Systems Genetics of DNA Damage Tolerance

Cisplatin, *RAD5* & CRISPR-mediated Nonsense

Eric Bryant

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

Systems Genetics of DNA Damage Tolerance – Cisplatin, *RAD5* & CRISPR-mediated Nonsense

Eric Bryant

DNA sequence information is constantly threatened by damage. In the clinic, intentional DNA damage is often used to treat cancer. Cisplatin, a first-line chemotherapy used to treat millions of patients, functions specifically by generating physical links within DNA strands, blocking DNA replication, and killing dividing cells. To maintain genome integrity, organisms have evolved the capacity to repair, respond, or otherwise resist change to the DNA sequence through a network of genetically encoded DNA damage tolerance pathways. In chapter 1, I present advances in experimental design and current progress for a systems genetics approach, using *Saccharomyces cerevisiae*, to reveal relationships between cisplatin tolerance pathways. Additionally, recent efforts to sequence thousands of cancer genomes have revealed recurrent genetic changes that cause overexpression of specific cisplatin tolerance genes. In chapter 2, I present a submitted manuscript that models overexpression of an essential cisplatin tolerance gene. This study uses a systems genetics approach to reveal the genetic pathways that are essential for tolerating this perturbation, which ultimately led to mechanistic insights for this gene. Convenient genome engineering in *Saccharomyces* has made this organism an ideal model to develop systems genetics concepts and approaches. In chapter 3, I present a published manuscript that demonstrates a new approach to disrupting genes by making site-specific nonsense mutations. Importantly, this approach does not require cytotoxic double-strand DNA breaks and is applicable to many model organisms for disrupting almost any gene, which may advance systems genetics into new model organisms. Systems genetics provides a framework for determining how DNA damage tolerance pathways act together to maintain cellular fitness and genome integrity. Such insights may one day help clinicians predict which cancers will respond to treatment, potentially sparing patients from unnecessary chemotherapy.

Contents

List of Figures

List of Tables

Acknowledgments

My PhD would not have been possible without the support, guidance, advice, collaborative spirit, and hard work of my mentors, students, collaborators, friends and family. I would first like to thank the many students I have had the opportunity to train and to learn from. This includes Alexandra Holden and Ryeen Islam who helped me with my first experiments in the Rothstein lab. Additionally, I would like to thank Saskia Craddock and Eva Hale for their great help at the bench. I would especially like to thank Isabel Sollohub for her great work on advancing a project on *CTF4* overexpression, and Emanuelle Rizk & Eden David for their hard work in constructing and validating strains for the cisplatin epistasis project. I would also like to thank the Columbia Biological Sciences PhD and Genetics & Development PhD rotation students that have helped me over the years. These students include, Maryam Mudasir, Divya Venkatesh, Laura Crowley, Zhengshan Hu and Tyler Perdue. I am also grateful for having the honor to work with a series of talented French masters students – Antoine Canat, François Dossin, Matthieu Haudiquet and Alice Serafin – who each contributed an incredible amount of time and effort to the cisplatin epistasis mapping project. Nearly all of the data presented for the cisplatin epistasis map was generated specifically by François Dossin, Matthieu Haudiquet, Alice Serafin, and Tyler Perdue. Much of my experimental design concepts for this project (i.e. designing the library, and optimizing drug concentrations) were developed through thoughtful discussions with François and Matthieu. In the Rothstein lab, I had the opportunity to learn from many outstanding visiting scientists, especially Luke Berchowitz, Peter Philippsen and Serge Gangloff. The members of the Rothstein lab – John Dittmar, Fraulin Joseph, So Jung Lee, Ivana Šunjevarić, Bob Reid, Michael Smith and Olga Marte – have been an amazing team to work with during my PhD, always offering help and kind advice without hesitation.

I would like to thank my co-authors on the *RAD5* overexpression project – Ivana Šunjevarić and Bob Reid. Both Ivana and Bob have played pivotal roles in conceiving, designing and performing many of the experiments presented in this chapter. For this publication, we thank Lorraine Symington, Luke Berchowitz, Alberto Ciccia and Chao Lu for critical reading of the manuscript.

I would also like to thank my co-authors on the CRISPR-mediated nonsense (iSTOP) publication – Pierre Billon, Sarah Joseph, Tarun Nambiar, Samuel Hayward and Alberto Ciccia. Specifically, I would like to thank Pierre and Alberto for sharing their great ideas, inviting me to collaborate on this project and offering me a co-first author role on the subsequent publication. For this publication, we thank Yannick Doyon for critical reading of the manuscript and for kindly communicating unpublished data. We also thank David Liu, Keith Joung and Feng Zhang for sharing reagents.

I would like to thank my thesis committee members, Songtao Jia and Elizabeth Miller, for their guidance and advice, as well as, Matthew Weitzman, for joining my defense committee. I would also like to thank my co-mentor, Alberto Ciccia, for supporting my growth as a scientist by inviting me to collaborate on several exciting projects in his lab. And of course, I would like to thank my primary mentor, Rodney Rothstein, who has given me the freedom to pursue my scientific interests. Importantly, Rodney has reminded me that, while science is an exciting and rewarding career worth much effort and sacrifice, family should always take priority in a well-balanced life. On that note, I would like to thank my family, friends and especially my wife, Larkin Mohn, for supporting me through this long and winding journey to become a scientist.

My PhD work was supported by the Biological Sciences department of Columbia University's Graduate School of Arts and Sciences. I received additional funding from three National Institute of Health (NIH) training grants in the Biological Sciences (T32GM008798), Cancer Biology (T32CA009503) and Precision Medicine & Translational Research (TL1TR001875). The projects presented in this thesis were also supported by NIH grants R01CA197774 & R01GM117064 (to Alberto Ciccia), and R35GM118180 & U54CA209997 (to Rodney Rothstein). Specific contributions from students and collaborators are indicated in figure legends. Cover art was generated from the genetic landscape of *Saccharomyces cerevisiae* using the rayshader R package (Morgan-Wall 2018). All computational work in my thesis was made possible by the open-source software community, especially contributors to R, Bioconductor, and the tidyverse (R Core Team 2018; Tenenbaum and Team 2017; Wickham 2017).

Dedication

For Larkin, with love.

Preface

Deoxyribonucleic acid (DNA) holds the instructions for life on Earth. Its code consists of only four letters – Adenine (A), Guanine (G), Thymine (T) and Cytosine (C) – the sequence of which determines ribonucleic acids (RNA) to be transcribed, and proteins to be translated in living cells (Dahm 2005; Avery et al. 1944; Crick et al. 1961; Nirenberg et al. 1966). Transcribing DNA into RNA and translating RNA into protein forms the foundation for the central dogma of molecular biology (Crick 1970), which emphasizes an important feature of DNA – its sequence can store the information necessary to reproduce the chemistry of life. This feature, along with fortuitous structural properties (Watson and Crick 1953), makes DNA a molecule that can not only reproduce the chemistry of life, but can also be reliably copied and passed to the next generation. However, to reliably and faithfully pass DNA to the next generation, cells must overcome a constant barrage of DNA damage (Lindahl 1993). Fortunately, the genomes of most cells contain instructions to build a large number of proteins that can work together to repair broken DNA, or enable DNA damage tolerance (Cleaver 2016). This genetically encoded capacity for cells to maintain genome integrity and viability in the face of damaged DNA is the focus of my thesis. In particular, I am interested in understanding how cells resist treatment to the DNA damaging agent, cisplatin – one of several platinum-based chemotherapeutic drugs that have been used to treat cancer in millions of patients (Galanski et al. 2005; Simon et al. 2000; Wu et al. 2004; Kelland 2007).

Decades of work by geneticists have revealed DNA damage tolerance to consist of hundreds of genes that operate in numerous pathways that collectively enable cells to resist DNA lesions caused by cisplatin treatment (reviewed by Ciccia and Elledge 2010; Heyer et al. 2010; Rocha et al. 2018). While the mechanisms for many of the major pathways responsible for cisplatin tolerance are well established, how these genetic pathways relate to each other as a system to maintain fitness and DNA integrity when responding to cisplatin treatment remains poorly understood. Moreover, many individual genes known to affect cellular fitness of cisplatin treated cells have no known role in any of the major cisplatin tolerance pathways. And,

while much is known about how loss of individual gene function can affect cisplatin tolerance, recent efforts to genomically characterize thousands of tumors has revealed recurrent DNA copy number amplification events that can result in overexpression of specific genes (Cancer Genome Atlas Research Network 2008). The genomic repercussions of gene overexpression, and how this could alter cisplatin tolerance, remains a largely unexplored area in genetics.

The work described here attempts to address some of these unexplored and poorly understood aspects of DNA damage tolerance using systems genetics approaches. **Chapter 1**, describes the experimental design and current results for an ongoing collaborative effort to map all cisplatin specific DNA damage tolerance pathway relationships by systematically measure fitness of double-mutants to identify genetic interactions (i.e. double-mutant combinations that cause an unexpected change in fitness). **Chapter 2**, is a submitted manuscript that utilizes a genetic interaction profile to reveal the genomic repercussions of *RAD5* overexpression, one of the most important cisplatin tolerance genes, whose human homolog is frequently overexpressed in specific cancers. **Chapter 3**, is a published manuscript that describes a new approach to efficiently disrupt the function of genes, which will open new avenues for studying the systems genetics of DNA damage tolerance in a diverse range of organisms. An introduction to the scientific and thesis context for each chapter is provided below in **Sections 0.1, 0.2 & 0.3**. Ultimately, understanding the fundamental genetic determinants of cisplatin tolerance may one day help clinicians predict which cancers will respond to cisplatin treatment, thus informing treatment decisions, and sparing patients from unnecessary chemotherapy (O'Connor 2015).

0.1 Epistasis and cisplatin tolerance

The field of genetics seeks to answer one simple question – how does DNA sequence, combined with environmental exposure, result in the observable features of an organism? In other words, how does genotype, plus environment, produce phenotype? Historically, geneticists have approached the "genotype to phenotype" question systematically using genetic screens. In such screens, changes to an organism's DNA sequence would be made randomly using some kind of DNA damaging agent to produce mutants. Mu-

tants that gained the phenotype of interest would be cataloged, and the tedious process of characterizing each mutant allele to identify the causative DNA sequence change would commence. The classic genetic screen began to lose prominence with the advent of high-throughput sequencing technologies that allowed scientists to produce the first complete sequence assemblies for the human genome and several model organisms (Lander et al. 2001). Having sequenced a genome, scientists could computationally predict all of an organism's protein coding genes – no known phenotype was necessary. Researchers could now perform their genetic screens in reverse. Instead of asking, "given a phenotype, what are the alleles?" scientists could now ask, "given the alleles, what are the phenotypes?" Thus, the reverse genetic screen was born.

Geneticists studying the budding yeast, *Saccharomyces cerevisiae*, were early adopters of the reverse genetic screen, which first required the construction of a large collection of defined alleles. Only a few short years after sequencing the yeast genome, every gene in yeast had been individually deleted to produce the non-essential gene disruption collection (Winzeler et al. 1999; Giaever et al. 2002). Remarkably, of the 6,000 protein coding genes in the yeast genome, approximately 5,000 were *not* essential for the organism to survive, which left a big open question – are these genes important for anything? The reverse genetic screen was the perfect tool for quickly answering this question. After exposing this strain collection to over 350 different chemicals, a group of researchers reassuringly found that indeed these genes were important, as 97% of genes showed a growth defect, or advantage, in at least one treatment condition (Hillenmeyer et al. 2008). This reverse genetic screening strategy, known as chemogenomic profiling, when used with a comprehensive collection of gene deletions, produces a list, or "profile", of genes that contribute to fitness when treated with a particular chemical. Importantly, information gained from chemogenomic profiling is a two-way street; as we learn more about gene function, a chemogenomic profile can tell us something about the biological impact of the chemical or treatment itself. For example, early genetic screens in yeast from the 1960s and 1970s had already identified genes that are essential to tolerate exposure to ultraviolet, x-ray, and gamma radiation (Snow 1967; Cox and Parry 1968; Lemontt 1971; Game and Mortimer 1974). These genes were given names starting with "*RAD*" and given a number (e.g. *RAD3*, *RAD6*, *RAD52*, etc.). Later work had determined that many of the *RAD* genes produced proteins that performed various DNA repair processes to combat DNA damage (Haynes and Kunz 1981). With this knowledge of *RAD* gene function

in mind, when we generate a new chemogenomic profile for an uncharacterized compound and find that its profile includes many *RAD* genes, we can infer that this new compound likely causes DNA damage. Thus, the more knowledge we have of gene function, the more powerful chemogenomic profiling becomes.

Cisplatin, being a critically important and clinically relevant drug, received plenty of attention during the heyday of chemogenomic profiling (Birrell et al. 2002; Wu et al. 2004; Lee et al. 2005; Huang et al. 2005; Hastie et al. 2006; Liao et al. 2007). However, while chemogenomic profiling has identified over 100 genes in yeast that are important for cisplatin tolerance, this list of genes does not tell us the biological process, or pathway, in which each gene participates. Moreover, a chemogenomic profile alone cannot tell us how the collection of pathways that affect fitness in response to cisplatin relate to each other as a complete system. For this, we leverage the power of double-mutants to identify which genes have shared contributions, or compensatory contributions to cisplatin sensitivity. This strategy was elegantly used by Brendel and Haynes (1973) to group genes known at the time to be involved in DNA repair into three major pathways, later defined as the *RAD3*, *RAD6* and *RAD52* epistasis groups. An epistasis group, in this context, refers to a set of mutations that share a contribution to the phenotype of DNA damage tolerance and can thus be grouped into shared damage tolerance pathways. Epistasis itself is a class of genetic interaction, a relationship between two genetic perturbations, that geneticists use to determine pathway membership. In the context of fitness, epistasis is most often observed when two deleterious genetic perturbations result in better fitness than expected when the two perturbations are combined. Thus, for clarity, I often describe such epistasis as a positive fitness interaction. In addition to identifying positive fitness interactions, double-mutant analysis can also identify negative fitness interactions – pairs of mutations that cause worse than expected fitness and indicate that each gene contributes to fitness in parallel compensatory pathways.

The major limitation to double-mutant analysis is time and energy. For example, in yeast, with a collection of 5,000 non-essential gene deletions, a researcher would have to perform approximately 12.5 million crosses to identify all possible fitness interactions among these gene deletions for a single treatment condition. Remarkably, after developing the Synthetic Genetic Array (SGA) method, researchers managed to use cleverness and robotics to produce a genetic interaction profile for nearly every gene deletion in budding

yeast (Tong et al. 2001; Tong and Boone 2006; Pan et al. 2006; Costanzo et al. 2010; Baryshnikova et al. 2010). The combined interaction profiles generate an interaction network that represents the genetic landscape of *Saccharomyces cerevisiae* (Usaj et al. 2017; Costanzo et al. 2016).

