We have six populations: yeast aged 6, 24, and 48 hours and daughters isolated from those three populations. The average replicative age for the mother cells at each time point are: 4, 9, and 17 generations respectively. A PCA analysis comparing the proteomic peptides completed on these 6 populations is in figure 4.4.3.1. The PCA show high reproducibility between biological replicates. PC1 separates based on cell age and PC2 separates mothers from daughters. It appears that all six populations are distinct from one another. There is a split between mother and daughter populations, however daughters isolated from the aged cells are more similar to the mother they were isolated from than the mother that was isolated at 6 hours. This indicates that daughter populations from aged mothers are not fully rejuvenated, as one might expect from previous data(Kennedy et al., 1994). However, it is unexpected that even at 24 hours, in which the cells only have a median of ~9 bud scars, there already is a distinct shift. Previous studies have shown that a lack of rejuvenation doesn't start until approximately 15 replications, and in fact have shown identical death curves for 8^{th} and 12^{th} buds of a mother. It could be that proteomic changes precede changes in viability. However, it could be that even a small contamination of aged cells is causing this. Due to some technical difficulties, daughters were not collected for the 6 and 48 hour time points, nor was live/dead staining performed. Figure 4.4.3.2 shows separation of mother and daughter cells at 24 hours as measured by bud scar florescence, however there is

some overlap. This must be kept in mind when analyzing data from these populations. In addition, live/dead staining done on the 24h time point shows that the mother population does contain more dead cells than the daughter population (figure 4.4.3.2). More experimentation and analysis needs to be done to identify the observed differences in the daughter population. Yet, even with these caveats, it is very exciting to see that on the global level there are distinct proteomic differences associated with age.

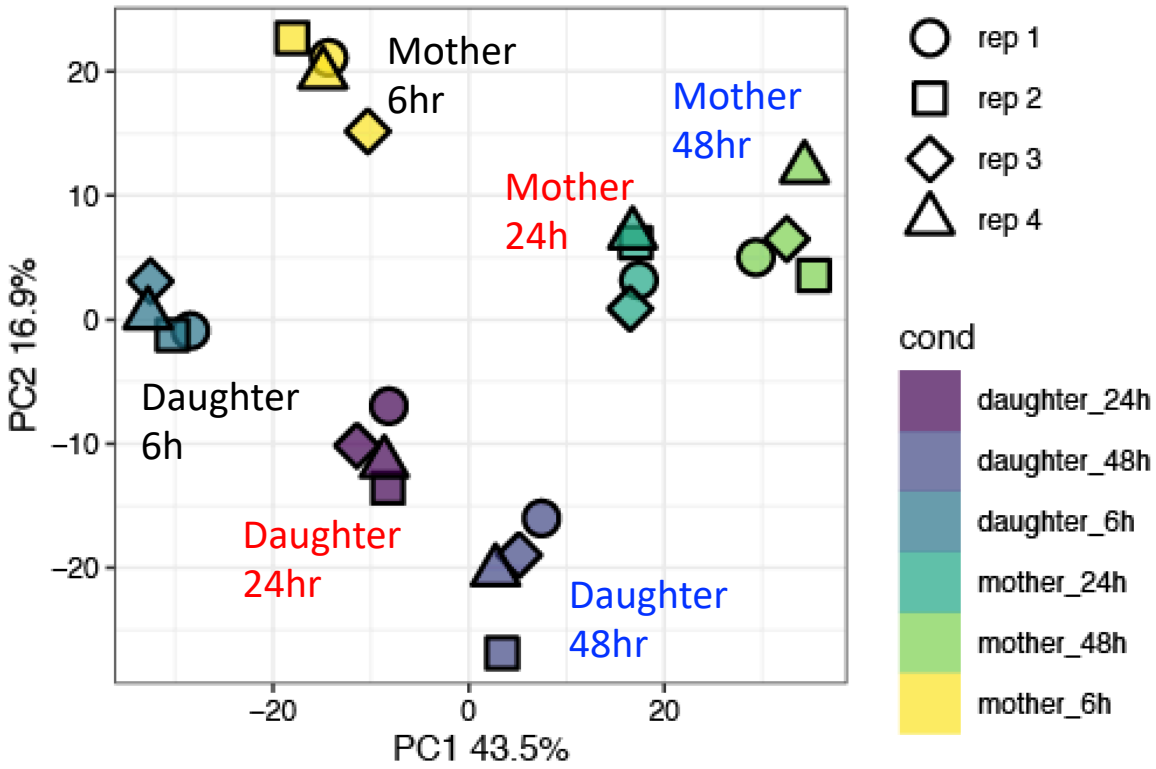
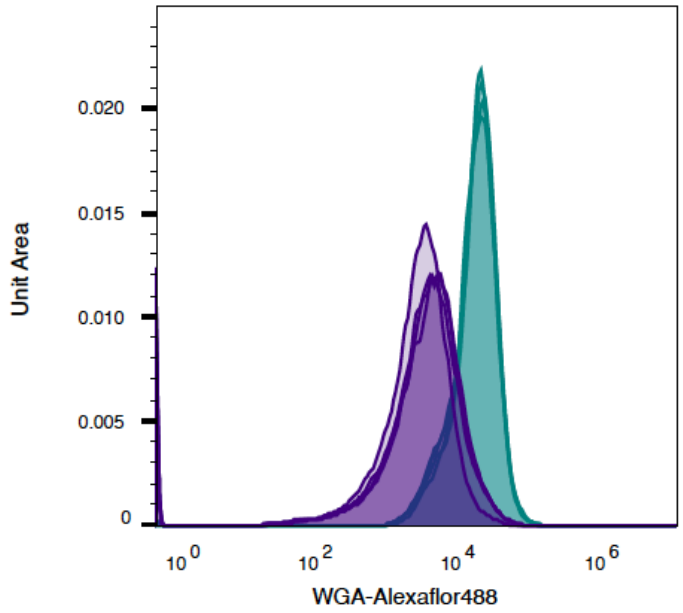
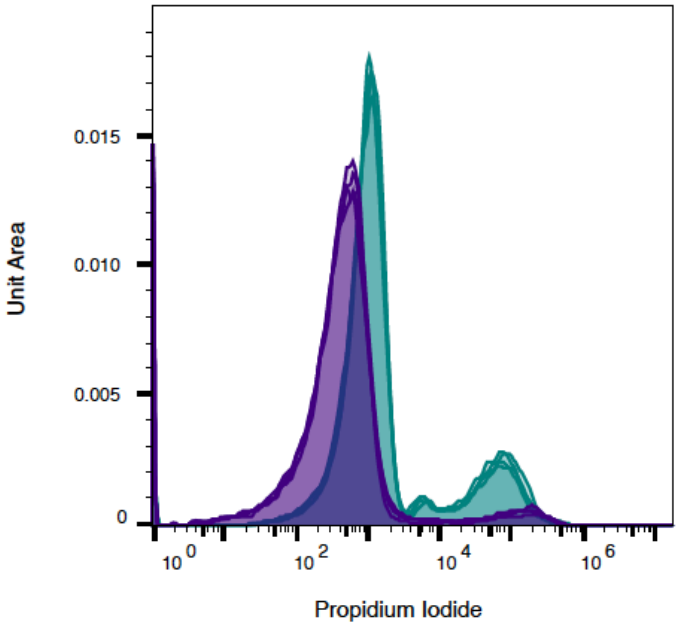


Figure 4.4.3.1 PCA analysis of mother and daughter populations isolated from stated timepoints Analysis carried out by Mario Leutert comparing whole proteome from stated populations.



	TimePoint	Isolated	Mode : FL1-A
	24	Daughter	5668
	24	Daughter	1.00
	24	Daughter	1.00
	24	Daughter	1.00
	24	Mothers	32768
	24	Mothers	30706
	24	Mothers	32768
	24	Mothers	32768



	TimePoint	Isolated	Mode : FL3-A
	24	Daughter	1.00
	24	Daughter	1.00
	24	Daughter	1.00
	24	Daughter	1.00
	24	Mothers	981
	24	Mothers	1117
	24	Mothers	1046
	24	Mothers	1192

Figure 4.4.3.2 FLOW cytometry reveals some overlap in replicative age between mother and daughter cells isolated from same time point and differences in viability Cells were stained with either WGA-Alexaflor-488 for bud scars or Propidium iodide for live/dead and >10,000 events were collected. Mother refers to the aged cells and daughter refers to the daughters collected from the aged cells.

4.5 Materials and Methods

4.5.1 Assembling circuit boards

Circuit boards and parts were ordered according to lists stored in building turbidostats folder on the Dunham lab drive. There are ReadMes in both folders for information on ordering. After the circuit boards arrive, it's important to use actual lead solder, as the nonleaded solder does not provide a resilient connection. I performed all soldering inside a chemical hood. The housing for the boards were 3D printed according to the files in the building turbidostats folder, using PLA plastic in a bright color, and then spray painted with matte black spray paint to avoid glare when measuring the absorbance. Once the circuit boards had been assembled and housed inside the 3D printed coverings and attached to the ring which holds the culture vessel, the only thing left to do is load on the firmware. The firmware to be loaded onto the Arduino is also in the building turbidostats folder, and a ReadMe is included with a link to how to set up the environment to be able to load the firmware onto the turbidostats. Once all that is completed, the circuitry is ready to be used.

4.5.2 Assembling motors

Boxer pumps were ordered https://www.boxerpumps.com/en/diaphragm-pumps-peristaltic-pumps/. Connections wire were soldered on to the pumps and attached to the controlling circuit boards. All parts for this are in the building turbidostats folder. One key point is that only the suggested tubing for this Boxer pump (it is also on the website) will work. Attempting to use a different OD/ID type of tubing will lead to leakage and

improper function! The connections on the pumps are a little sensitive, so use true lead solder and rubber shrink housing to keep them safe.

4.5.3 Assembling the culture vessels.

Hard tubing was fed through the ports on the Vic- cap, and then flexible silicone tubing was attached by putting it over the hard tubing and gluing it in place with waterproof adhesive. The length of the effluent hard tubing determines the level of water in the vessel, so it is useful to determine if it's at the correct level before gluing it in place. It is important to use the manufacturer's specified tubing for the tubing that goes through the pump. Initial attempts using what I thought was close enough in size tubing did not work, with media leaking in when the pump is not running. Once these have been made, setting up the tubing is very similar to the way it's set up in the chemostat manual.

4.5.4 Turbidostat growth conditions

Samples were harvested at the indicated times by collecting the effluent on ice over a ~15-min interval. Custom R scripts were used to convert the calculated influent amount to a doubling time. Run 1 parameters: four separate colonies, two each of rARSGC-1 (colonies a and b) and rARSD-1 (colonies a and b), were grown in WFC medium. Each 200-ml turbidostat was inoculated with 2 ml of an overnight culture. The cells were then allowed to grow until they reached OD_{600} 1. At this point, another sample was taken to manually measure the OD600 of the cells, which was then used to calibrate the turbidostat readings. The turbidostat was then set to maintain a constant OD_{600} of 1. At the same time each day, 15 ml of effluent from each turbidostat was collected on ice

(~10– 15 min). Cells were pelleted at 4000 3 g for 5 min and frozen at -20^{C} for downstream analysis. Glycerol stocks of 1 ml were also collected and stored at -80^{C}. Run 2 parameters: the OD600 was set to 0.5 to more accurately reflect log phase growth, and 50-ml samples of cells were collected to account for the decreased density and to provide adequate samples for additional assays. For doubling-time analysis, the system measures the on/off cycling of the media pump that is delivering medium at a known rate. A custom R script fits a line to the change in the overall time the pump is on over a 2-hr window. The slope of that line is then scaled by the known flow rate and volume of the vessel. This calculation produces a doubling time for that 2-hr interval.

4.5.5 Generation of CHEF gels

For CHEF gel analysis of chromosome sizes, 1–1.5 ml of saturated yeast cultures were embedded in 0.5% GTG (Genetic Technology Grade) agarose plugs and processed as described(Kwan et al., 2016). We used three CHEF gel running conditions in a CHEF DRII (Bio-Rad, Hercules, CA): the full chromosome run (which separates all chromosomes except for IV and XII) was 64 hr in 0.8% low electroendosmosis (LE) agarose with 0.53 TBE at 165 V, with switch times from 47 to 170 sec; the rDNA run (which separates larger chromosomes, such as IV and XII) was 68 hr in 0.8% LE agarose with 0.53 TBE at 100 V, with switch times from 300” to 900”; the Fsp1- digested chromosomes were run as for the full chromosome run but for only 44 hr. To collect DNA for two-dimensional (2D) gels, cells were grown to midlog phase, quick chilled in 0.1% (w/v) sodium azide and 200 mM EDTA, and concentrated by centrifugation. Agarose plugs from frozen cell pellets were prepared by the protocol available at http://

fangman-brewer.gs.washington.edu. Methods for restriction digestion of DNA in plugs and conditions for 2D gel electrophoresis of rDNA replication intermediates can be found in (Sanchez et al., 2017). Southern transfer and hybridization with 32P-dATP PCR-amplified probes were performed according to standard protocols. The split Southern/northern gel was performed as described in (Sanchez et al., 2017)

4.5.6 Whole genome analysis

Libraries from the turbidostat cultures were prepared by isolating genomic DNA using the Hoffman–Winston protocol (Hoffman and Winston, 1987). DNA concentration was assessed using Qbit. Aliquots of genomic DNA (50 ng) were prepared with the Nextera Library kit, according to the kit's instructions. Alignment and SNP/insertion/deletion variant calling WGS reads were aligned using BWA-mem (BWA/ 0.7.15) (Li, 2013) to the sacCer3 reference genome(Engel et al., 2014) then sorted and indexed using SAMtools/1.9 (Li et al., 2009). Duplicates were marked and removed using Picard tools (picard/2.6.0), resorted, and indexed using SAMtools. The insertion/deletions in the alignments were realigned using the GATK/3.7 package. Variants were called using freebayes/1.0.2-6-g3ce827d (Garrison and Marth, 2012)) with modified arguments (–pooled-discrete– pooled-continuous–report-genotype-likelihood-max–allelebalance-priors-off–min-alternate-fraction 0.1) and LoFreq/ 2.1.2 (Wilm et al., 2012) in a paired mode with their genetic ancestor. Called variants were filtered for uniqueness against their genetic ancestor(s) using bedtools/ 2.26.0. The variants were filtered for quality using bcftools/1.9. The filtered variants were annotated(Pashkova et al., 2013) and

manually inspected for accuracy using the Interactive Genomics Viewer (IGV) (Robinson et al., 2011).

4.5.7 Copy number and rearrangement analysis

The copy number was plotted using 1000-bp sliding windows (IGVtools) normalized by the mean total read depth across the genome (GATK/3.7) and manually inspected for changes in copy number with the sample's genetic ancestor. Copy number change breakpoints were manually inspected to determine the type of rearrangement using split and discordant reads, generated with BWA-mem (Li, 2013), SAMBLASTER/ 0.1.24 (Faust and Hall, 2014), and SAMtools.

4.5.8 Magnetically labeling yeast

Yeast was grown overnight in SMC+2% glucose+0.2%mannose. Overnight cells are diluted to an OD of ~ 0.1 in SMC+2% glucose+2%mannose. Cells are grown until they reach ~ OD 0.6 . The cells are then washed twice in 1X PBS to remove any amine (if there are excess amines in the media the beading will not work) and resuspended in 1X PBS + 30% PEG 3350. Cells are concentrated to ~ 50OD/ml and sonicated for ~30 seconds in bench top water bath sonicator to break up any clumps. 6.25mg of Sera-Mag SpeedBead Carboxylate-Modified Magnetic Particle (Hydrophilic) are washed 2X in 1XMES. The beads are then resuspended in a solution containing 1xMES + 100mM N-hydroxysuccinimide + 35mM -ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride and incubated at room temperature for 30 minutes. Once the beads have been incubated, they are washed on a magnet 3x with 1xMES. Cells in 1XPBS + 30% PEG are added and then incubated for 30 minutes. After labeling, cells are washed

twice in SMC+2% glucose+0.2%mannose, and then resuspended to a density of 2ODs/ml in SMC+2% glucose+0.2%mannose.