Unfortunately, this genetic landscape only reveals genetic relationships for pathways that contribute to fitness in standard laboratory conditions. To produce this network for the same drug treatment conditions used by Hillenmeyer et al. (2008) to reveal a phenotype for nearly every single gene deletion, one would have to phenotype all 12.5 million crosses 350 times, generating over 4 billion measurements. Even worse, gene deletions are only one type of genetic perturbation. To generate an interaction landscape for every single nucleotide change in yeast, and for 350 treatment conditions, one would need to generate more than a quadrillion fitness measurements ($>10^{15}$). Needless to say, a more focused approach for characterizing pathway relationships for genes involved in cisplatin tolerance is warranted. The strategy here is simple – focus on a set of genes that share a phenotype as single mutants and only generate crosses among this small set of genes. This strategy, when combined with SGA, is known as epistatic mini-array profiling, or E-MAP (Schuldiner et al. 2005, 2006; Collins et al. 2006, 2007; Fiedler et al. 2009; Braberg et al. 2013). And, when combined with a treatment it is referred to as a differential E-MAP, or dE-MAP (St. Onge et al. 2007; Bandyopadhyay et al. 2010; Guénolé et al. 2013; Srivas et al. 2013; Hustedt et al. 2015)

Chapter 1 presents an advancement in the experimental design for a cisplatin dE-MAP. Currently published dE-MAPs typically use only one, or two, fixed concentrations of drug treatment. However, detectable sensitivity for cisplatin tolerance mutants ranges across nearly two orders of magnitude of cisplatin concentration. For example, the most sensitive mutants, *rad5*Δ, *rad6*Δ and *rad18*Δ, have a measurable phenotype at 2 μM cisplatin, and are completely dead at 25 μM. Whereas, *psy4*Δ, a gene named for cisplatin sensitivity, has a barely detectable phenotype at 80 μM. To measure a genetic interaction, ideally the experiment uses a treatment condition where at least one of the two mutations has a phenotype, and neither mutation causes death. Therefore, to optimize detection of genetic interactions, we use a common “low” concentration where most strains will remain viable, and an optimized “high” concentration that produces a phenotype of the query mutation – the gene deletion being crossed to the mini-array of cisplatin tolerance genes. While this project

remains ongoing, we have successfully used this strategy to produce a high-resolution set of genetic interactions covering 70% of the cisplatin tolerance network.

0.2 Genomic repercussions of *RAD5* overexpression

When I began my PhD, I spent a few months meeting with professors looking for scientific inspiration. Having a background in *de novo* gene synthesis, I was naturally drawn to the topic of DNA recombination and repair, so I set up a meeting with Professor Rodney Rothstein – a leader in this field. During our meeting, Rodney described a project that had nothing to do with DNA repair. Instead, he described an ongoing project in his lab that aimed to identify genetic interactions that could be used to treat cancer. Fundamentally, the project was an extension of the genetic screening for cancer drug targets approach described by Hartwell et al. (1997). The described approach attempts to address a major problem in the current strategy for treating cancers, which is that typical chemotherapy and radiation treatment programs have many dangerous side-effects because they do not specifically kill cancer cells, rather they preferentially kill all dividing cells. The logic of the genetic approach described by Hartwell, begins with the observation that there are numerous recurring genetic alterations that are observed in cancer, which represent an an attack surface for selectively killing cancer cells. Geneticists had long observed a phenomenon they called synthetic lethality, wherein combining two otherwise harmless mutations results in cell death (i.e. an extreme example of a negative fitness genetic interaction). So the theory goes, if a cancer has a mutation and we know the synthetic lethal interactions of that mutation, then we can selectively target the cancer cell by inhibiting a synthetic lethal interactor with a drug. This drug would selectively kill only cancer cells harboring the interacting mutation, while leaving normal cells unharmed. Remarkably, this approach has received some traction with the development of clinically approved PARP inhibitors, which exploit a synthetic lethal genetic interaction between *PARP* and commonly observed mutations in the breast cancer susceptibility genes, *BRCA1* and *BRCA2* (Ashworth and Lord 2018; Lord et al. 2015). Needless to say, I found the synthetic lethality approach to targeted cancer therapy to be incredibly interesting, so I asked to join the Rothstein lab.

The first major challenge to using synthetic lethality for treating cancer is to identify common genetic al-

terations in cancer. Then, by pairing these common genetic alterations with a collection of mutant strains one can identify synthetic lethal interactions. When I started in the Rothstein lab, the project I joined was attempting to leverage yeast genetics to identify synthetic lethal interactions for genes that are commonly overexpressed in PTEN deficient cancer. After performing a number of these genetic screens, it became clear that the data we were using to identify these overexpressed genes could not distinguish between genes that are overexpressed in individual cells, and genes that are just overrepresented in the population of cells used to measure expression (e.g. S-phase regulated genes will be overrepresented in a sample that has more dividing cells). Sadly, this issue undermined the initial motivation for performing my first synthetic lethal screens. Undeterred, I looked to other sources to find data for frequent genetic alterations in cancer. Around this time, The Cancer Genome Atlas consortium (Cancer Genome Atlas Research Network 2008) had released data for thousands of genomically characterized tumors. Importantly, these data included not only gene expression information, but also copy number information. After diving into the data, it became clear that each cancer cohort had distinct patterns of recurrent focal and chromosomal copy number amplifications resulting in overexpression of specific genes. Importantly, because gene copy number amplification is a physical change that occurs within individual cells, the change in expression could be attributed to an actual increase in transcript levels within cells rather than just a population level change.

In **Chapter 2**, I present a submitted manuscript that was motivated by the observation that the *RAD5* human homolog, *HLTF*, incurs frequent copy number amplification and corresponding overexpression in several types of human cancer (typically squamous cell carcinoma). *RAD5* is a member of the most essential cisplatin tolerance pathway, and, when overexpressed, we found that Rad5 bypasses canonical regulatory signals to cause cisplatin sensitivity and genome instability by driving hyperactive recombination at replication forks.

0.3 CRISPR-mediated nonsense

Yeast has been an ideal model organism for studying genetic interactions at scale to identify genetic determinants of phenotype, and to reveal pathway relationships between genes. One reason yeast has achieved

such traction as a genetic model arises from the ease with which scientists can modify the yeast genome in the lab. The ability to make targeted changes in an organism's genome emerged from pioneering work in the field of genetic recombination, which led to the development of one-step gene disruption – a fast and efficient approach to knockout a gene (Rothstein 1983; Baudin et al. 1993; Wach et al. 1994). Eventually, with the release of the *S. cerevisiae* genome (Goffeau et al. 1996; Goffeau and Vassarotti 1991) – the first complete DNA sequence assembly of any eukaryotic organism – work began to construct a comprehensive collection of yeast knockout mutants representing nearly all of the roughly 5,000 non-essential genes found in budding yeast (Winzeler et al. 1999; Giaever et al. 2002). This collection paved the way for the development of reverse genetic screening approaches, which utilize a large set of defined mutants to identify those that give rise to a particular phenotypic feature, such as cisplatin sensitivity (see **Chapter 1**), or a genetic interaction profile (see **Chapter 2**). In principle, one-step gene disruption should be feasible in any recombination proficient organism, which includes most of the tree of life. In fact, adaptation of the one-step gene disruption approach in mice was the basis of the 2007 Nobel Prize in Physiology or Medicine (Vogel 2007; Koller et al. 1990; Zijlstra et al. 1990, 1989; Koller and Smithies 1989), and the mouse genetics community has followed suit in sequencing the mouse genome (Waterston et al. 2002), and constructing their own large collections of knockout mice (Grimm 2006). However, one-step gene disruption, and other recombination mediated genome engineering methods, rely on rare DNA repair events, rendering these approaches inefficient in most organisms – entire PhDs could be dedicated to constructing a single knockout mouse. Thus, at the beginning of my PhD in the fall of 2011, yeast remained a dominant model in the field of systems genetics and was the focus of my research on the genetic relationships between cisplatin tolerance pathways and the genetic repercussions of *RAD5* overexpression.

However, in the spring of 2013, when I was working as a teaching assistant for Professor Ron Prywes and Professor Songtao Jia's course in molecular biology, a fellow teaching assistant, Ling Ye, informed me of a new advancement in the field of genome engineering that he was sure would change the future of genetics research. This new advancement in genome engineering technology that Ling described to me is the now famous CRISPR-Cas9 system, which has indeed changed the face of genetics and biomedical research (Jinek et al. 2012; Gasiunas et al. 2012; Deltcheva et al. 2011; Sapranauskas et al. 2011; Garneau et

al. 2010; Cong et al. 2013; Mali et al. 2013). The elegance of the CRISPR-Cas9 system for genome engineering is that it consists of a single protein (Cas9) and a short programmable guide RNA that targets Cas9 to a specific location in the genome where, upon recognition of the target sequence, it will generate a DNA double-strand break (DSB).

Suddenly, that rare DNA repair event necessary for recombination mediated genome engineering could be efficiently induced, opening up new avenues for modifying the genomes of almost any organism. Moreover, this simple system was amenable to simple modifications which allowed scientists to localize other proteins to specific DNA sequences without causing DNA breaks. Using this approach, Komor et al. (2016) and others, modified the CRISPR system to localize an enzyme that would directly convert cytosines (C) to thymines (T) in a small window of five nucleotides near the guide targeted sequence without having to create a DSB – an attractive feature for those interested in modifying cells that might already be deficient in DNA repair. When this system, known as CRISPR-mediated base editing, was first described, I had recently begun a collaboration with Dr. Alberto Ciccia's lab to help with sequence analysis for a pilot run of a CRISPR-Cas9 based reverse genetic screen. Dr. Ciccia and a postdoctoral research associate in his lab, Dr. Pierre Billon, realized that CRISPR-mediated base editing of C to T could be used to disrupt gene function by converting four specific codons to premature stop codons in protein coding genes. What was unclear though, was whether this approach for knocking out gene function would be generalizable given the limited number of targetable codons and other technical restrictions dictated by the CRISPR system.

Chapter 3 is a published manuscript I co-authored with Dr. Billon and Dr. Ciccia addressing the scope and feasibility of CRISPR-mediated base editing for disrupting gene function (Billon et al. 2017). The title of the original publication – *CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons (iSTOP)* – sums up the ultimate conclusion; CRISPR-mediated C to T base editors allow DSB free induction of stop codons at 23 different codons per gene on average (i.e. 1 iSTOP targetable codon for every 26 codons in the human genome). Thus, iSTOP provides a new tool alongside CRISPR-Cas9, and classic one-step gene disruption, that no longer relies on DSBs and recombination to efficiently knockout gene function. iSTOP, and other gene editing tools, will enable scientists to look beyond

the humble baker's yeast, to many other branches of life, to gain insight from systems genetics into the genetic determinants of life.

Chapter 1

Epistasis and cisplatin tolerance

Cisplatin, is one of several platinum-based drugs used in an estimated 50% of all chemotherapeutic regimens. Its anticancer activity derives predominantly from intrastrand crosslinks, which physically link bases on the same strand of DNA, causing distortions that impede replication fork progression in rapidly dividing cancer cells. In *Saccharomyces cerevisiae*, over 100 genes have been identified to be important for cisplatin tolerance or resistance. Many of these genes are known to function in major DNA damage tolerance pathways. However, there remain a number of individual genes that are known to affect cisplatin sensitivity, but have no known role in any of the major cisplatin tolerance pathways. And, while much of the major cisplatin tolerance pathways are independently well studied, how these pathways relate to each other to maintain fitness and genome integrity remains poorly understood. To address these unknown aspects of cisplatin tolerance we are constructing a cisplatin differential epistasis mini-array profile (dE-MAP). This approach uses double-mutant colony fitness to reveal novel pathway membership and provides insight into how damage tolerance pathways work together to establish fitness in the face of cisplatin crosslinks. The experimental design of our dE-MAP differs from prior epistasis mapping strategies in that it optimizes cisplatin concentrations tested for each query mutant crossed to the mini-array of mutant strains. This optimization enhances detection of genetic interactions by ensuring that each query mutation has been screened at a concentration of cisplatin that induces a phenotype. Currently 89 queries, of 202 strains present in the mini-array, have been scored for fitness interactions, resolving approximately 70% of the network. Both well characterized and novel pathway relationships have begun to emerge. Current results indicate that, genetically, cisplatin mimics UV damage, and that postreplication repair is a central node in the cisplatin tolerance network.

1.1 Introduction

cis-Diamminedichloroplatinum (II), commonly known as cisplatin, is the first discovered member of the platinum-based group of chemotherapeutics used to treat millions of cancer patients worldwide (Galanski et al. 2005; Simon et al. 2000; Wu et al. 2004; Kelland 2007). The ability of cisplatin to inhibit growth of dividing cells was discovered during studies of electromagnetism's effect on *Escherichia coli* cell division. In this study, Dr. Barnett Rosenberg realized that growth of *E. coli* in his experiments was not in fact inhibited by electromagnetic fields, but by a compound, cisplatin, leaching into the buffered solution from platinum electrodes (Rosenberg et al. 1965). From this insight, Rosenberg made the logical leap that if cisplatin could inhibit growth of *E. coli*, perhaps it could also inhibit growth of cancer cells. Testing this hypothesis, Rosenberg demonstrated that tumors in mice were indeed responsive to cisplatin treatment (Rosenberg et al. 1969). Cisplatin and its derivatives are now used daily in the clinic and have inspired decades of scientific inquiry into the mechanism of action for, and acquired resistance to, this critically important small molecule (Kelland 2007).