4.5.9 Building and loading the MADs

The MADs were built as described in (Hendrickson et al., 2018). If I were to build them again, I would mill a 4th hole for sample addition. The MADs are allowed to fill with SMC+2% glucose+0.2%mannose. The labeled yeast cells are then added through the air inlet so that the final OD in the MAD is 50 OD. The cells are given 5 minutes to attach to the magnets ringing the MAD before the pumps are turned on to a rate of approximately 1 replacement volume/ hour.

4.5.10 Sampling the MADs

The MADs were run for the stated amount of time. Every ~12 hours and 1 hour before collection, the media and air pump were turned off and each MAD was taken out and vortexed, to remove any daughter cells which may have been stuck between the labelled cells. At each stated time point, the MAD was harvested, where the aged mother cells were further purified by washing ~10 times on a magnet. Daughter cell samples were isolated from these washes and should represent new daughters. Samples were either snap frozen in liquid nitrogen for proteome analysis or placed in PBS + 1M Sorbitol + 4mM EDTA for Bud scar counting and Live/Dead staining. Staining was done by mixing 200ul of cells with 1 μL 5 mg/mL Wheat Germ Agglutinin, Alexa Fluor 488 Conjugate for budscars and 1 μL of 5 mg/ml propidium iodide for live/dead and waiting 30 minutes at RT, then washing twice with 1X PBS. Flow cytometry data was done by collecting >10k events on the C6. Microscopy was done in the Keck center

on a Leica inverted widefield scope. Bud scars were counted by looking at the images obtained by microscopy.

4.5.11 Cell lysis, protein reduction, alkylation and digestion

Frozen cell pellets were resuspended in lysis buffer composed of 8 M urea, 150 mM NaCl, and 100 mM HEPES pH 8.2. Cells were lysed by 3 cycles of bead beating (30-s beating, 1-min rest) with zirconia/silica beads. Lysate protein concentration was measured by BCA assay. Proteins were reduced with 5 mM dithiothreitol (DTT) for 30 min at 55°C and alkylated with 15 mM iodoacetamide in the dark for 15 min at room temperature. The alkylation reaction was quenched by incubating with additional 5 mM DTT for 15 min at room temperature. Lysate was diluted with 100 mM HEPES pH 8.2 to 1.5 M Urea and Proteins were digested overnight using endopeptidase Lys-C (1:100 enzyme to protein). The next day digests were acidified with TFA and desalted over a C18 sep pak cartdrige and peptides were dried.

4.5.12 TMT labeling

We employed a Tandem Mass Tag (TMT) 10plex systems (Thermo Fisher) to multiplex our samples for identification and quantification by mass spectrometry. We chose a TMT labeling scheme where we distributed the five biological replicates to five different TMTplexes. Each plex contained 6 channels for the different biological conditions and 2 reference channels with pooled samples. The remaining 2 channels were not used. TMT10plex reagents were resuspended and aliquoted as indicated by the manufacturer. Peptides corresponding to 25 μg proteins were resuspended in 40ul of 100mM HEPES pH 8.2, 30% ACN. 25 μl of the corresponding TMT10plex reagent in

100% ACN were used for each labeling reaction and mixed with the peptide sample. Labeling was performed for 1h at room temperature and quenched with 5% hydroxylamine. Samples were pooled and desalted over a MCX SepPak and dried.

4.5.13 Mass spectrometry data acquisition

Dried peptide samples were dissolved in 4% formic acid, 3% acetonitrile and analyzed by an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher) equipped with an Easy1200 nanoLC system (Thermo Fisher). Peptides were loaded onto a 100 µm ID × 3 cm precolumn packed with Reprosil C18 3 µm beads (Dr. Maisch GmbH), and separated by reverse-phase chromatography on a 100 µm ID × 30 cm analytical column packed with Reprosil C18 1.9 µm beads (Dr. Maisch GmbH) and housed into a column heater set at 50°C. Peptides were separated by a 90-min gradient ranging from 10 to 35% acetonitrile in 0.125% formic acid. The instrument method included full MS1 orbitrap, ion trap MS2 and quantitative SPS-MS3 orbitrap scans.

4.5.14 Mass spectrometry data analysis

DDA-MS/MS spectra were searched with Comet(Eng et al., 2013) against the *S. cerevisiae* proteome. Constant modification of cysteine carbamidomethylation (57.021463 Da) and variable modification of methionine oxidation (15.994914 Da) were used for all searches. Search results were filtered to a 1% FDR at PSM level using Percolator (Käll et al., 2007). TMT reporter ion intensities were quantified using in-house software. Further data processing, statistical analysis and visualization were performed in R using the MSstatsTMT package

(https://www.mcponline.org/content/mcprot/early/2020/07/17/mcp.RA120.002105.full.pdf) and tidyverse packages.

4.6 Discussion

In this chapter I describe the implementation of two robust culturing techniques in the lab. The turbidostat was run for ~300 hours without issue, with the main limiting factor being the media usage. I was able to successfully evolve suppressors to a major genome edit which affected growth rate- the rDNA locus deletion. It appears that the turbidostat correctly tracked the increase in size of the rDNA origin which affected growth rate. This is exciting for many reasons. First, with more work I believe we could use this data to calculate the fitness effect of a sweeping mutation. I do believe that the data would be most simple to interpret if the strong selection-weak mutations (SSWM) assumptions are met(reviewed in Orr, 2005). It would be difficult to deconvolute the effect of many smaller mutations since the read out- the amount of media needed to dilute the culture- is on a population scale. If these assumptions were met, we could see in real time a mutation sweeping the population, while at the same time estimate the fitness benefit. This would require additional control experiments with strains having known and calculated fitness differences. We would need to understand how the turbidostat output should be used to calculate the fitness effect as well as the ability to determine the difference between hard and soft sweeps. In addition to calculating the fitness effect of a beneficial mutation arising within an isogenic population, the turbidostat could also be used to determine the fitness effects of gene manipulations in a pooled format. A future use of the turbidostat could be deducing the fitness effect of a

variant in a pool by the effect it has on the overall growth rate. One could potentially determine the fitness effects of many variants at once by comparing their frequency in the pool to the overall growth rate, as long as they were no complex interactions between populations. Once you've achieved a fitness parameter for every variant in your pool, it would also be a natural next step to see if any of them affect the lifespan of yeast. We have successfully implemented the MADs in the lab, and they are available for further experimentation. We have shown that we are capable of growing large quantities of cells- important for accurate coverage of a library and generate enough material to do comprehensive proteomics. It would then be absolutely no stretch to have enough material for comprehensive tracking of variants which have an effect on aging. I believe this would also require some additional control experiments to determine how often one must sample to pick up potentially small differences in life span. However, with 16 parallel magnetic stats set up, the ability to collect many replicates is easy and straightforward. Both of these devices produce interesting and robust data on their own and are very amenable to being paired.

4.7 Acknowledgements

Joseph Sanchez generated the rDNA deletion and GC replacement strains. Bonny Brewer did initial experiments in batch culture with these strains and generated the CHEF gels and southern blots. Chris Large generated the variant analysis pipeline and the scripts to plot copy number. Mario Leutert counted bud scars and completed the proteomic analysis on the aged cells, as well as helped with putting together and

running the MADs. Method sections 4.5.5 -4.5.7 were copied from(Sanchez et al., 2019). Mario Leutert wrote the methods for section 4.5.11-4.5.14.

Chapter 5: Conclusions and future directions

5.1 Multiplexed mutation rate: What we've learned and improvements for future experiments

In the three continuous culture systems developed during my thesis, the types of questions that can be asked are endless. All the methods developed within my thesis are meant to be multi-purpose, and have been built robustly enough to be used in other applications. The multiplexed mutation rate assay I developed is capable of measuring the mutation rate of ~40 variants without the use of a barcode. The limitation is likely the error rate of the polymerase used in Illumina sequencing, and thus new technology is needed to drive down sequencing error rate in order to study more non-barcoded variants at once. In addition, the rise in accuracy and drop in cost of long read sequencing, developed by Pacific Biosciences (Wenger et al., 2019), in combination with the variant analysis pipeline generated by Cindy Yeh, Clara Amorosi, and Soyeon Showman in the Dunham lab, may allow for greater accuracy when using unbarcoded plasmids. A key limitation to using short read Illumina sequencing is that it does not allow for tracking of wt or null sequences, as it uses the mutation within the gene itself to identify the allele. Use of long read sequencing would also allow for the addition of wild type and full null sequences in the unbarcoded pools because full haplotype information would be retained. This would allow for better internal controls in unbarcoded experiments.

In chapter 3, I assayed new variants of Msh2 and increased the capacity of the multiplexed mutation rate assessment assay by barcoding these variants. The addition of barcodes allowed me to assay more variants as the entire haplotype information was encoded in 13 base pairs which was easily covered by a double-ended Illumina sequencing read. However, I saw a similar spread of mutation rate values as with the unbarcoded variants, suggesting it didn't help accuracy. This could be driven by evolutionary forces working on mutations in the background of a lineage, but lineages with large increases or decreases in total frequency of the population were removed from the analysis. This would appear to indicate that mutation rate noise is likely inherent to the biology of mismatch repair, rather than a technical defect of the experimental setup. There has already been evidence using traditional fluctuation assays that this is the case (Gou et al., 2019). However, the multiplexed mutation rate assay took place in an environment in which active evolution could occur, which may affect the mutation rate calculation. If I were to rebuild the mutation rate assay from the ground up, I would consider doing so in a background strain which had already evolved maximal fitness in a specific condition while maintaining the full deletion mutation rate, such as a strain which has already evolved in the turbidostat. This would obviously not eliminate possible epistatic interactions, and it wouldn't prevent evolution from acting on the variants themselves but may be a good starting point.

As an example of this, I saw that the strain in the turbidostat in Fig 4.3.1 appeared to have had a beneficial mutation sweep the population. It would be interesting to isolate a clone (or clones) from the end of that population and continue to

evolve it to see if we can get a longer range of collection before there is evidence of a sweep. Currently we can only collect between 12-45 generations before we see evidence of evolution, and a longer assay period may reduce noise in the mutation rate reading. In chapter 3, I discuss a 5-fold increase in expected mutation rate of wt barcodes in the wt background, and a 10 fold increase in the deletion background. This is really unexpected and may be caused by the downstream analysis. A higher-than-expected mutation rate would indicate that there are variants that contribute to the mutation rate which were not assayed. Individual clone sequencing of these populations doesn't indicate any unbarcoded variants, so as of right now, it is unclear how the analysis might be affecting mutation rate. In addition to the analysis pipeline, the 5-10-fold increase in mutation rate of WT barcodes, may not be a processing quirk, but rather a further indication of actual biology. It would be illuminating to switch to rich medium and run the experiment at the true growth rate of yeast, to see if that has an effect on any of the mutation rate calculations. In order to do this, one would first have to verify that canavanine resistance is neutral in any future conditions, as it would make mutation rate calculations invalid if there was selection on mutants. However, changing the medium in which the assay is completed would potentially provide information on the environmental effect on mutation rate.

5.2 Alternative uses of mixed mutator pools of Msh2 variants

In chapter 3, I generated a pool of Msh2 variants with differing mutation rates which could provide insight into the effects of mutation rate on other processes. One such example is the effect of mutation rate on the evolution of a culture. In chapter 2, I

show a short-term growth defect associated with the *msh2Δ* mutant. However, in the long run the *msh2Δ* mutant may have access to a greater mutational landscape of potential genetic mutations than a wt *MSH2* strain. A previous study on mismatch repair proteins Mlh1 and Pms1 indicates that the variation caused by a temporary increase in mutation rate may be beneficial (Bui et al., 2015). Work in chapter 3 provided mutation rate calculations for all variants in the pool which could be compared to the final evolved population of the pool to determine the effect of mutation rate on evolutionary dynamics of an actively growing mixed culture. To determine the difference between one off and recurrent events, we would need to repeat the evolution many times, in which case perhaps implementing the eVolver (Wong et al., 2018) or other small volume turbidostat may be necessary. With the barcoded Msh2 variant pool generated in chapter 3, we could easily track and determine the impact of clonal interference on an evolving population, and whether a pool of mixed mutators affects evolutionary behavior.

In addition to addressing hypotheses in evolutionary behavior, the mixed mutator pool of Msh2 variants could provide information on the dynamics of anti-microbial resistance. We observed that the mixed population of Msh2 variants gains resistance to canavanine at the median mutation rate of the pool when in a permissive environment. However, an additional line of inquiry could be to ask if the population as a whole developed resistance faster or slower than isogenic populations in sub-minimum inhibitory concentrations. I would expect an isogenic population of a higher mutator strain would develop resistance fastest, however this strain may not be able to achieve the same type of clean sweep because it could be more likely to also gain a detrimental

mutation. In addition to experiments using the developed methods in this thesis , we could add a spatial separation element by using water in oil emulsions (reviewed in Schaerli and Hollfelder, 2009) or extra-large agar plates (Baym et al., 2016) that may significantly change the dynamics of resistance if there isn't direct initial competition between the mutator strains. A pool of mixed mutators would allow for a deeper investigation into how the mutation rate parameter affects evolution, especially since in our system we can keep other parameters, such as population size, constant.

5.3 Suppressor screens and beyond: Uses of the turbidostat

In addition to combining reagents and hypotheses generated during chapters 2,3 and 4, I hope that the turbidostat is used to address questions outside the scope of my thesis. In chapter 4, I describe the implementation of the turbidostat and subsequent use to determine suppressors to reduced rDNA copies by evolution over a 2-week span. The main barrier to running the turbidostat for an extended time frame is the large volumes of media needed, resulting in extensive media preparation time. For long term evolution experiments reducing the culture volume would increase ease of use.

In chapter 4 I observed a correlation between increased rDNA array size and increased growth rate. A natural follow up would be to ask if the increase in the rDNA locus seen in the experiment is just a function of generations, or if population size may play an effect. It is not clear whether reducing the population size would have extended the time to the full increase in size of the rDNA locus. A larger population size does provide a greater set of possible mutations to select upon, but also more clonal interference and competition between lineages. I would like to play with the population

size parameter for experimental evolution experiments, which is simple accomplish in the turbidostat. The turbidostat allows for fine control of growth dynamics, which makes it ideal for studying evolution, but also could be instrumental in screens.

The turbidostat is ideally suited for screening library fitness. As an example of this, I previously attempted to select for the most functional centromeric sequence on a plasmid in order to produce a more faithfully segregating plasmid to use in future experiments. This was a small library containing tiling fragments from the 16 centromeric sequences in yeast. While I was ultimately unsuccessful due to issues with the experimental setup and the downstream analysis, the turbidostat is well-suited to this type of screen. The turbidostat allows the user to fine tune selection over time and can yield a better separation between strains with different fitness values. There are many different variant libraries which may benefit from selection within the turbidostat. First, Soyeon Showman is expanding on work by Elisa Wong, testing a library of CYP2D6 variants, a human liver enzyme important in the drug metabolism. These variants have different growth rates in the presence of tamoxifen, an anti-breast cancer drug (Chan et al., 2020), and a batch culture growth assay has already been developed in the lab, but switching the assay to the tubidostat may provide improved results. The turbidostat could potentially tease out even small differences in function of CYP2D6 variants when run in tamoxifen containing media. The turbidostat contains morbidostat functionality that would allow for the gradual increases in tamoxifen concentration over time, potentially feeding back on growth rate and cumulative resistance. Additionally, Renee Geck is developing an assay in *S. cerevisiae* to determine the effects of variation

in G6PD, an enzyme important in red blood cells and a common Mendelian deficiency in humans. Growth is a read out of function in her assay, which would be a straight forward use of the turbidostat as running pool of variants is what it was initially designed for (McGeachy et al., 2019).