Cisplatin causes damage to DNA primarily by causing intrastrand crosslinks (90% of lesions) typically between adjacent guanines on the same strand (Wong et al. 2010). Interstrand DNA crosslinks, covalent linkage between the two strands of DNA, can also occur, but is considered a minor contribution to the total number of lesions (5%) (Eastman 1986). While cisplatin is an immensely effective treatment for many tumors, acquired resistance can develop through a number of vaguely identified mechanisms (Siddik 2003; Kelland 2007). Known mechanism of cisplatin resistance fall into three basic categories, (1) reduced cisplatin uptake, (2) increased repair, and (3) increased lesion bypass. In *Saccharomyces cerevisiae*, individual genetic requirements for cisplatin resistance have been well established through several independently performed chemogenomic screens utilizing the complete yeast gene disruption collection (Birrell et al. 2002; Wu et al. 2004; Lee et al. 2005; Huang et al. 2005; Hastie et al. 2006; Liao et al. 2007). Genes required for cisplatin tolerance can be grouped into the following canonical pathways (reviewed by Ciccia and Elledge 2010; Heyer et al. 2010; Boiteux and Jinks-Robertson 2013; Rocha et al. 2018):

- **Lesion bypass** via Post Replication Repair (PRR). PRR involves post-translational modifications performed by Rad6-Rad18 and Rad5-Ubc13-Mms2 on proliferating cell nuclear antigen (PCNA – Pol30 in budding yeast) at stalled DNA replication forks. These modifications determine choice between two PRR sub-pathways.
  - Error-prone PRR performed by translesion synthesis (TLS) (e.g. Rev1, Rev3 & Rev7) that allows synthesis across from damaged bases.
  - Error-free template switching (TS) that allows replication bypass on a nascent sister chromatid. This pathway is thought to be activated by Rad5-Ubc13-Mms2 to engage homology directed repair (HDR) machinery (e.g. Rad51, Rad52, Sgs1 etc.).
- **Lesion repair** via Nucleotide Excision Repair (NER) (e.g. Rad1, Rad10, Rad14 etc.). NER can be categorized into two sub-pathways, differing in their mechanism for lesion recognition, but sharing downstream repair machinery:
  - Transcription Coupled NER (TC-NER), which surveils DNA during transcription (involving Rad26).
  - Global Genomic NER (GG-NER), which can recognize lesions independently from transcription (Involving Rad23 & Rad4).

Of course, other DNA repair mechanisms also exist including Mismatch Repair (MMR), Non-Homologous End Joining (NHEJ), and Base Excision Repair (BER). And, each pathway, whether it be PRR, TLS, TS, NER, MMR, NHEJ, or BER, is important in the context of specific DNA lesions. While all of these pathways are independently well studied, how the various repair and damage tolerance pathways relate to each other to establish cisplatin tolerance and fitness remains unclear. Moreover, more than 100 genes have been identified as important for cisplatin tolerance, yet the role of many of these genes in establishing a cisplatin tolerance network remains to be determined.

To address these gaps in understanding, we are currently constructing a cisplatin differential epistasis mini-array profile (dE-MAP) (St. Onge et al. 2007; Bandyopadhyay et al. 2010; Guénolé et al. 2013; Srivas et al. 2013; Hustedt et al. 2015) – see schematized experiment in **Figure 1.1**. This variation on the standard E-MAP is a systems genetics approach for optimizing detection of genetic interactions – pairs of mutant

alleles that have an unexpected phenotypic consequence. In simple terms, an E-MAP just involves making all pair-wise combinations for a set of alleles to identify synergistic or epistatic relationships to a phenotype (Schuldiner et al. 2005, 2006; Collins et al. 2006, 2007; Fiedler et al. 2009; Braberg et al. 2013). Typically, the Synthetic Genetic Array (SGA) method is used to generate the double mutants, and colony size or density is used for the phenotype as a proxy for cellular fitness (Tong et al. 2001; Tong and Boone 2006). In this context, a synergistic relationship between two mutant alleles represents unexpectedly low fitness in the double-mutant (sometimes referred to as synthetic lethality, or a negative fitness interaction), and the alleles are interpreted as having parallel compensatory contributions to fitness (i.e. the alleles impact closely related parallel pathways). Conversely, an epistatic relationship between two mutant alleles represents unexpectedly high fitness in the double-mutant (sometimes referred to as suppression, or a positive fitness interaction), and the alleles are interpreted as having a shared contribution to fitness (i.e. the alleles impact the same pathway). The differential extension of the E-MAP approach involves scoring the phenotype after exposure to a treatment. Constructing a dE-MAP reveals (1) the quantitative fitness for all alleles in the set as single mutants, with and without exposure to treatment, (2) genetic interactions that exist between alleles, with and without exposure to treatment. Importantly, being able to account for existing double-mutant interactions and drug sensitivity allows detection of treatment specific genetic relationships.

A number of dE-MAPs have been constructed for varying size sets of alleles and under different treatment conditions (mostly DNA damaging agents). Notable examples cited above include (in order of publication):

1. A 26 query × 26 array treated with 0.002% methyl methanesulfonate (MMS) (St. Onge et al. 2007).
2. A 418 query × 418 array treated with 0.002% MMS (Bandyopadhyay et al. 2010).
3. A 55 query × 2,022 array treated with 0.01% MMS, 5 µg/ml camptothecin, or 75 µM zeocin [2013-01-Gunol-vanAttikum].
4. A 37 query × 1,397 array treated with 20, or 80 J/m^2 of ultraviolet radiation (Srivas et al. 2013).
5. A 35 query × 1,311 array treated with 20, or 100 mM hydroxyurea (Hustedt et al. 2015).

Here we present the design, execution, current insights and progress in constructing a 202 query × 202 array cisplatin dE-MAP. Importantly, we highlight a key limitation in published dE-MAP studies. These studies all

use fixed treatment concentrations, which impacts the resolution for detecting genetic interactions, and affects interpretation of these existing dE-MAP networks (**Figure 1.2**). We propose a solution by optimizing cisplatin concentrations to generate a phenotype in for the query allele, while minimizing lethality caused by alleles in the array. Results from the cisplatin dE-MAP will reveal how multiple genetic pathways function as a network to establish a robust system for tolerating cisplatin mediated DNA damage.

1.2 Results

1.2.1 Designing the cisplatin strain collection

The E-MAP and dE-MAP methods are powerful systems genetics approaches for identifying genetic relationships because they leverage the observation that most genetically interacting alleles share a common phenotype (Schuldiner et al. 2005; Collins et al. 2006). Using this observation, we can focus our search for cisplatin specific genetic relationships by identifying genes that share a common phenotype. The phenotype we are specifically interested in is that of cisplatin sensitivity. Therefore, we first identified the set of genes that, when deleted, cause a change in cisplatin sensitivity. These genes were identified from three sources:

1. Genes annotated in the *Saccharomyces* genome database as being sensitive, or resistant to cisplatin.
2. Genes identified by other groups as modifying cisplatin sensitivity using the CLIK algorithm to identify lenient screen cutoffs (Birrell et al. 2002; Wu et al. 2004; Lee et al. 2005; Huang et al. 2005; Hastie et al. 2006; Dittmar et al. 2013).
3. Our own chemogenomic screens of the *MAT***a** and *MATα* gene disruption collections at cisplatin concentrations of 0, 50, 100 and 125 μM.

Since some mutants might only show cisplatin sensitivity when paired with deletion of a redundant pathway, we also bootstrapped genes that had evidence for being involved in DNA replication, recombination or repair using the following two sources of evidence:

1. Gene ontology annotation to DNA replication, recombination, or repair.

2. Genes having shown strong genetic interactions in previously published differential epistasis maps of other DNA damaging treatments (MMS, camptothecin, zeocin, ultraviolet radiation, and hydroxyurea) (Bandyopadhyay et al. 2010; Guénolé et al. 2013; Srivas et al. 2013; Hustedt et al. 2015).

From this set, some members of overrepresented pathways were removed (e.g. 6 subunits of the vacuolar membrane ATPase were removed from the set). Dubious ORFs were also removed. The resulting mini-array consisted of 202 strains, allowing for 182 wildtype control strains to be spread evenly across a single 16×24, 384 strain plate. All query and array strains were independently verified as described in the **Methods 1.4.1**.

1.2.2 Strategy for optimizing detection of genetic interactions

Before beginning the dE-MAP experiment, we were faced with the challenge of selecting the best concentration of cisplatin to use for detecting interactions. Previous dE-MAP studies were performed using one (sometimes two) fixed concentrations of a treatment (Bandyopadhyay et al. 2010; Guénolé et al. 2013; Srivas et al. 2013; Hustedt et al. 2015). However, when we assayed our mini-array to determine the optimal range for the cisplatin dE-MAP, we observed sensitivity across nearly two orders of magnitude of cisplatin concentration (e.g. *rad5*Δ and *rad18* show sensitivity at 2.5 μM, whereas other mutants did not display a phenotype until 125 μM). No pair of cisplatin concentrations would be sufficient to detect genetic relationships for both the most sensitive mutant strains, and weakly sensitive strains. Unfortunately, performing a full cisplatin titration for every round of SGA was also not feasible as each additional treatment condition multiplies the number of plates to an unmanageable scale. Instead, as a compromise, we decided to optimize the cisplatin concentrations based on the phenotype of each query strain. The final concentrations, in addition to untreated, would correspond to a low concentration that would minimize lethality of strains in the array, and a high concentration that would produce a cisplatin sensitive phenotype in the query strain. The resulting protocol is schematized in **Figure 1.1**. The choice for optimizing the high cisplatin concentration was determined based on the observation that colony fitness interactions, under the multiplicative null model,

are best detected when one or both alleles have a fitness phenotype, but neither allele has caused lethality. This is illustrated in **Figure 1.2**. We therefore performed a pilot round of SGA by crossing a *leu2::NatMX* control query to the mini-array and measuring cisplatin sensitivity at concentrations ranging from 5 to 100 µM. Since all strains in the array are *leu2*, this reference query allowed us to identify the single mutant cisplatin dose response profiles for each allele as they would be observed at the end of SGA.

1.2.3 Single-mutant cisplatin sensitivity

From the cisplatin titration results we observed very few strains that showed strong sensitivity to 25 µM cisplatin (**Figure 1.3 left**). Interestingly, the most acutely sensitive strains were *rad5*Δ, *rad18*Δ, and *rad6*Δ – all members of postreplication repair (PRR) that are involved in post-translationally modifying PCNA (*S. cerevisiae* Pol30) to regulate how replication forks respond to a DNA lesion (Xiao et al. 2000). This observation alone underscores the importance of enabling DNA replication forks to respond appropriately when encountering cisplatin mediated damage. Furthermore, when increasing the concentration to 80 µM, we observed more than 55 strains that displayed at least a 35% growth defect (**Figure 1.3 right**). At 80 µM, we also observed complete lethality for mutants of genes in both the error-free and error-prone branches of PRR corresponding to sister-chromatid template switching (TS – *rad51*Δ, *rad52*Δ, *rad54*Δ, *rad57*Δ, *sgs1*Δ) and translesion synthesis (TLS – *rev1*Δ, *rev3*Δ, *rev7*Δ, *pol32*Δ) respectively. Additionally, the MRX complex (*mre11*Δ, *rad50*Δ, *xrs2*Δ), involved in an early step of homology directed repair (HDR), and nucleotide excision repair (NER – *rad1*Δ, *rad10*Δ, *rad2*Δ, *rad4*Δ, *rad14*Δ), responsible for excising damaged DNA bases, were both inviable when treated with 80 µM cisplatin. Lastly, the *slx5*Δ and *slx8*Δ complex, a SUMO-targeted ubiquitin ligase, known to degrade HDR/TS members *rad52*Δ and *rad57*Δ, also showed near lethality at this higher cisplatin concentration. All of these genes, owing to having the strongest response to cisplatin treatment, represent the core cisplatin tolerance pathways.

1.2.4 Insights from the static E-MAP

In constructing the cisplatin dE-MAP, a static untreated E-MAP is also generated, allowing detection of genetic fitness relationships in standard laboratory growth conditions. When focusing on the major cisplatin tolerance pathways, we see many genetic relationships that reaffirm known relationships between repair pathways (**Figure 1.4 left**). For example, we see epistasis amongst the MRX complex and strand invasion components of HDR. We also detect some epistasis between members of NER (e.g. between *rad2*Δ & *rad14*Δ and between *rad14*Δ & *rad1*Δ). Unexpectedly, given the well established role of MRX in recognizing broken ends to initiate resection for HDR, we see some considerable positive interactions between MRX and NER (e.g. between *rad50*Δ and *rad2*Δ, *rad14*Δ, *rad10*Δ and *rad4*Δ). HDR and NER are thought to act as parallel repair pathways and thus one would reasonably predict that there would be negative fitness interactions between these two pathways. Counter to the observed epistasis between MRX and NER, crossover resolution shows evidence for synergy between HDR and NER, as expected (e.g. between *sgs1*Δ, *top3*Δ and NER). Without cisplatin challenge, we see no evidence for epistasis between members of error-prone PRR (*rev3*Δ & *rev7*Δ), but they do share a consistent pattern of negative genetic interactions with the error-free PRR branch, and Rad5's E2 ubiquitin ligase partners, *mms2*Δ & *ubc13*Δ. This is consistent with the canonical model of Rad5-Ubc13-Mms2 polyubiquitinating PCNA to activate error-free template switching as an alternative to error-prone translesion synthesis. With respect to PRR, we see *rad18*Δ and *rad5* show negative interactions with NER, MRX, and HDR, indicating that PRR acts redundantly to these processes to maintain fitness in standard growth conditions. Overall, with only a few exceptions, the untreated E-MAP largely confirms some of the well established genetic relationships between the core cisplatin tolerance pathways. However, the network is relatively sparse, and many known pathway relationships are not revealed through epistasis, likely owing to a lack of fitness phenotypes for some of these mutant alleles, thus limiting detection of positive genetic relationships (**Figure 1.2 left**).

1.2.5 Insights from the cisplatin differential E-MAP

The cisplatin dE-MAP reveals pathway relationships specific to cisplatin treatment after accounting for genetic interactions observed in the untreated condition (**Methods 1.4.5**). Again focusing on the core cisplatin tolerance pathways, we see clear epistasis and synergistic relationships emerge (**Figure 1.4 right**). For example, epistasis amongst most of NER becomes very clear (purple box including *rad2*Δ, *rad14*Δ, *rad10*Δ & *rad1*Δ). Remarkably, the dE-MAP is also able to resolve NER into its two canonical pathways, transcription-coupled and global-genomic (TC-NER & GG-NER) represented by *rad26*Δ and *rad23*Δ respectively. Combining these mutations results in synergy, indicative of parallel compensatory contributions to cisplatin tolerance. And, when looking at the remaining profile of synergistic interactions, GG-NER appears to be more essential than TC-NER for maintaining cisplatin tolerance. With respect to the epistatic relationship between MRX and NER observed in the E-MAP, this relationship is reversed in the context of cisplatin. MRX and NER clearly synergize, acting in parallel compensatory pathways for responding to cisplatin damage. In fact, NER represents a clearly distinct parallel pathway to all of the other core cisplatin tolerance pathways. Since NER can act independently to excise intrastrand DNA crosslinks, it stands to reason that defects in NER increase the effective dose of cisplatin resulting in an increased burden on the other mechanisms for cisplatin tolerance.