5.4 Pure populations of aged cells and their uses

The implementation of the MAD in collaboration with the Villén lab allowed us to study the proteomic changes associated with age. We found that aged yeast and the daughter cells isolated from aged populations have distinct proteome profiles from young yeast cells. There are many follow-up experiments we are investigating with this proteomics data, as we are currently validating genes of interest found and investigating changes seen in kinase networks using phospho-proteomic data. However, these experiments were all done using wild type FY4 yeast, a common lab strain. In the future, we would like to screen libraries of different strain backgrounds or gene knockouts to investigate genetic background effects. For example, the yeast deletion collection (reviewed in Giaever and Nislow, 2014), is a set of *S. cerevisiae* strains were each strain has a single gene knocked out, would be a prime candidate for the MADs. In the MAD, there is no growth competition because daughters are washed away and cannot compete with the mother cells. There is reduced competition between mother cells once inoculated, as their numbers don't increase over the course of the experiment and there is an environment of excess nutrients. With the MAD, one could directly probe the effect of each non-essential gene on aging. In addition, the telomeric amplicon library (Sunshine et al., 2015a), which is a *S. cerevisiae* library which contains

aneuploidies of all chromosomes in different lengths, would be an obvious choice to directly test the impacts of aneuploidy on aging. The MAD is a multi-use device for generating relatively pure populations of aged cells and is a very powerful method on its own. However, I think the true power of the method development I completed during my thesis is how all of these techniques can be used together.

5.5 Combing continuous culture techniques

The techniques and reagents I developed during my thesis can all be combined for a greater understanding of the effect of mutation rate variation. For example, one could take the pool of differential *MSH2* mutators, determine their fitness in the turbidostat, and compare that to their mutation rate measured using the multiplexed mutation rate assessment in the chemostat. This same pool of variants could be used to investigate the effect of mutation rate on age in the MAD. With these three devices (the chemostat, turbidostat, and MAD), we can determine a population's mutation rate, measure its fitness, and investigate how the population ages. As mentioned previously, determining the effects of gene deletions on age using the yeast deletion collection would generate interesting hypotheses, however additional information could be obtained by assaying their fitness and mutation rate in the turbidostat and with the multiplexed mutation rate protocol in the chemostat. Strains which have neutral fitness in a glucose-limited chemostat have already been determined (Sunshine et al., 2015b) which is the first requirement for the multiplexed mutation rate assessment. Additionally, previous work has looked at the bulk contribution of all deletion collection variants to loss of function mutations in *CAN1* (Huang et al., 2003), however they did not study

mutation rate. Many studies (as reviewed in Delneri, 2011) have looked at the fitness effects of the deletion collection, and it has been competed in the turbidostat (Pir et al., 2012). Ultimately, a combination of all these assays could give us information about a gene's effect on mutation rate, aging, and overall growth dynamics, all in the same strain background. In conclusion, the three assays I implemented during my thesis work well in combination and have already produced genome-wide data on various biological processes in *S. cerevisiae*.

Literature Cited

Abildgaard, A.B., Stein, A., Nielsen, S.V., Schultz-Knudsen, K., Papaleo, E., Shrikhande, A., Hoffmann, E.R., Bernstein, I., Gerdes, A.-M., Takahashi, M., Ishioka, C., Lindorff-Larsen, K., Hartmann-Petersen, R., 2019. Computational and cellular studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome. eLife 8, e49138. https://doi.org/10.7554/eLife.49138

Adams, J., Paquin, C., Oeller, P.W., Lee, L.W., 1985. Physiological characterization of adaptive clones in evolving populations of the yeast, *Saccharomyces cerevisiae*. Genetics 110, 173–185.

Anderson, P.A., 1953. Automatic Recording of the Growth Rates of Continuously Cultured Microorganisms. J. Gen. Physiol. 36, 733–737. https://doi.org/10.1085/jgp.36.6.733

Arduino - Home [WWW Document], n.d. URL https://www.arduino.cc/ (accessed 12.7.20).

Bahadori, M., Azizi, M.H., Dabiri, S., 2018. Recent Advances on Nucleolar Functions in Health and Disease. Arch. Iran. Med. 21, 600–607.

Baym, M., Lieberman, T.D., Kelsic, E.D., Chait, R., Gross, R., Yelin, I., Kishony, R., 2016. Spatiotemporal microbial evolution on antibiotic landscapes. Science 353, 1147–1151. https://doi.org/10.1126/science.aag0822

Blount, Z.D., Lenski, R.E., Losos, J.B., 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362. https://doi.org/10.1126/science.aam5979

Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R., 1987. [10] 5-Fluoroorotic acid as a selective agent in yeast molecular genetics, in: Methods in Enzymology. Elsevier, pp. 164–175. https://doi.org/10.1016/0076-6879(87)54076-9

Boiteux, S., Jinks-Robertson, S., 2013. DNA Repair Mechanisms and the Bypass of DNA Damage in *Saccharomyces cerevisiae* . Genetics 193, 1025–1064. https://doi.org/10.1534/genetics.112.145219

Bouvet, D., Bodo, S., Munier, A., Guillerm, E., Bertrand, R., Colas, C., Duval, A., Coulet, F., Muleris, M., 2019. Methylation Tolerance-Based Functional Assay to Assess Variants of Unknown Significance in the *MLH1* and *MSH2* Genes and Identify Patients With Lynch Syndrome. Gastroenterology 157, 421–431. https://doi.org/10.1053/j.gastro.2019.03.071

Bowen, N., Kolodner, R.D., 2017. Reconstitution of *Saccharomyces cerevisiae* DNA polymerase ε-dependent mismatch repair with purified proteins. Proc. Natl. Acad. Sci. 114, 3607–3612. https://doi.org/10.1073/pnas.1701753114

Bui, D.T., Dine, E., Anderson, J.B., Aquadro, C.F., Alani, E.E., 2015. A Genetic Incompatibility Accelerates Adaptation in Yeast. PLOS Genet. 11, e1005407. https://doi.org/10.1371/journal.pgen.1005407

Chan, C.W.H., Law, B.M.H., So, W.K.W., Chow, K.M., Waye, M.M.Y., 2020. Pharmacogenomics of breast cancer: highlighting CYP2D6 and tamoxifen. J. Cancer Res. Clin. Oncol. 146, 1395–1404. https://doi.org/10.1007/s00432-020-03206-w

Chen, C., Contreras, R., 2007. Identifying genes that extend life span using a high-throughput screening system. Methods Mol. Biol. Clifton NJ 371, 237–248. https://doi.org/10.1007/978-1-59745-361-5_18

Chernoff, Y. o., Vincent, A., Liebman, S. w., 1994. Mutations in eukaryotic 18S ribosomal RNA affect translational fidelity and resistance to aminoglycoside antibiotics. EMBO J. 13, 906–913. https://doi.org/10.1002/j.1460-2075.1994.tb06334.x

Delneri, D., 2011. Competition Experiments Coupled with High-Throughput Analyses for Functional Genomics Studies in Yeast, in: Castrillo, J.I., Oliver, S.G. (Eds.), Yeast Systems Biology: Methods and Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 271–282. https://doi.org/10.1007/978-1-61779-173-4_16

Demogines, A., Wong, A., Aquadro, C., Alani, E., 2008. Incompatibilities Involving Yeast Mismatch Repair Genes: A Role for Genetic Modifiers and Implications for Disease Penetrance and Variation in Genomic Mutation Rates. PLoS Genet. 4, e1000103. https://doi.org/10.1371/journal.pgen.1000103

Denoth Lippuner, A., Julou, T., Barral, Y., 2014. Budding yeast as a model organism to study the effects of age. FEMS Microbiol. Rev. 38, 300–325. https://doi.org/10.1111/1574-6976.12060

Drost, M., Lützen, A., van Hees, S., Ferreira, D., Calléja, F., Zonneveld, J.B.M., Nielsen, F.C., Rasmussen, L.J., de Wind, N., 2013. Genetic screens to identify pathogenic gene variants in the common cancer predisposition Lynch syndrome. Proc. Natl. Acad. Sci. U. S. A. 110, 9403–9408. https://doi.org/10.1073/pnas.1220537110

Drost, M., Tiersma, Y., Thompson, B.A., Frederiksen, J.H., Keijzers, G., Glubb, D., Kathe, S., Osinga, J., Westers, H., Pappas, L., Boucher, K.M., Molenkamp, S., Zonneveld, J.B., van Asperen, C.J., Goldgar, D.E., Wallace, S.S., Sijmons, R.H., Spurdle, A.B., Rasmussen, L.J., Greenblatt, M.S., de Wind, N., Tavtigian, S.V., 2019. A functional assay–based procedure to classify mismatch repair gene variants in Lynch syndrome. Genet. Med. 21, 1486–1496. https://doi.org/10.1038/s41436-018-0372-2