While the matrix approach to visualizing the dE-MAP is rich in detail, it can be difficult to discern overall pathway relationships. As an alternative visualization, we have distilled the major patterns of genetic requirements between the core cisplatin tolerance pathways as a summary network diagram shown in **Figure 1.5**. From this diagram, we see that HDR, acting in conjunction with crossover resolution, forms a pathway with many members acting redundantly to both TLS, NER and PRR. We also see that NER, as mentioned above acts completely independently of all other tolerance pathways, synergizing strongly with defects in HDR, crossover resolution, PRR, MRX and DNA replication fork stability. Interestingly, the early PRR regulatory members, Rad5 and Rad18, form a central pathway, bridging crossover resolution, TLS, MRX and fork stability into a larger meta pathway of cisplatin tolerance. This coordination of multiple tolerance pathways may explain why *rad5*Δ and *rad18*Δ show the strongest sensitization to cisplatin (**Figure 1.3 left**).

1.2.6 Current progress and novel pathway membership

Currently 89 query mutations have been crossed to the array with four complete biological replicates. The analysis of the core cisplatin tolerance pathways described above does not include many of the alleles included in the current network. To visualize pathway memberships for all alleles in the network, dimension reduction was performed using t-SNE (van der Maaten and Hinton 2008), to cluster genes in two dimensions (**Figure 1.6**). From this clustering we can see evidence for some novel, or unappreciated pathway memberships, in some of the core pathways. For example, we see *saw1*Δ clustering with TLS, which is surprising given its known role in recruiting NER members, Rad1 & Rad10, to perform a role in single-strand annealing and double-strand break repair (Li et al. 2013). Additionally, we see Slx8, part of the Slx5-Slx8 SUMO-targeted ubiquitin ligase complex clustering with the error-free HDR branch of PRR. Since Slx5-Slx8 is known to target Rad52, and Rad57 for degradation (Xie et al. 2007), this clustering suggests that an inability to degrade Rad52 and Rad57 at the correct time may be just as important as having these proteins at all for efficient HDR. Another remarkable feature, briefly mentioned above, is that the MRX complex (acting early in HDR), and crossover resolution (acting late in HDR), are easily distinguished genetically from the strand invasion components of HDR. While crossover resolution tends to cluster independently, MRX tends to cluster with chromatin remodelers, fork stability, and sister-chromatid cohesion factors. This observation provides support for a role of MRX acting to stabilize replication forks, possibly by tethering sister strands and other DNA ends that may be generated during PRR at replication forks (Mirzoeva and Petrini 2003; Tittel-Elmer et al. 2009). Lastly, while much of the literature for the PRR member, Rad5, focuses on its role in poly-ubiquitinating PCNA to activate error-free sister-chromatid recombination (Hoege et al. 2002; Branzei et al. 2004), loss of the Rad5 E2 ubiquitin ligase partners (Ubc13-Mms2) is genetically distinct from loss of Rad5 itself. Some might argue that this could be due to Rad5 playing a role in recruiting TLS polymerases (Kuang et al. 2013). However, this would predict that *rad5*Δ should be genetically similar to TLS mutants. This is not the case, rather, one of the major distinguishing features between deletion of *RAD5* and its E2 partners is having opposite genetic interactions with MRX – *rad5*Δ shows evidence for epistasis with MRX whereas deletion of the Rad5 E2 partners shows stronger evidence for synergy. This observation

could suggest that Rad5, one of the most critical cisplatin tolerance proteins, may work with MRX to promote stability of a cisplatin challenged replication fork.

1.3 Discussion

Cisplatin is an essential first-line chemotherapeutic effective in treating many types of cancer. Encoded within cells, cancer or otherwise, are multiple cisplatin tolerance pathways that enable repair, or bypass, of intrastrand crosslinks on DNA (Deans and West 2011). In yeast, over 100 genes take part in enabling cisplatin tolerance, some of which have no known role in this process. This study has provided a clearer picture of how the core cisplatin tolerance pathways relate to each other to create a cisplatin tolerance network.

1.3.1 NER is the dominant *repair* pathway

Importantly, nucleotide excision repair (NER), and not base excision repair (BER), or mismatch repair (MMR), is clearly the dominant pathway for *repair* of the lesions caused by cisplatin. This observation makes sense given that 90% of DNA lesions caused by cisplatin are thought to occur between bases on the same strand (Eastman 1986). These interstrand crosslinks are amenable to repair by NER, whereas MMR can recognize but not repair these lesions (Rocha et al. 2018). Additionally, a cisplatin crosslink, even when on the same strand, creates a bulky adduct that would likely be preferentially recognized by NER, rather than BER, which typically repairs individually damaged bases that do not distort the DNA double-helix (Todd and Lippard 2010; Baute and Depicker 2008; Deans and West 2011). The overall single-mutant cisplatin sensitivity profiles are consistent with these notions – NER is essential, whereas BER and MMR are not (**Figure 1.3**). In some sense, cisplatin treatment can be likened to ultraviolet radiation induced DNA damage, which often covalently links adjacent thymines on the same strand. In fact, there have been some proposals to treat cancers with NER inhibitors specifically as a combination therapy with cisplatin (Gentile et al. 2016), and a UV based assay has been used to detect efficacy of such inhibitors (Szalat et al. 2018).

1.3.2 PRR is the most critical pathway for cisplatin tolerance

NER, and specifically global-genomic NER, clearly plays an important role in repairing cisplatin lesions. Additionally, our cisplatin differential epistasis mini-array profile (dE-MAP) has revealed unequivocally that NER acts in parallel to all of the other core cisplatin tolerance pathways (**Figure 1.4**). This is consistent with a model whereby the efficiency of NER dictates the effective damage burden passed to the lesion bypass machinery. While NER is important for *repairing* cisplatin lesions, it cannot be understated how acutely sensitive mutations in the *lesion bypassing* postreplication repair pathway are to the lowest doses of cisplatin treatment (**Figure 1.3**). The current dE-MAP provides a clue as to why this may be the case. Rad5 and Rad18 act with multiple core components of the cisplatin tolerance network (**Figure 1.5**).

1.3.3 Genetic evidence supporting a role for MRX in stabilizing replication forks

In constructing both the E-MAP and cisplatin dE-MAP we observed some interesting genetic relationships for the MRX complex. Specifically, in untreated conditions, MRX and NER appear to share a contribution to fitness, suggesting some overlapping role in maintaining growth of colonies **Figure 1.4**. Also, MRX in the cisplatin dE-MAP clusters away from other HDR members and with factors more associated with maintaining the stability of replication forks **Figure 1.6**. A role for MRX in stabilizing forks, rather than in double-strand break end tethering and resection, has been proposed (Mirzoeva and Petrini 2003; Tittel-Elmer et al. 2009), but this role is often under-appreciated.

1.3.4 Clinical relevance and future directions

The current cisplatin dE-MAP is approximately 70% complete and is already providing a clearer picture of how the cisplatin tolerance network is structured. Importantly, understanding the structure of the cisplatin tolerance network will help researchers and clinicians develop strategies to treat tumors with defects in specific damage tolerance pathways (Rocha et al. 2018). Identifying fitness dependencies that occur between repair pathways has already resulted in the development of synthetic lethal approaches for treating repair

deficient tumors, with PARP inhibitors recently being approved for treating certain *BRCA* mutant cancers (Ashworth and Lord 2018; Lord et al. 2015). Understanding how these synthetic lethal relationships might arise in the context of chemotherapy may lead to new avenues for drug development and could certainly lead to to improved treatment decisions. Indeed, at least one group has leveraged genetic relationships similar to what we have identified in our dE-MAP, by pairing PARP inhibition with cisplatin in cell lines that have low expression of a critical NER gene (Cheng et al. 2013). Our network would predict what this study observed, that inhibiting any other cisplatin tolerance pathway should synergize with low NER activity.

Future directions for this project will seek to compare the cisplatin dE-MAP to previously published dE-MAPs generated for other DNA damaging treatments to identify genetic relationships specific to cisplatin treatment. And, with the development of new CRISPR-based pooled screening approaches for generating combinatorial knockouts in other model systems (Shen et al. 2017), I look forward to taking some of the lessons learned from mapping epistasis in yeast into human cell lines. Two of these lessons I would like to share here:

1. Most interacting genes share a common phenotype. Therefore, use chemogenomics (or another equally simple assay) to define a focused set of single-mutants that share a common phenotype before pursuing a combinatorial genetic interaction screen.
2. Detection of genetic interactions can critically depend on environmental conditions. Therefore, carefully consider treatment conditions for optimizing detection of genetic relationships based on a validated genetic interaction null model.

1.4 Methods

1.4.1 Strain construction and validation

Strains in the collection were streak-purified from the *MAT***a** *KanMX* gene disruption collection and validated by PCR using primers upstream and downstream of the integrated *KanMX* marker. Additional *KanMX*-

internal and *GENE*-internal primers were also used to confirm both presence of the *KanMX* marker and absence of the gene's open reading frame. Strains that failed to validate were reconstructed using the approach described by the yeast gene deletion consortium (http://www-sequence.stanford.edu/group/yeast_deletion_project/protocols.html). Query strains were constructed using the PCR product of the *KanMX* cassette generated during strain validation, which included approximately 250 bases of homology on either side of the cassette. This PCR product was co-transformed with *NatMX* product to simultaneously integrate and marker-swap *KanMX* with *NatMX*. Transformants were then validated using genomic primers outside of the transformed region of homology. Additional *NatMX*-internal and *GENE*-internal primers were also used to confirm presence of the *NatMX* marker and absence of the gene. The E-MAP approach also has an internal strain validation mechanism in that each query allele is present in the array. Therefore, when crossing the *NatMX* query to the same allele marked with *KanMX*, the two markers will fail to independently segregate and lethality will be observed. Related to **Section 1.2.1**.

1.4.2 Synthetic genetic array

Differential epistasis mapping was performed using the synthetic genetic array (SGA) method to construct pairwise double-mutants among the strains in the cisplatin collection (Tong et al. 2001; Tong and Boone 2006; Costanzo et al. 2010; Baryshnikova et al. 2010). This method begins by mating the *MAT***a** collection of *KanMX* marked single gene disruption mutants to a single *MATα* query strain containing a *NatMX* marked single gene disruption. Diploids are then selected by plating on media containing geneticin (G418) and nourseothricin (NAT) to select for presence of the *KanMX* and *NatMX* markers respectively. Haploids are then generated by plating on sporulation media (SPO). After sporulation, strains are pinned to a series of increasingly selective conditions to select for haploids having the same genetic background and both *KanMX* and *NatMX* marked gene disruptions. Mating type specific markers are selected for using S-AEC and canavanine to select for *lyp1*Δ and *can1*Δ respectively (lysine and arginine are removed from the media to encourage uptake of S-AEC and canavanine into the cell). Mating type **a** is selected for by removing histidine which is complemented by *MAT***a** specific expression from the *STE2* promoter of *HIS5*, a histidine

biosynthesis gene from *Schizosaccharomyces pombe* that complements *his3*. After selecting for double mutant haploids, strains are pinned to the final growth media containing concentrations of cisplatin optimized to detect both positive (epistatic / positive) and negative (synergistic / negative) genetic interactions based on the sensitivity of the query strain to the drug. At least four biological replicates were generated for each query strain by repeating the experiment from independent colonies in at least two separate rounds of SGA. Every batch of SGA included a *leu2::NatMX* query to account for variation in cisplatin potency from batch to batch. The strain collection contains 178 *his3::KanMX* strains placed at the border of the grid (excluded from analysis) and evenly distributed throughout the center of the place (used for spatial normalization of colony growth). 0 µM cisplatin, *leu2::NatMX*, and *his3::KanMX* are also used as controls for estimating fitness of WT, and single mutants, in the presence and absence of cisplatin. Relevant strain genotypes are provided below.

- **Genotype of strain collection (BY4741):**
  - *MAT***a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 A::KanMX*
- **Genotype of query strain (Y8205):**
  - *MATα his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::STE2p-Sp_HIS5 lyp1Δ::STE3p-LEU2 Q::NatMX*
- **Genotype at end of SGA**
  - *MAT***a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::STE2p-Sp_HIS5 lyp1Δ::STE3p-LEU2 A::KanMX Q::NatMX*

1.4.3 Detailed SGA protocol

All incubation steps take place in a 30°C incubator. SC and YPD are standard laboratory recipes with 2% dextrose. Concentrations of additives are 60 mg/L of Can, 50 mg/L of S-AEC, 200 mg/L of G418, and 100 mg/L of clonNAT. The SPO recipe is described by Tong et al. (2001), which consists of 2% agar, 1% potassium acetate, 0.1% yeast extracts, 0.05% glucose, and is supplemented with uracil, histidine, and leucine. Robotic pinning is performed using a Singer RoToR robot.

1. **Streaking out:** 3-4 days prior to starting SGA, take query strains out of -80°C freezer and streak for single colonies on YPD +NAT plates.

2. **Liquid cultures:** Once single colonies have grown, inoculate one colony per query into 5 mL of YPD. Grow overnight in a shaking incubator. On that same day, robotically pin copies of the strain library on YPD + G418 plates. Print at least one copy for every 8 queries.

3. **Lawns:** The next morning, pellet the 5 mL cultures, re-suspend in 400uL of sterile water, and spread 400 uL with sterile glass beads on YPD rectangular plates. Allow lawns to grow overnight. *Tip: quickly invert the plates when removing glass beads to prevent gaps in the lawn.*

4. **Mating:** The next morning, robotically pin the lawns and the strain collection to YPD.

    - Run **replicate many** program, using the strain collection as the source, and YPD plates as targets. Turn on **recycle** and **revisit**. 384 short pinned pads should be used.
    - Run **replicate** program using each lawn as source, and each YPD plate (previously pinned with strain collection) as target. **DO NOT** recycle, or revisit. 384 short pinned pads should be used.

5. **Diploid Selection 1:** The next morning, replicate each cross onto YPD +G418 +NAT.

6. **Diploid Selection 2:** The next morning, repeat **step 5**.

7. **Sporulation:** The next morning, replicate each plate onto SPO +G418 (G418 is added to prevent contamination and does not hinder sporulation).

8. **Haploid Selection 1:** One week later, replicate each plate onto SC -His -Arg -Lys +Can +S-AEC plates

9. **Haploid Selection 2:** Two days later, repeat **step 8**

10. **Haploid Selection 3:** The next morning, replicate each plate onto SC -His -Arg -Lys +Can +S-AEC **+G418**

11. **Haploid Selection 4:** The next morning, replicate each plate onto SC -His -Arg -Lys +Can +S-AEC **+G418 +NAT**

12. **Prepare cisplatin media:** The next day, cisplatin media should be prepared.

13. **Pin to final plates:** The next day.

    - **384-well microtiter plate preparation:** Fill 384-well plates with 80 µL sterile water. *Prepare as many plates as there are independent crosses.*
    - **Resuspension:** Pin from the last haploid selection plate, to the 384-well microtiter plate from, using 384 long pin pads, and an extensive resuspension program.
    - **1 to 4 liquid to agar pinning:** Pin from the newly resuspended 384 well plate to the destination agar plates. The first agar plate pinned from a microtiter plate should be discarded as the four technical replicates are often inconsistent for the first pinnings. *384 long pin pads should be used. Pinning should be done from a 384 to a 1536 density format. For a given cross, use the same pad across all cisplatin plates. This step should be performed as quickly as possible to avoid settling of the yeast in the 384 well plates.*

1.4.4 Cisplatin media preparation

Cisplatin media should be prepared one day prior to use to ensure maximum stability of the drug.

A brief protocol for making the final media is included below.

1. **Prepare SC:** Prepare a sufficient volume of SC to cover the number of plates required as well as an additional plate per independent cross.

2. **Prepare cisplatin solution:** One hour before pouring plates, prepare a 0.9% NaCl 1 mg/mL cisplatin solution. *To ease solubilization, the solution can be warmed a few minutes in a 50°C water bath.*

3. **Media preparation:** In a large flask, mix the appropriate volumes of autoclaved SC media, cisplatin solution, G418, and NAT after the media cools to below 52°C.

4. **Plate pouring:** Pour 60 mL of media per plate and let dry overnight on an even surface. *Covering the plates with foil may prevent photodegradation of the drug and helps keep the plate surface level.*

1.4.5 Scoring interactions

Expected fitness of the query mutation q combined with the array mutation a, qa (E^{qa}) is calculated as the multiplicative fitness contributions of each single mutant (F^q, and F^a) scaled to fitness of wildtype ($F^{wt} \equiv 1$).

$$E^{qa} = F^q \cdot F^a$$

And the genetic interaction magnitude between q and a (S^{qa}) is computed as the difference between expected and observed fitness of the double mutant:

$$S^{qa} = E^{qa} - F^{qa}$$

In practice, plate normalization must be performed to account for systematic differences in growth between plates. This has the effect of also dividing out the contribution of the query mutation, and the contribution of the drug to colony fitness. Thus, after plate normalization, the resulting value is the marginal contribution of the array mutation to fitness, which, in the absence of a genetic or drug interaction, should be equivalent to the plate normalized measurement of the array strain in the reference context (i.e. *leu2::NatMX* query, or 0 μM cisplatin control conditions). Using P to denote plate normalized colony measurements, genetic interaction magnitude can be directly calculated with:

$$S^{qa} = P^{qa} - P^a$$

For the E-MAP (related to **Figure 1.4 left**), S^{qa} is calculated by aggregating values of P^{qa}, and P^a from the 0 μM untreated condition using every observation from running a given query allele (providing P^{qa}) and the *leu2::NatMX* reference allele (providing P^a) through SGA. A t-distribution based 95% confidence interval error (ε), with $\alpha = 0.05$ and $n - 1$ degrees of freedom, can be calculated for both P^{qa} and P^a given the standard deviation (s), and the number of observations for each value (n) with:

$$\varepsilon = t_{n-1}^{\alpha} \cdot \frac{s}{\sqrt{n}}$$

The confidence interval for S^{qa} is then derived by error propagation calculated with:

$$S^{qa} \pm \sqrt{\left(\varepsilon^{qa}\right)^2 + \left(\varepsilon^{a}\right)^2}$$

For the dE-MAP (related to **Figure 1.4 right**), the array allele cisplatin sensitivity D^{ac} is calculated by pairing observations within biological replicates to minimize spurious differences that can occur between replicates. This is done by:

$$D^{ac} = median(P^{ac} - P^{a})$$

Similarly, the array allele cisplatin sensitivity in the context of a query allele D^{qac} is calculated with:

$$D^{qac} = median(P^{qac} - P^{qa})$$

Errors can be directly calculated using the standard deviation and number of complete observations of paired differences.

Using the two different array strain cisplatin sensitivities, the cisplatin specific genetic interaction between query and array alleles D_{cispt}^{qa} is calculated with:

$$D_{cispt}^{qa} = D^{qac} - D^{ac}$$

Errors can then be propagated as described above given the errors for D^{qac}, and D^{ac}.

Finally (related to **Figure 1.4**), to generate the E-MAP and dE-MAP networks, interactions can be aggregated since a cross of gene deletion X with gene deletion Y should be equivalent to a cross between Y and X

(i.e. the network is an undirected graph). And, since interactions observed across cisplatin concentrations should either be near zero or trend in the same direction (positive or negative) from zero, fitness differences are aggregated using Stouffer's method for combining Z-scores (Stouffer et al. 1949):

$$\frac{\sum Interactions}{\sqrt{n}}$$

This essentially averages the observed interactions, while slightly enhancing the aggregated interaction value if multiple observations fall on the same side of zero. I should note that, while the above analysis pipeline works well for recapitulating 30 years of genetics in DNA repair pathways (i.e. it passes the "sanity" check), scoring interactions for the E-MAP and dE-MAP would likely benefit from more advanced statistical modeling.

1.5 Figures and tables

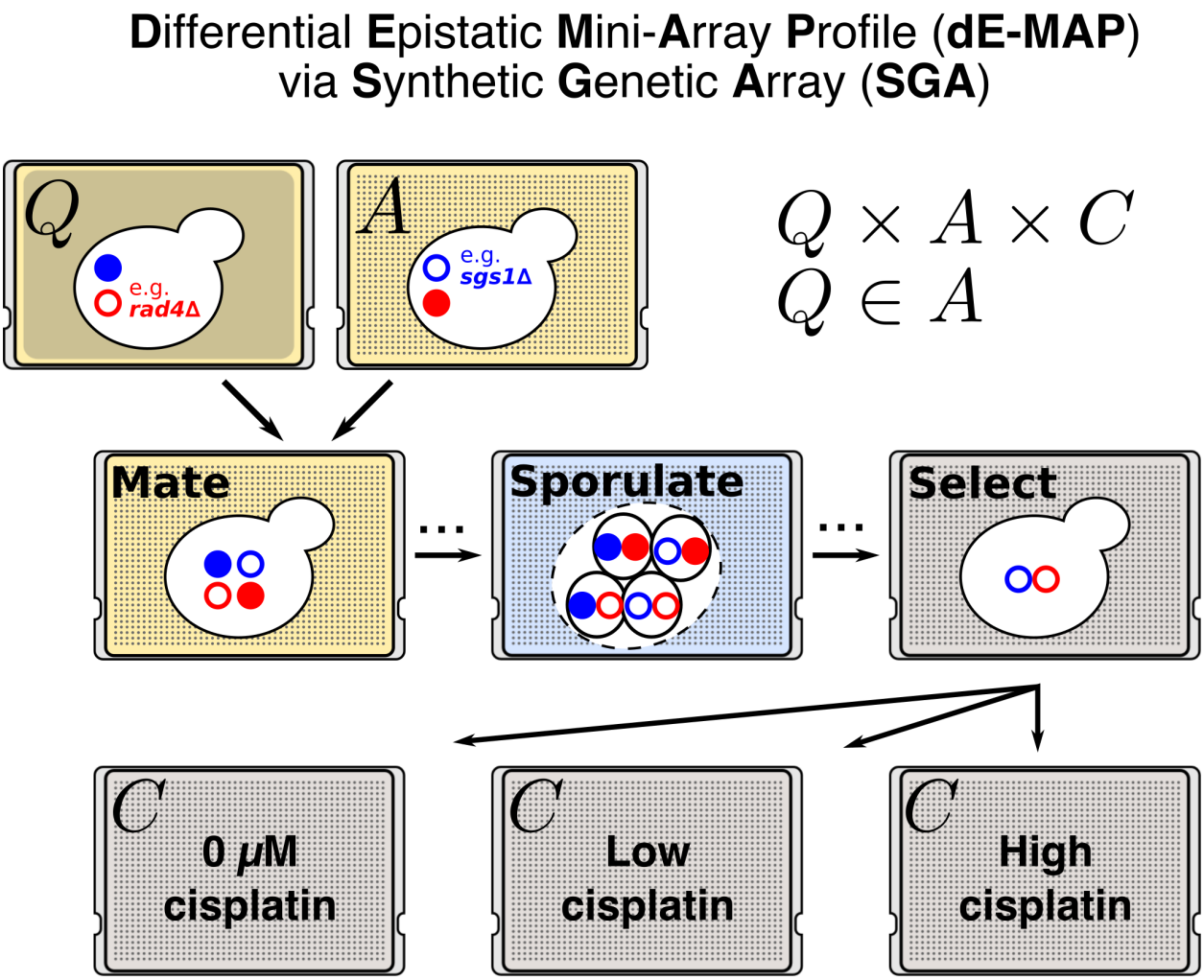

Figure 1.1 – **Differential epistasis profiling.** A differential epistasis mini-array profile (dE-MAP) can be used to detect treatment-specific genetic relationships for a focused set of mutant alleles (A). To generate a dE-MAP, a set of mutant query strains (Q), is crossed to a mini-array of mutant strains (A) (i.e. $Q \times A$). Since the query alleles are all represented in the set of array alleles ($Q \in A$), the complete set of crosses generates all pair-wise double-mutant combinations of alleles in A. In the final step, double-mutants are replicated to three concentrations of cisplatin (C), 0 μM, low, and high (i.e. $Q \times A \times C$). The high concentration of cisplatin is optimized for Q to produce a phenotype in Q as a single mutant. This optimization enhances the likelihood of detecting positive genetic interactions for Q. The low concentration is optimized to prevent total lethality of cisplatin sensitive strains in A, which enhances the likelihood of detecting any interactions with mutants in A that are very sensitive to cisplatin. Related to **Sections** 1.1 **&** 1.2.2.

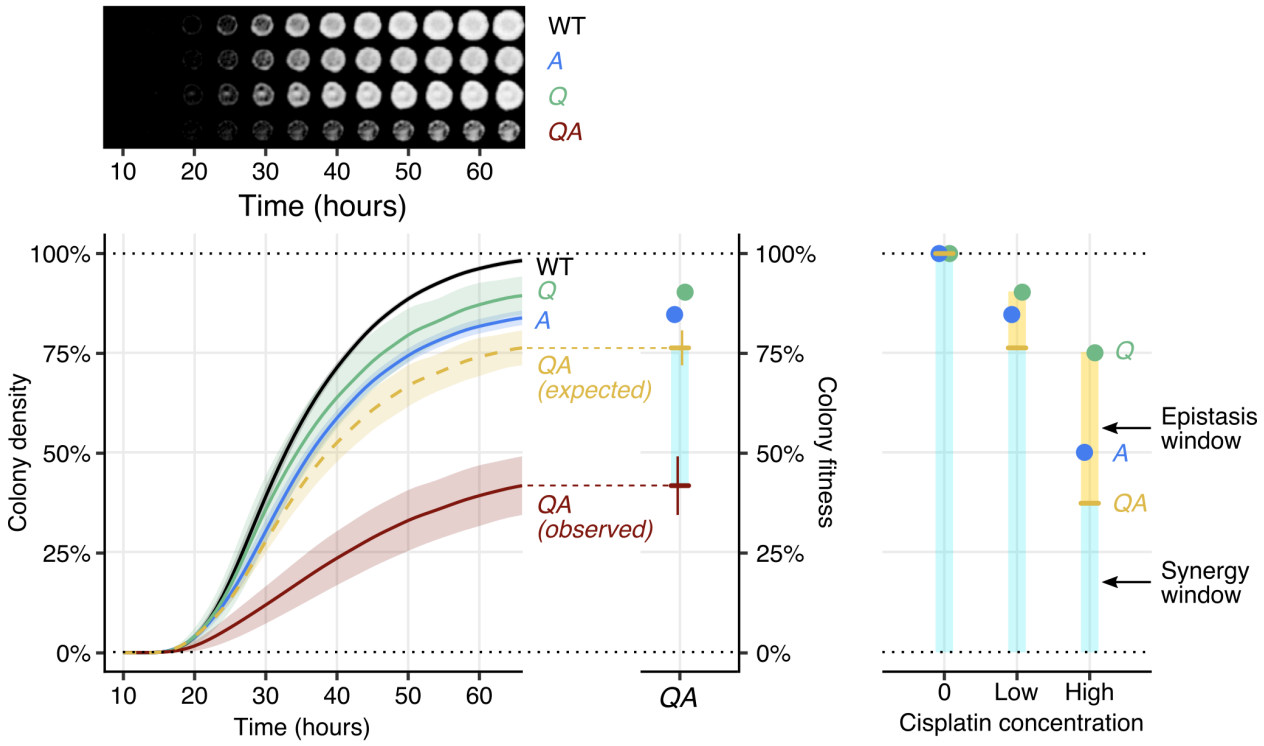

Figure 1.2 – **Optimizing interaction detection.** Colony density at maximal growth is used as a proxy for fitness. Growth of representative colonies for wildtype (WT), a query mutant (Q), an array mutant (A) and a query-array double mutant (QA) are shown with quantitation of colony density over time shown below. In nutrient limiting conditions, colony density plateaus and the relative fitness of colonies remains proportionally stable such that the exact growth time does not have a major impact on relative fitness measurements. Using endpoint growth as an example, Q and A synergize in this instance to have an observed double-mutant fitness (QA, or F^{qa}) considerably lower than expected (red bar). This synergy represents a negative fitness interaction between alleles Q and A. Expected fitness of the double mutant (E^{qa}) is modeled using a multiplicative model (i.e. $E^{qa} = F^{q} \cdot F^{a}$ where $F^{wt} \equiv 1$). Given this model for expected fitness, the windows of opportunity to detect epistasis (yellow box), or synergy (blue box), will critically depend on whether Q, or A have a fitness phenotype (i.e. if $F^{q} \neq 1$, or $F^{a} \neq 1$). On the right is an example showing how the windows for detecting interaction change depending on the phenotype of Q and A. Related to **Sections 1.1 & 1.2.2**.