Drost, M., Zonneveld, J. e B.M., van Dijk, L., Morreau, H., Tops, C.M., Vasen, H.F.A., Wijnen, J.T., de Wind, N., 2010. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum. Mutat. 31, 247–253. https://doi.org/10.1002/humu.21180

Drost, M., Zonneveld, J.B.M., Hees, S. van, Rasmussen, L.J., Hofstra, R.M.W., Wind, N. de, 2012. A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. Hum. Mutat. 33, 488–494. https://doi.org/10.1002/humu.22000

Drotschmann, K., Clark, A.B., Kunkel, T.A., 1999a. Mutator phenotypes of common polymorphisms and missense mutations in *MSH2*. Curr. Biol. 9, 907–910. https://doi.org/10.1016/S0960-9822(99)80396-0

Drotschmann, K., Clark, A.B., Tran, H.T., Resnick, M.A., Gordenin, D.A., Kunkel, T.A., 1999b. Mutator phenotypes of yeast strains heterozygous for mutations in the *MSH2* gene. Proc. Natl. Acad. Sci. 96, 2970–2975. https://doi.org/10.1073/pnas.96.6.2970

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Ekkers, D.M., Santos, F.B. dos, Mallon, C.A., Bruggeman, F., Doorn, G.S. van, 2020. The omnistat: A flexible continuous-culture system for prolonged experimental evolution. Methods Ecol. Evol. 11, 932–942. https://doi.org/10.1111/2041-210X.13403

Eng, J.K., Jahan, T.A., Hoopmann, M.R., 2013. Comet: an open-source MS/MS sequence database search tool. Proteomics 13, 22–24. https://doi.org/10.1002/pmic.201200439

Engel, S.R., Dietrich, F.S., Fisk, D.G., Binkley, G., Balakrishnan, R., Costanzo, M.C., Dwight, S.S., Hitz, B.C., Karra, K., Nash, R.S., Weng, S., Wong, E.D., Lloyd, P., Skrzypek, M.S., Miyasato, S.R., Simison, M., Cherry, J.M., 2014. The Reference Genome Sequence of

Saccharomyces cerevisiae: Then and Now. G3 Genes Genomes Genet. 4, 389–398. https://doi.org/10.1534/g3.113.008995
Faust, G.G., Hall, I.M., 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics 30, 2503–2505. https://doi.org/10.1093/bioinformatics/btu314
FelixKrueger, 2019. A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data: FelixKrueger/TrimGalore.
Foster, P.L., 2006. Methods for Determining Spontaneous Mutation Rates. Methods Enzymol. 409, 195–213. https://doi.org/10.1016/S0076-6879(05)09012-9
Fowler, D.M., Araya, C.L., Gerard, W., Fields, S., 2011. Enrich: software for analysis of protein function by enrichment and depletion of variants. Bioinformatics 27, 3430–3431. https://doi.org/10.1093/bioinformatics/btr577
Fox, M.S., 1998. Some Recollections and Reflections on Mutation Rates. Genetics 148, 1415–1418.
Fox, M.S., 1955. Mutation Rates of Bacteria in Steady State Populations. J. Gen. Physiol. 39, 267–278.
Gammie, A.E., Erdeniz, N., Beaver, J., Devlin, B., Nanji, A., Rose, M.D., 2007. Functional Characterization of Pathogenic Human MSH2 Missense Mutations in *Saccharomyces cerevisiae*. Genetics 177, 707–721. https://doi.org/10.1534/genetics.107.071084
Ganley, A.R.D., Ide, S., Saka, K., Kobayashi, T., 2009. The Effect of Replication Initiation on Gene Amplification in the rDNA and Its Relationship to Aging. Mol. Cell 35, 683–693. https://doi.org/10.1016/j.molcel.2009.07.012
Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio.
Giaever, G., Nislow, C., 2014. The Yeast Deletion Collection: A Decade of Functional Genomics. Genetics 197, 451–465. https://doi.org/10.1534/genetics.114.161620
Gordon, A.S., Rosenthal, E.A., Carrell, D.S., Amendola, L.M., Dorschner, M.O., Scrol, A., Stanaway, I.B., DeVange, S., Ralston, J.D., Zouk, H., Rehm, H.L., Larson, E., Crosslin, D.R., Leppig, K.A., Jarvik, G.P., 2019. Rates of Actionable Genetic Findings in Individuals with Colorectal Cancer or Polyps Ascertained from a Community Medical Setting. Am. J. Hum. Genet. 105, 526–533. https://doi.org/10.1016/j.ajhg.2019.07.012
Gou, L., Bloom, J.S., Kruglyak, L., 2019. The Genetic Basis of Mutation Rate Variation in Yeast. Genetics 211, 731–740. https://doi.org/10.1534/genetics.118.301609
Gramelsberger, G., 2018. Continuous culture techniques as simulators for standard cells: Jacques Monod's, Aron Novick's and Leo Szilard's quantitative approach to microbiology. Hist. Philos. Life Sci. 40, 23. https://doi.org/10.1007/s40656-017-0182-x
Gresham, D., Dunham, M.J., 2014. The enduring utility of continuous culturing in experimental evolution. Genomics 104, 399–405. https://doi.org/10.1016/j.ygeno.2014.09.015
Gupta, S., Provenzale, D., Llor, X., Halverson, A.L., Grady, W., Chung, D.C., Haraldsdottir, S., Markowitz, A.J., Slavin, T.P., Hampel, H., CGC, Ness, R.M., Weiss, J.M., Ahnen, D.J., Chen, L.-M., Cooper, G., Early, D.S., Giardiello, F.M., Hall, M.J., Hamilton, S.R., Kanth, P., Klapman, J.B., Lazenby, A.J., Lynch, P.M., Mayer, R.J., Mikkelson, J., CGC, Peter, S., Regenbogen, S.E., Dwyer, M.A., CGC, Ogba, N., 2019. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 2.2019. J. Natl. Compr. Cancer Netw. JNCCN 17, 1032–1041. https://doi.org/10.6004/jnccn.2019.0044

Hendrickson, D.G., Soifer, I., Wranik, B.J., Kim, G., Robles, M., Gibney, P.A., McIsaac, R.S., 2018. A new experimental platform facilitates assessment of the transcriptional and chromatin landscapes of aging yeast. eLife 7, e39911. https://doi.org/10.7554/eLife.39911

Hoffman, C.S., Winston, F., 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. Gene 57, 267–272. https://doi.org/10.1016/0378-1119(87)90131-4

Hope, E.A., Amorosi, C.J., Miller, A.W., Dang, K., Heil, C.S., Dunham, M.J., 2017. Experimental Evolution Reveals Favored Adaptive Routes to Cell Aggregation in Yeast. Genetics 206, 1153–1167. https://doi.org/10.1534/genetics.116.198895

Hope, E.A., Dunham, M.J., 2014. Ploidy-Regulated Variation in Biofilm-Related Phenotypes in Natural Isolates of Saccharomyces cerevisiae. G3 Genes Genomes Genet. 4, 1773–1786. https://doi.org/10.1534/g3.114.013250

Houlleberghs, H., Goverde, A., Lusseveld, J., Dekker, M., Bruno, M.J., Menko, F.H., Mensenkamp, A.R., Spaander, M.C.W., Wagner, A., Hofstra, R.M.W., Riele, H. te, 2017. Suspected Lynch syndrome associated *MSH6* variants: A functional assay to determine their pathogenicity. PLOS Genet. 13, e1006765. https://doi.org/10.1371/journal.pgen.1006765

Huang, M.-E., Rio, A.-G., Nicolas, A., Kolodner, R.D., 2003. A genomewide screen in Saccharomyces cerevisiae for genes that suppress the accumulation of mutations. Proc. Natl. Acad. Sci. 100, 11529–11534. https://doi.org/10.1073/pnas.2035018100

Janssens, G.E., Meinema, A.C., González, J., Wolters, J.C., Schmidt, A., Guryev, V., Bischoff, R., Wit, E.C., Veenhoff, L.M., Heinemann, M., 2015. Protein biogenesis machinery is a driver of replicative aging in yeast. eLife 4, e08527. https://doi.org/10.7554/eLife.08527