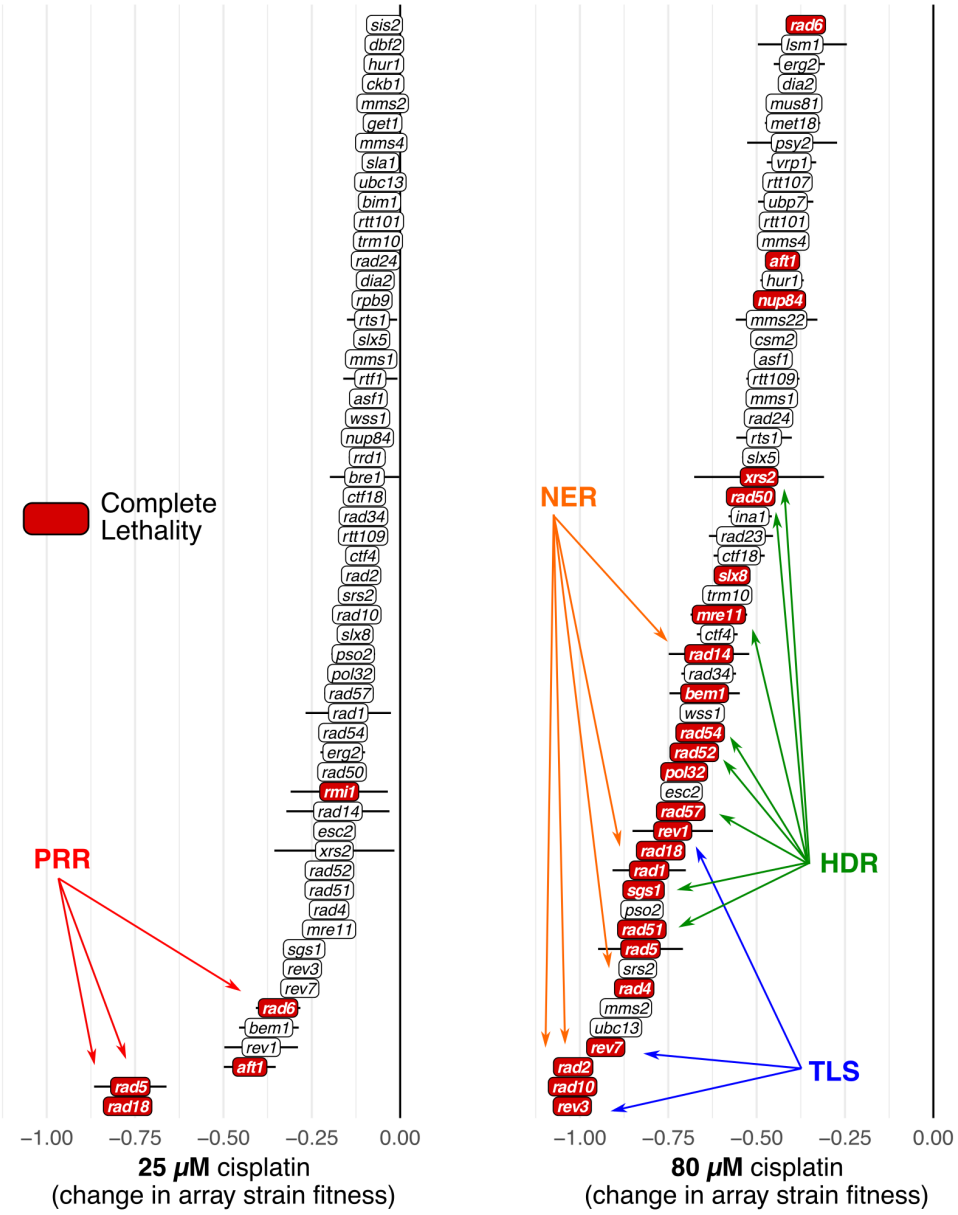

Figure 1.3 – **Single mutant cisplatin sensitivity.** Single mutant strains have major differences in cisplatin sensitivity. The top 56 most cisplatin sensitive mutants at 25 μM are shown in the left panel. The top 56 most cisplatin sensitive mutants at 80 μM are shown in the right panel. Change in fitness at the designated concentration of cisplatin is shown. A negative change in fitness is indicative of cisplatin sensitivity (i.e. a negative fitness interaction with cisplatin). Mutants that have died are shaded red (i.e. no colony growth). Strains are ordered by estimated change in fitness. Horizontal bars are 95% confidence intervals of this estimate. Major repair pathways with significant amounts of lethality are annotated with arrows. PRR = postreplication repair. NER = nucleotide excision repair. TLS = translesion synthesis. HDR = homology-directed repair. Data is from dE-MAP SGA experiments performed by François Dossin, Matthieu Haudiquet, Alice Serafin, Tyler Perdue, Eden David and Eric Bryant. Related to **Table 1.1 & Sections 1.2.3 & 1.2.5**.

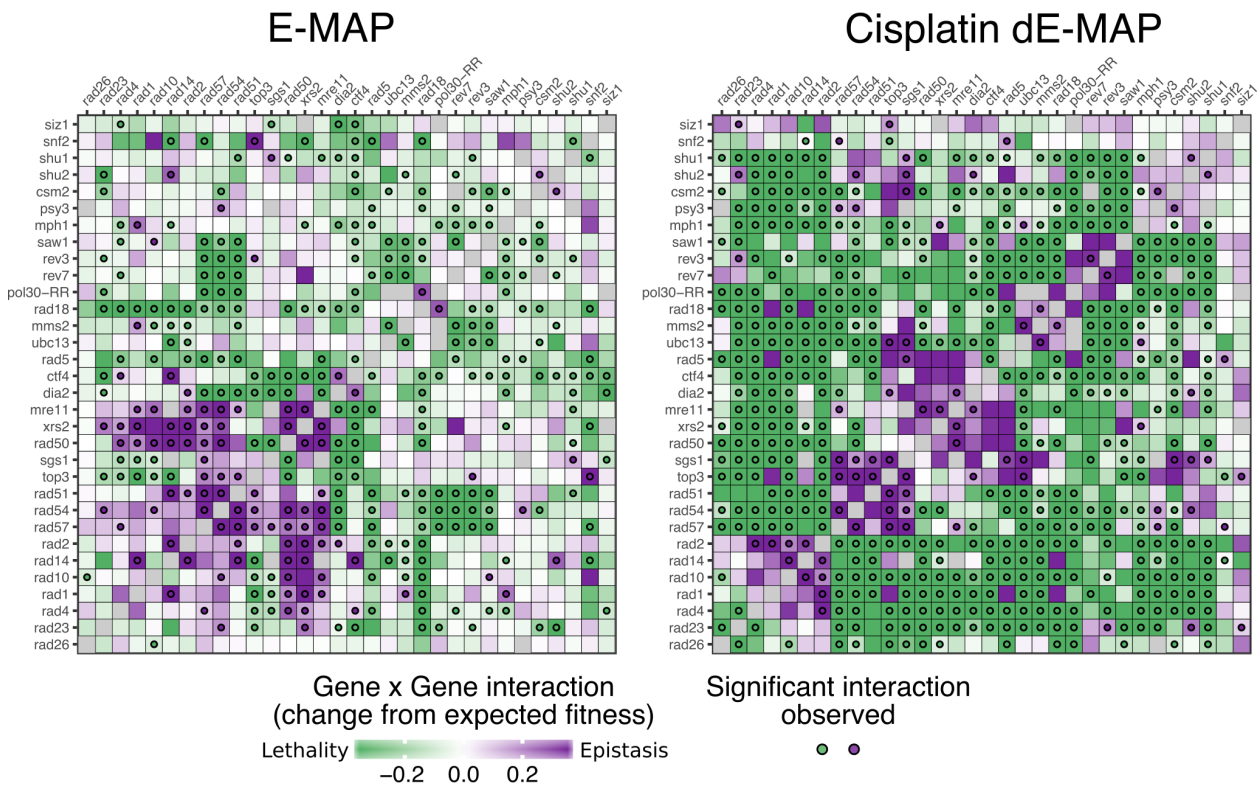

Figure 1.4 – **Epistasis profiles for core cisplatin tolerance pathways.** Boxes represent a double-mutant combination indicated by gene labels on the x and y axes. Boxes are shaded to indicate the change from expected fitness with purple indicating epistasis (i.e. a positive fitness interaction), and green indicating synergy (i.e. a negative fitness interaction). Boxes with circles indicate that a t-distribution based 95% confidence interval of the estimated fitness difference was either above zero (purple), or below zero (green) (see **Methods** 1.4.5). The interaction matrix was mirrored by aggregating undirected interactions between Q and A alleles for all cisplatin concentrations considered (E-MAP, cisplatin = 0μ; dE-MAP cisplatin ≠ 0μ). Data is from SGA experiments performed by François Dossin, Matthieu Haudiquet, Alice Serafin, Tyler Perdue, Eden David and Eric Bryant. Related to **Tables 1.2 & 1.3 and Sections 1.2.4 & 1.2.5**. Genes were clustered manually into pathways with the following order:

1. Transcription-coupled nucleotide excision repair (TC-NER) (*rad26*Δ).
2. Global-genomic nucleotide excision (GG-NER). (*rad4*Δ, *rad23*Δ)
3. Nucleotide excision repair (NER) (*rad1*Δ, *rad10*Δ, *rad14*Δ, *rad2*Δ).
4. Homology directed repair (HDR) (*rad57*Δ, *rad54*Δ, *rad51*Δ).
5. Crossover resolution (*top3*Δ, *sgs1*Δ).
6. MRX (HDR) (*rad50*Δ, *mre11*Δ, *xrs2*Δ).
7. Fork stability / sister chromatid cohesion (*dia2*Δ, *ctf4*Δ).
8. Postreplication repair (PRR) (*rad5*Δ, *ubc13*Δ, *mms2*Δ, *rad18*Δ, *pol30*^{RR}). *pol30*^{RR} is an allele of *POL30* that cannot be ubiquitinated by PRR (Hoege et al. 2002).
9. Translesion synthesis (TLS) (*rev3*Δ, *rev7*Δ). *saw1*Δ has a similar profile to TLS.
10. SHU complex (HDR) (*psy3*Δ, *csm2*Δ, *shu2*Δ, *shu1*Δ). *mph1*Δ has a similar profile to the SHU complex.
11. *snf2*Δ and *siz1*Δ were added for reference as mutants with few, if any, interactions.

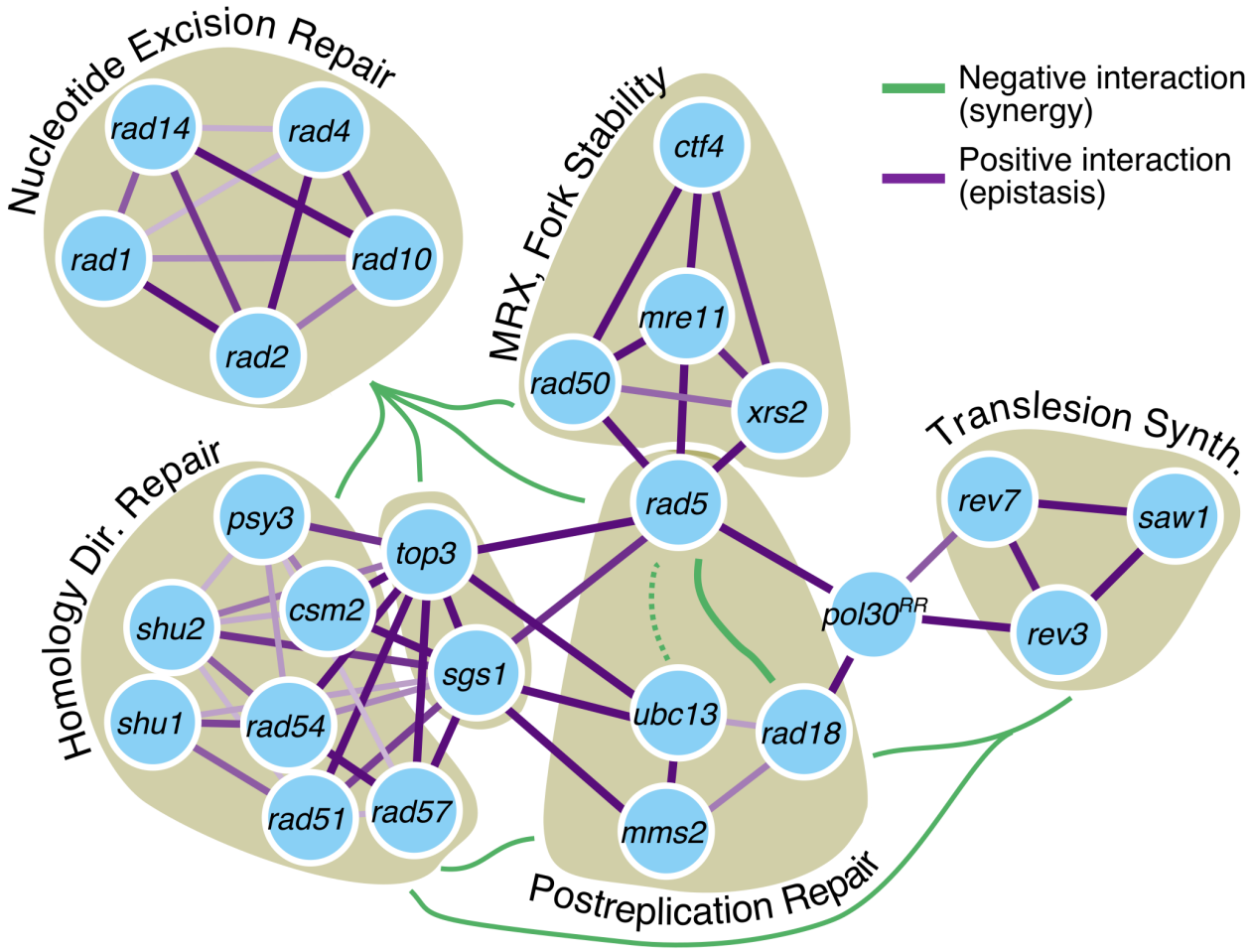

Figure 1.5 – **Summary network of core cisplatin tolerance pathways.** Major pathways from **Figure 1.4** and the general trends in genetic relationships are shown here as a network with blue nodes representing mutant alleles, clustered cisplatin tolerance pathways by tan shaded regions. Purple edges represent shared contribution to cisplatin tolerance between alleles (i.e. same pathway, epistasis, positive fitness interaction). Green edges are drawn between pathways representing parallel contributions to cisplatin tolerance (i.e. closely related redundant pathways, synergy, negative fitness interaction). All pathways are labeled except for crossover resolution (*top3*, *sgs1*). Some edges have been manually trimmed due to noisy underlying colony measurements. Data is from dE-MAP SGA experiments performed by François Dossin, Matthieu Haudiquet, Alice Serafin, Tyler Perdue, Eden David and Eric Bryant. Related to **Section 1.2.5**.

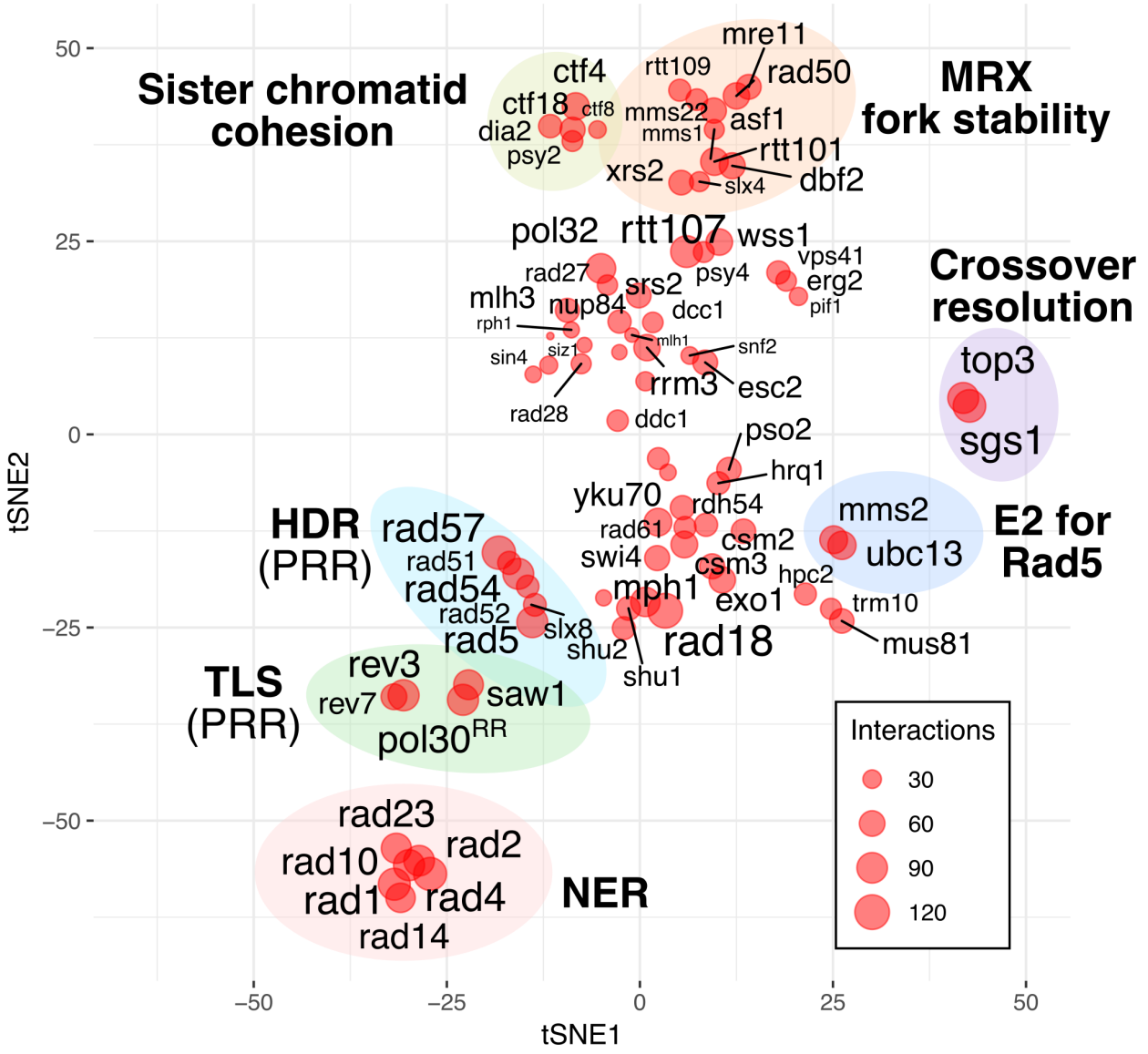

Figure 1.6 – **Clusters of completed query mutants.** Using the dE-MAP for the current complete network, genes were clustered based on their cisplatin specific genetic relationships (only a subset of the complete dE-MAP is presented for **Figures 1.4 & 1.5**). Clustering is presented in two-dimensions, where dimension reduction of the complete network matrix was performed using the t-SNE algorithm (van der Maaten and Hinton 2008). The size of each circle represents the number of interactions detected for a given allele. Some major pathways and complexes are shaded with colored ovals and labeled. Abbreviations stand for homologous recombination (HR), postreplication repair (PRR), translesion synthesis (TLS), and nucleotide excision repair (NER). E2 for Rad5 is labeling the *mms2* and *ubc13* alleles, whose protein products act together as the E2 ubiquitin ligase for Rad5's E3 ubiquitination activity on Pol30. *pol30^{RR}* is an allele of *POL30* that cannot be ubiquitinated by PRR. Data is from dE-MAP SGA experiments performed by François Dossin, Matthieu Haudiquet, Alice Serafin, Tyler Perdue, Eden David and Eric Bryant. Related to **Section 1.2.6**.

Table 1.1 – Cisplatin treated array allele fitness interaction

Header	Description
`A`	Array strain name (*KanMX* disrupted gene name).
`n_a`	Number of measured values for the array strain (*leu2::NatMX* query control, and 0 μM cisplatin control).
`P_a_est`	Mean of center 60% of plate normalized array strain (i.e. estimated fitness of array strain relative to WT).
`ci95_a`	95% confidence interval for the estimated fitness of the array strain.
`C`	Cisplatin concentration (μM).
`n_ac`	Number of measured values for the array strain at the cisplatin concentration, C (*leu2::NatMX* query control).
`P_ac_est`	Mean of center 60% of plate normalized array strain at the cisplatin concentration C.
`ci95_ac`	95% confidence interval for the estimated fitness of the array strain at the cisplatin concentration C.
`P_c_diff`	Difference between array strain fitness with and without cisplatin treatment (`P_ac_est - P_a_est`).
`P_c_diff_ci95`	95% confidence interval for `P_c_diff` calculated by error propagation as the square root of the sum of squared errors (`sqrt(ci95_a^2 + ci95_ac^2)`).
`P_c_diff_upr`	Upper confidence bound (`P_c_diff + P_c_diff_ci95`).
`P_c_diff_lwr`	Lower confidence bound (`P_c_diff - P_c_diff_ci95`).

Related to **Figure 1.3**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/cispt-01.csv.

Table 1.2 – Untreated array and query allele fitness interactions

Header	Description
`A`	Array strain name (*KanMX* disrupted gene name).
`n_a`	Number of measured values for the array strain (*leu2::NatMX* query control, and 0 μM cisplatin control).
`P_a_est`	Mean of center 60% of plate normalized array strain (i.e. estimated fitness of array strain relative to WT).
`ci95_a`	95% confidence interval for the estimated fitness of the array strain.
`Q`	Query strain name (*NatMX* disrupted gene name).
`n_qa`	Number of measured values for the array-query double mutant (0 μM cisplatin control).
`P_qa_est`	Mean of center 80% of plate normalized array strain sizes in the context of the query mutation.
`ci95_qa`	95% confidence interval for the estimated fitness of the array strain in the context of the query mutation.
`P_q_diff`	Difference between array strain fitness with and without query mutation (`P_qa_est - P_a_est`).
`P_q_diff_ci95`	95% confidence interval for `P_q_diff` calculated by error propagation as the square root of the sum of squared errors (`sqrt(ci95_a^2 + ci95_qa^2)`).
`P_q_diff_upr`	Upper confidence bound (`P_q_diff + P_q_diff_ci95`).
`P_q_diff_lwr`	Lower confidence bound (`P_q_diff - P_q_diff_ci95`).

Related to E-MAP in **Figure 1.4**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/cispt-02.csv.

Table 1.3 – Cisplatin treated array and query allele fitness interactions

Header	Description
A	Array strain name (*KanMX* disrupted gene name).
Q	Query strain name (*NatMX* disrupted gene name)
C	Cisplatin concentration (μM).
Pa_n	Number of plate normalized observations of array allele in the *leu2::NatMX* 0 μM cisplatin control conditions.
Paq_n	Number of plate normalized observations of an array-query double mutant in the 0 μM cisplatin control condition.
Pac_n	Number of plate normalized observations of array allele with cisplatin treatment in the *leu2::NatMX* control condition.
Paqc_n	Number of plate normalized observations of an array-query double mutant with cisplatin treatment.
Saq_n	Number of complete, paired within batch, array-query fitness differences (i.e. N for `Saq = Paq - Pa`).
Dac_n	Number of complete, paired within batch, array-cisplatin fitness differences (i.e. N for `Dac = Pac - Pa`).
Daqc_n	Number of complete, paired within batch, array-cisplatin fitness differences in the context of a query mutation (i.e. N for `Daqc = Paqc - Paq`).
n_a	Number of measured values for the array strain (*leu2::NatMX* query control, and 0 μM cisplatin control).
P_a_est	Mean of center 60% of plate normalized array strain (i.e. estimated fitness of array strain relative to WT).
Pa_sd, Paq_sd, Pac_sd, Paqc_sd, Saq_sd, Dac_sd, Daqc_sd	Standard deviations of non-missing measurements of Pa, Paq, Pac, Paqc, Saq, Dac and Daqc respectively.
Pa_ci95, Paq_ci95, Pac_ci95, Paqc_ci95, Saq_ci95, Dac_ci95, Daqc_ci95	As above but for the 95% confidence interval derived from a t-distribution using the following R formula (`(sd / sqrt(n)) * qt(0.975, df = n - 1)`)
Saq	Median array fitness change in query background (i.e. `median(Paq - Pa)`), where Paq and Pa are plate normalized array strain fitness values in the query and untreated control respectively).
Dac	Median array fitness change in cisplatin treated condition. (i.e. `median(Pac - Pa)`), where Pac and Pa are plate normalized array strain fitness values in the cisplatin treated and untreated control respectively).
Daqc	Median array fitness change in query background and cisplatin treated condition. (i.e. `median(Paqc - Paq)`), where Paqc and Paq plate normalized array-query double mutant strain fitness values in the cisplatin treated and untreated control respectively).
Differential	The fitness differential corresponding to `Daqc - Dac`. This value represents the change in an array strain's cisplatin sensitivity caused by having the query mutation in the background (note that wildtype and query fitness and sensitivity to cisplatin is already accounted for during plate normalization, thus this is the genetic interaction between the query and array allele in the context of cisplatin treatment).

Related to dE-MAP **Figure 1.4**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/cispt-03.csv.

Table 1.4 – Complete mirrored E-MAP

Header	Description
`from`	"From" allele in the network
`to`	"To" allele in the network
`any_sig`	Were any of interaction estimates from **Table 1.2** considered "significant" from the confidence interval.
`EMAP_diff`	Aggregated E-MAP differential. When "significant" interactions were present these were aggregated with the following formula: `mean(diff) / sqrt(n)`. When no "significant" interactions were present, all observations were aggregated by the same method.

Related to E-MAP in **Figure 1.4**. Data are aggregated from **Table 1.2**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/cispt-04.csv.

Table 1.5 – Complete mirrored dE-MAP

Header	Description
`from`	"From" allele in the network
`to`	"To" allele in the network
`any_sig`	Were any of interaction estimates from **Table 1.2** considered "significant" from the confidence interval.
`dEMAP_diff`	Aggregated dE-MAP differential. When "significant" interactions were present these were aggregated with the following formula: `mean(diff) / sqrt(n)`. When no "significant" interactions were present, all observations were aggregated by the same method.

Related to dE-MAP in **Figure 1.4**. Data are aggregated from **Table 1.3**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/cispt-05.csv.

Chapter 2

Genomic repercussions of *RAD5* overexpression

Postreplication repair (PRR) enables cells to replicate DNA in the presence of cisplatin and other DNA crosslinking agents that are widely used in the treatment of cancer. In this study, we highlight overexpression of a PRR gene, *HLTF*, as a common event observed in squamous cell, ovarian, and esophageal carcinomas. We elucidate the repercussions of *HLTF* overexpression through the functionally conserved *Saccharomyces cerevisiae* homolog, *RAD5*. Utilizing comprehensive yeast mutant collections, we identify genes essential for tolerating *RAD5* overexpression ($RAD5^{OE}$). We present a genetic landscape enrichment analysis that reveals $RAD5^{OE}$ to cause a requirement for recombination, crossover resolution, and efficient DNA replication. A comparison of the genetic requirements for *rad5*Δ and $RAD5^{OE}$ identifies a specific requirement for crossover resolving genes in $RAD5^{OE}$ cells. These genetic observations are supported by the finding that, while both *rad5*Δ and $RAD5^{OE}$ cause cisplatin sensitivity, only $RAD5^{OE}$ can drive recombination events in a region of repetitive DNA. Remarkably, $RAD5^{OE}$ induced recombination does not require other PRR pathway members, nor does it require the PCNA modification sites involved in PRR regulation. Rather, the $RAD5^{OE}$ phenotype depends largely on a conserved DNA-binding domain implicated in binding 3′ DNA ends. Analysis of DNA replication from a single origin supports a model in which misexpression of *RAD5* causes a bypass of PRR regulatory events resulting in aberrant template switching during DNA replication. Furthermore, our results reveal a possible role for *HLTF* in driving genome instability during carcinogenesis, and suggest platinum-based chemotherapy as an approach for targeting cells with altered PRR.

2.1 Introduction

Postreplication Repair (PRR) is essential for damage tolerance during DNA replication. Cells without this pathway are acutely sensitive to intrastrand DNA crosslinking agents such as cisplatin – one of several platinum-based drugs used in an estimated 50% of all chemotherapeutic regimens (Galanski et al. 2005; Simon et al. 2000; Wu et al. 2004; Kelland 2007). PRR enables damage tolerance during replication either through recruitment of specialized polymerases that can synthesize across damaged bases in a process

called translesion synthesis, or through bypassing lesions by continuing synthesis on undamaged sister chromatid DNA in a process called template switching (reviewed by Boiteux and Jinks-Robertson 2013).

Translesion synthesis and template switching are sometimes referred to as error-prone and error-free PRR respectively owing to high rates of nucleotide misincorporation by translesion synthesis polymerases across from damaged bases (Xiao et al. 2000). However, template switching can cause errors through ectopic recombination, which can drive loss or gain of information in repetitive DNA (Lin et al. 2011; Huang et al. 2002). Moreover, fork stalling and template switching resulting in microhomology-mediated break-induced replication is the proposed driving mechanism behind disease-associated human genome rearrangements (Lee et al. 2007; Zhang et al. 2009). In the absence of PRR, replication forks encountering a DNA lesion are more prone to fork collapse which may also result in a recombinogenic double-stranded DNA break and chromosomal rearrangements (Motegi et al. 2008). Entry into PRR is mediated by monoubiquitination of proliferating cell nuclear antigen (PCNA), performed in *S. cerevisiae* by the Rad6-Rad18 complex (human UBE2A/UBE2B-RAD18) (Ulrich and Jentsch 2000). This initial modification of PCNA is thought to recruit translesion synthesis polymerases to a stalled replication fork. Subsequent regulation of translesion synthesis/template switching pathway choice is then thought to be mediated by the Mms2-Ubc13-Rad5 complex (human UBE2N-UBE2V2-HLTF), which can polyubiquitinate PCNA and perform replication fork remodeling to initiate template switching (Hoege et al. 2002; Branzei et al. 2004).