Jia, X., Burugula, B.B., Chen, V., Lemons, R.M., Jayakody, S., Maksutova, M., Kitzman, J.O., 2020. Massively parallel functional testing of *MSH2* missense variants conferring Lynch Syndrome risk. bioRxiv 2020.06.03.133017. https://doi.org/10.1101/2020.06.03.133017

Johnson, R.E., Kovvali, G.K., Prakash, L., Prakash, S., 1996. Requirement of the Yeast MSH3 and MSH6 Genes for MSH2-dependent Genomic Stability. J. Biol. Chem. 271, 7285–7288. https://doi.org/10.1074/jbc.271.13.7285

Kadyrov, F.A., Dzantiev, L., Constantin, N., Modrich, P., 2006. Endonucleolytic Function of MutLα in Human Mismatch Repair. Cell 126, 297–308. https://doi.org/10.1016/j.cell.2006.05.039

Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., Modrich, P., 2007. Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease. J. Biol. Chem. 282, 37181–37190. https://doi.org/10.1074/jbc.M707617200

Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., MacCoss, M.J., 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods 4, 923–925. https://doi.org/10.1038/nmeth1113

Kennedy, B., Austriaco Jr., N.R., Guarente, L., 1994. Daughter cells of Saccharomyces cerevisiae from old mothers display a reduced life span. J. Cell Biol. 127, 1985–1993.

Kubitschek, H.E., Bendigkeit, H.E., 1964. Mutation in continuous cultures. I. Dependence of mutational response upon growth-limiting factors. Mutat. Res. Mol. Mech. Mutagen. 1, 113–120. https://doi.org/10.1016/0027-5107(64)90013-2

Kunkel, T.A., Erie, D.A., 2005. Dna Mismatch Repair. Annu. Rev. Biochem. 74, 681–710. https://doi.org/10.1146/annurev.biochem.74.082803.133243
Kwan, E.X., Wang, X.S., Amemiya, H.M., Brewer, B.J., Raghuraman, M.K., 2016. rDNA Copy Number Variants Are Frequent Passenger Mutations in Saccharomyces cerevisiae Deletion Collections and de Novo Transformants. G3 Genes Genomes Genet. 6, 2829–2838. https://doi.org/10.1534/g3.116.030296
Lang, G.I., Murray, A.W., 2008. Estimating the Per-Base-Pair Mutation Rate in the Yeast Saccharomyces cerevisiae. Genetics 178, 67–82. https://doi.org/10.1534/genetics.107.071506
Lang, G.I., Parsons, L., Gammie, A.E., 2013. Mutation Rates, Spectra, and Genome-Wide Distribution of Spontaneous Mutations in Mismatch Repair Deficient Yeast. G3 Genes Genomes Genet. 3, 1453–1465. https://doi.org/10.1534/g3.113.006429
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923
Leutert, M., Rodríguez-Mias, R.A., Fukuda, N.K., Villén, J., 2019. R2-P2 rapid-robotic phosphoproteomics enables multidimensional cell signaling studies. Mol. Syst. Biol. 15. https://doi.org/10.15252/msb.20199021
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Li, S., Qian, D., Thompson, B.A., Gutierrez, S., Wu, S., Pesaran, T., LaDuca, H., Lu, H.-M., Chao, E.C., Black, M.H., 2020. Tumour characteristics provide evidence for germline mismatch repair missense variant pathogenicity. J. Med. Genet. 57, 62–69. https://doi.org/10.1136/jmedgenet-2019-106096
Lindstrom, D.L., Gottschling, D.E., 2009. The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae. Genetics 183, 413–422, 1SI-13SI. https://doi.org/10.1534/genetics.109.106229
Longo, V.D., Shadel, G.S., Kaeberlein, M., Kennedy, B., 2012. Replicative and Chronological Aging in Saccharomyces cerevisiae. Cell Metab. 16, 18–31. https://doi.org/10.1016/j.cmet.2012.06.002
Luria, S.E., Delbrück, M., 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28, 491–511.
Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., Hitchins, M.P., 2015. Milestones of Lynch syndrome: 1895-2015. Nat. Rev. Cancer 15, 181–194. https://doi.org/10.1038/nrc3878
Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C.R., Dopman, E.B., Dickinson, W.J., Okamoto, K., Kulkarni, S., Hartl, D.L., Thomas, W.K., 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. 105, 9272–9277. https://doi.org/10.1073/pnas.0803466105
Marsischky, G.T., Filosi, N., Kane, M.F., Kolodner, R., 1996. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev. 10, 407–420. https://doi.org/10.1101/gad.10.4.407

Martinez, S.L., Kolodner, R.D., 2010. Functional analysis of human mismatch repair gene mutations identifies weak alleles and polymorphisms capable of polygenic interactions. Proc. Natl. Acad. Sci. 107, 5070–5075. https://doi.org/10.1073/pnas.1000798107
McGeachy, A.M., Meacham, Z.A., Ingolia, N.T., 2019. An Accessible Continuous-Culture Turbidostat for Pooled Analysis of Complex Libraries. ACS Synth. Biol. 8, 844–856. https://doi.org/10.1021/acssynbio.8b00529
Mendillo, M.L., Hargreaves, V.V., Jamison, J.W., Mo, A.O., Li, S., Putnam, C.D., Woods, V.L., Kolodner, R.D., 2009. A conserved MutS homolog connector domain interface interacts with MutL homologs. Proc. Natl. Acad. Sci. 106, 22223–22228. https://doi.org/10.1073/pnas.0912250106
Miles, A., 2019. A fast Python and command-line utility for extracting simple statistics against genome positions based on sequence alignments from a SAM or BAM file.: alimanfoo/pysamstats.
Miller, A.W., Befort, C., Kerr, E.O., Dunham, M.J., 2013. Design and use of multiplexed chemostat arrays. J. Vis. Exp. JoVE e50262. https://doi.org/10.3791/50262
Monod, J., 1950. Selected Papers in Molecular Biology by Jacques Monod. Ann. Inst. Pasteur Paris 390–410.
Mortimer, R.K., Johnston, J.R., 1959. Life Span of Individual Yeast Cells. Nature 183, 1751–1752. https://doi.org/10.1038/1831751a0
Müller, I., 1985. Parental age and the life-span of zygotes of Saccharomyces cerevisiae. Antonie Van Leeuwenhoek 51, 1–10. https://doi.org/10.1007/BF00444223
Myers, J., Clark, L.B., 1944. Culture Conditions and the Development of the Photosynthetic Mechanism. J. Gen. Physiol. 28, 103–112.
Nielsen, S.V., Stein, A., Dinitzen, A.B., Papaleo, E., Tatham, M.H., Poulsen, E.G., Kassem, M.M., Rasmussen, L.J., Lindorff-Larsen, K., Hartmann-Petersen, R., 2017. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. PLOS Genet. 13, e1006739. https://doi.org/10.1371/journal.pgen.1006739
Novick, A., Szilard, L., 1951. Experiments on Spontaneous and Chemically Induced Mutations of Bacteria Growing in the Chemostat. Cold Spring Harb. Symp. Quant. Biol. 16, 337–343. https://doi.org/10.1101/SQB.1951.016.01.025
Novick, A., Szilard, L., 1950a. Description of the Chemostat. Science 112, 715–716.
Novick, A., Szilard, L., 1950b. Experiments with the chemostat on spontaneous mutations of bacteria. Proc. Natl. Acad. Sci. 36, 708–719.
Orr, H.A., 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6, 119–127. https://doi.org/10.1038/nrg1523
Paquin, C., Adams, J., 1983. Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. Nature 302, 495–500. https://doi.org/10.1038/302495a0
Paquin, C.E., Adams, J., 1983. Relative fitness can decrease in evolving asexual populations of <i>S. cerevisiae <i/>. Nature 306, 368–370.
Pashkova, N., Gakhar, L., Winistorfer, S.C., Sunshine, A.B., Rich, M., Dunham, M.J., Yu, L., Piper, R.C., 2013. The Yeast Alix Homolog Bro1 Functions as a Ubiquitin Receptor for Protein Sorting into Multivesicular Endosomes. Dev. Cell 25, 520–533. https://doi.org/10.1016/j.devcel.2013.04.007

Pavlov, Y.I., Mian, I.M., Kunkel, T.A., 2003. Evidence for Preferential Mismatch Repair of Lagging Strand DNA Replication Errors in Yeast. Curr. Biol. 13, 744–748. https://doi.org/10.1016/S0960-9822(03)00284-7
Peltomäki, P., 2016. Update on Lynch syndrome genomics. Fam. Cancer 15, 385–393. https://doi.org/10.1007/s10689-016-9882-8
Pir, P., Gutteridge, A., Wu, J., Rash, B., Kell, D.B., Zhang, N., Oliver, S.G., 2012. The genetic control of growth rate: a systems biology study in yeast. BMC Syst. Biol. 6, 4. https://doi.org/10.1186/1752-0509-6-4
Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A., Modrich, P., 2010. PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc. Natl. Acad. Sci. 107, 16066–16071. https://doi.org/10.1073/pnas.1010662107
Pronobis, M.I., Deuitch, N., Peifer, M., 2016. The Miraprep: A Protocol that Uses a Miniprep Kit and Provides Maxiprep Yields. PLOS ONE 11, e0160509. https://doi.org/10.1371/journal.pone.0160509
Rañola, J.M.O., Liu, Q., Rosenthal, E.A., Shirts, B.H., 2018. A comparison of cosegregation analysis methods for the clinical setting. Fam. Cancer 17, 295–302. https://doi.org/10.1007/s10689-017-0017-7
Rath, A., Mishra, A., Ferreira, V.D., Hu, C., Omerza, G., Kelly, K., Hesse, A., Reddi, H.V., Grady, J.P., Heinen, C.D., 2019. Functional interrogation of Lynch syndrome-associated MSH2 missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. Hum. Mutat. 40, 2044–2056. https://doi.org/10.1002/humu.23848
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–423. https://doi.org/10.1038/gim.2015.30
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754
Sanchez, J.C., Kwan, E.X., Pohl, T.J., Amemiya, H.M., Raghuraman, M.K., Brewer, B.J., 2017. Defective replication initiation results in locus specific chromosome breakage and a ribosomal RNA deficiency in yeast. PLOS Genet. 13, e1007041. https://doi.org/10.1371/journal.pgen.1007041
Sanchez, J.C., Ollodart, A., Large, C.R.L., Clough, C., Alvino, G.M., Tsuchiya, M., Crane, M., Kwan, E.X., Kaeberlein, M., Dunham, M.J., Raghuraman, M.K., Brewer, B.J., 2019. Phenotypic and Genotypic Consequences of CRISPR/Cas9 Editing of the Replication Origins in the rDNA of Saccharomyces cerevisiae. Genetics 213, 229–249. https://doi.org/10.1534/genetics.119.302351
Schaerli, Y., Hollfelder, F., 2009. The potential of microfluidic water-in-oil droplets in experimental biology. Mol. Biosyst. 5, 1392–1404. https://doi.org/10.1039/B907578J

Shcheprova, Z., Baldi, S., Frei, S.B., Gonnet, G., Barral, Y., 2008. A mechanism for asymmetric segregation of age during yeast budding. Nature 454, 728–734. https://doi.org/10.1038/nature07212

Shirts, B.H., Konnick, E.Q., Upham, S., Walsh, T., Ranola, J.M.O., Jacobson, A.L., King, M.-C., Pearlman, R., Hampel, H., Pritchard, C.C., 2018. Using Somatic Mutations from Tumors to Classify Variants in Mismatch Repair Genes. Am. J. Hum. Genet. 103, 19–29. https://doi.org/10.1016/j.ajhg.2018.05.001

Shor, E., Schuyler, J., Perlin, D.S., 2019. A Novel, Drug Resistance-Independent, Fluorescence-Based Approach To Measure Mutation Rates in Microbial Pathogens. mBio 10. https://doi.org/10.1128/mBio.00120-19

Sims, A.L., Jordan, R.C., 1942. An Accurate Automatic Syringe Mechanism. J. Sci. Instrum. 19, 58–61. https://doi.org/10.1088/0950-7671/19/4/303

Sims, A.L., Jordan, R.C., 1941. A simple bacterial colony counting device. J. Sci. Instrum. 18, 243–243. https://doi.org/10.1088/0950-7671/18/12/407

Smeal, T., Claus, J., Kennedy, B., Cole, F., Guarente, L., 1996. Loss of Transcriptional Silencing Causes Sterility in Old Mother Cells of S. cerevisiae. Cell 84, 633–642. https://doi.org/10.1016/S0092-8674(00)81038-7

Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., Fowler, D.M., 2017. Variant Interpretation: Functional Assays to the Rescue. Am. J. Hum. Genet. 101, 315–325. https://doi.org/10.1016/j.ajhg.2017.07.014

Stein, A., Fowler, D.M., Hartmann-Petersen, R., Lindorff-Larsen, K., 2019. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. Trends Biochem. Sci. 44, 575–588. https://doi.org/10.1016/j.tibs.2019.01.003

Strand, M., Earley, M.C., Crouse, G.F., Petes, T.D., 1995. Mutations in the MSH3 gene preferentially lead to deletions within tracts of simple repetitive DNA in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 92, 10418–10421. https://doi.org/10.1073/pnas.92.22.10418

Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D., 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365, 274–276. https://doi.org/10.1038/365274a0

Sunshine, A.B., Payen, C., Ong, G.T., Liachko, I., Tan, K.M., Dunham, M.J., 2015a. The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects. PLoS Biol. 13. https://doi.org/10.1371/journal.pbio.1002155

Sunshine, A.B., Payen, C., Ong, G.T., Liachko, I., Tan, K.M., Dunham, M.J., 2015b. The fitness consequences of aneuploidy are driven by condition-dependent gene effects. PLoS Biol. 13, e1002155. https://doi.org/10.1371/journal.pbio.1002155

Takahashi, C.N., Miller, A.W., Ekness, F., Dunham, M.J., Klavins, E., 2015. A low cost, customizable turbidostat for use in synthetic circuit characterization. ACS Synth. Biol. 4, 32–38. https://doi.org/10.1021/sb500165g

Thayer, N.H., Leverich, C.K., Fitzgibbon, M.P., Nelson, Z.W., Henderson, K.A., Gafken, P.R., Hsu, J.J., Gottschling, D.E., 2014. Identification of long-lived proteins retained in cells undergoing repeated asymmetric divisions. Proc. Natl. Acad. Sci. U. S. A. 111, 14019–14026. https://doi.org/10.1073/pnas.1416079111

The Nobel Prize in Chemistry 2015 [WWW Document], n.d. . NobelPrize.org. URL https://www.nobelprize.org/prizes/chemistry/2015/modrich/lecture/ (accessed 11.30.20).

Thompson, B.A., Goldgar, D.E., Paterson, C., Clendenning, M., Walters, R., Arnold, S., Parsons, M.T., Michael D, W., Gallinger, S., Haile, R.W., Hopper, J.L., Jenkins, M.A., Lemarchand, L., Lindor, N.M., Newcomb, P.A., Thibodeau, S.N., Colon Cancer Family Registry, Young, J.P., Buchanan, D.D., Tavtigian, S.V., Spurdle, A.B., 2013. A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the Colon Cancer Family Registry. Hum. Mutat. 34, 200–209. https://doi.org/10.1002/humu.22213

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. 37, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Whelan, W.L., Gocke, E., Manney, T.R., 1979. The *CAN1* Locus of *Saccharomyces cerevisiae*: Fine-Structure Analysis and Forward Mutation Rates. Genetics 91, 35–51.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis, Use R! Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-98141-3

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., Nagarajan, N., 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40, 11189–11201. https://doi.org/10.1093/nar/gks918

Wong, B.G., Mancuso, C.P., Kiriakov, S., Bashor, C.J., Khalil, A.S., 2018. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat. Biotechnol. 36, 614–623. https://doi.org/10.1038/nbt.4151

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinforma. Oxf. Engl. 30, 614–620. https://doi.org/10.1093/bioinformatics/btt593