Given the importance of PRR in responding to lesions caused by commonly used chemotherapeutic drugs, we sought to determine whether this pathway incurs any frequent genetic alterations in cancer genomes. In this study, we perform an analysis of primary tumor samples from The Cancer Genome Atlas (TCGA), which reveals *HLTF* to be commonly amplified and overexpressed in esophageal, uterine and several types of squamous cell carcinoma (Cancer Genome Atlas Research Network 2008). This observation corroborates previous work showing overexpression of *HLTF* at the RNA level in several cancer cell lines and at the protein level in hypopharyngeal squamous cell carcinomas, as well as during early stages of carcinogenesis in a kidney tumor model (Gong et al. 1997; Debauve et al. 2006; Capouillez et al. 2008). Paradoxically, *HLTF* is also found to be frequently silenced by promoter methylation in colorectal cancers (Moinova et al. 2002), leading to conflicting descriptions of *HLTF* as both a tumor suppressor and driver. Ultimately, the role of variable *HLTF* expression in carcinogenesis remains unclear.

Human HLTF and its budding yeast homolog, Rad5, are remarkably conserved in domain structure and biochemical function. HLTF and Rad5 share the eponymous HIP116, Rad5p N-terminal (HIRAN) domain – an ancient region of amino acid sequence homology that has been shown to form an OB-fold structure

which can bind to the 3′ end of single-stranded DNA (ssDNA) (Iyer et al. 2006; Kile et al. 2015; Hishiki et al. 2015; Shin et al. 2018). Downstream of the HIRAN domain, both HLTF and Rad5 have SWI2/SNF2 helicase sequence homology interrupted by a RING finger domain (Johnson et al. 1992; Iyer et al. 2006). Biochemically, both proteins have ATP dependent DNA translocase activity, and are capable of reversing model replication forks *in vitro* (Blastyák et al. 2007, 2010). Additionally, both proteins physically interact with a Rad6-Rad18 counterpart and cooperate with an Mms2-Ubc13 E2 ubiquitin ligase heterodimer to mediate Lys-63-linked PCNA polyubiquitination (Ulrich and Jentsch 2000; Ulrich 2003; Unk et al. 2008). Genetically, *HLTF* is capable of partially complementing *rad5*Δ in ultraviolet radiation (UV) treated yeast, albeit in a translesion synthesis deficient background (Unk et al. 2008).

The *RAD5* gene was initially discovered on the basis of UV sensitivity, and reduced forward mutation rates in forward genetic screens performed in the late 1960s and early 1970s and much progress has been made since in elucidating its roles in PRR (Snow 1967; Cox and Parry 1968; Lemontt 1971; Game and Mortimer 1974). While it is clear that Rad5 contributes to regulating PRR through PCNA modification, the mechanism of PCNA polyubiquitination in promoting template switching is unclear. Furthermore, while many studies have considered the effect of *RAD5* loss of function alleles, little is known about the consequences of *RAD5* overexpression – a genetic alteration frequently observed for *HLTF* in human cancers. In this study, we reveal the genomic repercussions of *RAD5* overexpression ($RAD5^{OE}$) using synthetic dosage interaction (SDI) profiling with the yeast gene disruption and temperature-sensitive mutant collections. We present a genetic landscape enrichment analysis analysis indicating that efficient DNA replication, recombination, and crossover resolution are essential in the context of $RAD5^{OE}$. We demonstrate that *rad5*Δ and $RAD5^{OE}$ are both sensitive to cisplatin treatment. However, $RAD5^{OE}$, but not *rad5*Δ, requires crossover resolution and drives recombination in the rDNA tandem repeat array. Remarkably, we observe no requirement for upstream PRR signaling events on PCNA, and a previously characterized Rad5 RING domain mutant, known to disrupt polyubiquitination activity (Chen et al. 2005), has no effect on the $RAD5^{OE}$ phenotype. We find that a Rad5 ATPase mutant, defective in translocase activity (Blastyák et al. 2007), only partially relieves the consequences of $RAD5^{OE}$, whereas a Rad5 HIRAN point mutation, identified using a homology model of the HLTF ssDNA binding HIRAN domain, nearly eliminates the $RAD5^{OE}$ phenotype. By quantitative physical analysis of DNA replication intermediates, we demonstrate that $RAD5^{OE}$ promotes formation of crossover products near an early firing replication origin. These products persist in the absence of a RecQ helicase involved in crossover resolution and depend on a functional Rad5 HIRAN domain. Our data support a model whereby $RAD5^{OE}$ bypasses PRR regulatory events to promote aberrant template switching via HIRAN domain-mediated fork remodeling. The genomic ramifications of $RAD5^{OE}$ presented here inform

the mechanistic consequences of HLTF misexpression and its potential effect on carcinogenesis, genome instability, and response to platinum-based chemotherapy.

2.2 Results

2.2.1 The *RAD5* ortholog, *HLTF*, is overexpressed in specific cancer types

Deletions of PRR regulating genes in *S. cerevisiae* (i.e. *rad5*Δ, *mms2*Δ, *ubc13*Δ, *rad18*Δ, and *rad6*Δ) cause acute cisplatin sensitivity. Given the importance of PRR in responding to this common chemotherapeutic, we sought to determine whether the corresponding PRR regulating human genes (i.e. *HLTF*, *UBE2V2*, *UBE2N*, *RAD18*, and *UBE2A/B*) are genetically altered in cancer genomes. Using somatic copy number and mRNA expression data from The Cancer Genome Atlas (TCGA), we assessed the frequency of amplification and deletion events in all available cancer cohorts. We observe *UBE2V2* (8q11.21), an ortholog of yeast *MMS2*, to have a mild increase in amplification events (>5% of samples) in 8 different cohorts. Additionally, we observe *RAD18* (3p25.3) amplified in some bladder cancers (9%), and deep deletions of *RAD18* were observed in kidney renal clear cell carcinomas (11%) (**Figure 2.2A & Table 2.1**). Most strikingly, we find *HLTF* (3q24), a human ortholog of yeast *RAD5*, to be commonly amplified in lung (26%), head & neck (14%), and cervical squamous cell carcinomas (12%), as well as ovarian serous cystadenocarcinoma (14%), and esophageal carcinomas (15%) (**Figure 2.1A**). We note that transcript abundance correlates well with gene copy number for all PRR genes considered in this analysis. Overexpression of HLTF protein was previously observed by immunohistochemistry in an independent study of hypopharyngeal squamous cell carcinomas (Capouillez et al. 2008). Since little is known about the consequences of overexpressing either human *HLTF*, or the functionally conserved *RAD5*, we systematically identified genes that become essential when *RAD5* is overexpressed ($RAD5^{OE}$) in the model organism, budding yeast – an approach we previously used to identify cancer-specific pathway dependencies (Reid et al. 2016; Sustic et al. 2018).

2.2.2 $RAD5^{OE}$ causes a requirement for DNA replication and recombination genes

To study the consequences of $RAD5^{OE}$, we constructed a galactose-inducible $RAD5^{OE}$ system on a CEN plasmid. Characterization of Rad5 protein levels by immunoblot indicated an approximately 100- to 500-fold increase in Rad5 abundance compared to endogenous expression levels (**Figure 2.2C**). This observation is also supported by fluorescence intensity comparing an endogenously-tagged Turquoise2-Rad5 (TQ2-Rad5)

with galactose-induced overexpression of TQ2-Rad5 (**Figure 2.2D**). Both endogenous and overexpressed Rad5 localize almost exclusively to the nucleus. These observations are consistent with estimates of Rad5 having low abundance – around 100 Rad5 molecules per cell – in wildtype yeast (Kulak et al. 2014).

We next performed two reverse genetic screens for *RAD5*^{OE} synthetic lethality. For this experiment, we used selective ploidy ablation to pair a *RAD5*^{OE} plasmid with 4,876 non-essential yeast gene disruption mutants, and 525 temperature-sensitive mutants (Reid et al. 2011) – see experimental diagram in **Figure 2.1B**. Fitness effects were scored by comparing growth of arrayed colonies on solid agar for *RAD5*^{OE} compared to an empty-vector control. Reduced colony density in the *RAD5*^{OE} condition relative to the vector control for a given mutant strain is indicative of a *RAD5*^{OE} synthetic dosage interaction (SDI), with extreme examples indicative of synthetic dosage lethality (SDL) (Kroll et al. 1996).

To summarize the genetic interactions of *RAD5*^{OE}, we performed a landscape enrichment analysis. This approach to gene-set visualization takes advantage of the yeast genetic interactome provided by *The Cell Map*, to cluster genes in a two-dimensional landscape using the t-SNE dimension reduction algorithm (Usaj et al. 2017; Costanzo et al. 2016; van der Maaten and Hinton 2008). Clusters in this landscape represent major genetic features and biological processes included in our genetic screens (**Figure 2.1C left**). Using this landscape, we then visualize where *RAD5*^{OE} SDIs cluster. By permutation test, we observe two significant clusters of *RAD5*^{OE} SDI hits ($p < 0.001$) (**Figure 2.1C middle**). The primary cluster highlights DNA replication, DNA damage checkpoint, homology-directed repair (HDR), and crossover resolution as sub-clusters that are essential when *RAD5* is overexpressed (**Figure 2.1C right**). *RAD5*^{OE} SDI screen hits found elsewhere on the landscape are shown in **Figure 2.2B**. Landscape coordinates and interaction scores for all strains are provided in **Tables 2.2, 2.3, 2.4 & 2.5**.

To validate our screen results, we developed a quantitative colony growth assay to determine the magnitude of each genetic interaction (see **Methods 2.4.5** & **Figure 2.3A left**). From the quantitative growth measurements, we observed wildtype strains with *RAD5*^{OE} to grow at approximately 80% of an empty vector control strain (compare WT to *RAD5*^{OE} in **Figure 2.3A left**). Expected growth values for each mutant strain were calculated by multiplying each single mutant growth effect (without *RAD5*^{OE}) to the *RAD5*^{OE} effect in wildtype (yellow bars in **Figure 2.3A**). Observed growth for each mutant with *RAD5*^{OE} and difference from expected growth is shown (red bars and blue columns in **Figure 2.3A right**). Mutants in the 9-1-1 replication checkpoint complexes (*rad17*Δ/*mec3*Δ/*rad24*Δ) and a mutant involved in fork protection (*csm3*Δ) showed only small but reproducible *RAD5*^{OE} SDIs that were difficult to discern in a standard serial-dilution drop assay. Notably, the strongest interactions were observed for HDR and crossover resolution mutants.

Specifically, HDR machinery (e.g. *rad51*Δ, *rad52*Δ and the MRX complex *mre11*Δ/*rad50*Δ/*xrs2*Δ), as well as two crossover resolution complexes (*sgs1*Δ/*top3*Δ and *mms4*Δ/*mus81*Δ) were critically important in *RAD5*^{OE} cells (**Figure 2.3A & 2.3B**).

2.2.3 *RAD5*^{OE} and *rad5*Δ have similar genetic interaction profiles

When analyzing the landscape of *RAD5*^{OE} SDIs, we noted overlap with previously published genetic interactions for *rad5*Δ (Stark et al. 2006). Indeed, we observed 17 shared genetic interactions for *RAD5*^{OE} and *rad5*Δ, all of which are present in the two significant landscape enrichment clusters. This overlap included essentially all of the HDR members of the primary cluster, as well as the replication checkpoint and some DNA replication associated mutants (**Figure 2.4A, green**). There were, however, some revealing differences in genetic interaction profiles. The major differences being that *rad5*Δ uniquely requires nucleotide excision repair (NER) proteins (i.e. Rad4, Rad23, Rad1 & Rad14), Ctf18-RFC complex proteins (i.e. Ctf18, Ctf8 and Dcc1), and translesion synthesis proteins (i.e. Rev3 & Rev7) (**Figure 2.5A & Table 2.6**). Conversely, *RAD5*^{OE} uniquely requires crossover resolution proteins (e.g. Sgs1, Top3, Rmi1, Mms4 & Mus81) (**Figure 2.4A**). To confirm the *RAD5*^{OE} specific requirement for crossover resolution, we measured colony fitness to quantify an interaction with *sgs1*Δ in a *rad5*Δ and *RAD5*^{OE} background. Indeed, we observed no enhanced requirement for *SGS1* in the *rad5*Δ background, and a strong enhanced requirement for *SGS1* in the *RAD5*^{OE} background, which indicates that *RAD5*^{OE} does not simply disrupt Rad5 function (**Figure 2.4B**). Based on these interaction profiles, we hypothesize that *RAD5*^{OE} likely drives the formation of crossovers, possibly by promoting replication fork collapse, or aberrant fork regression and template switching.

2.2.4 *RAD5*^{OE} and *rad5*Δ cause sensitivity to DNA replication fork blocking drugs

RAD5 deletion causes dramatic cisplatin sensitivity, and given the considerable overlap of *RAD5*^{OE} and *rad5*Δ genetic interactions, we reasoned that *RAD5*^{OE} might also sensitize cells to cisplatin treatment. Using the same quantitative spot assay used to measure genetic interactions, *RAD5*^{OE} and *rad5*Δ both show cisplatin sensitivity (50% inhibitory concentrations (IC_{50}) of 75 and 45 μM, respectively) (**Figure 2.4C**). *RAD5*^{OE} can suppress *rad5*Δ cisplatin sensitivity, conferring the same IC_{50} observed when overexpressing *RAD5* in a wildtype background (**Figure 2.5B**). This result again suggests that *RAD5*^{OE} produces functional Rad5 protein, but still causes a limited capacity to resist DNA crosslinks. *RAD5*^{OE} also causes sensitivity to two other replication inhibitors, hydroxyurea (HU), and 4-nitroquinoline 1-oxide (4NQO) (**Figure 2.5C**). However,

RAD5^{OE} is not sensitive to DNA double-strand break (DSB)-inducing ionizing radiation, at a dose sufficient to kill *rad51*Δ cells (**Figure 2.5D,E**). We conclude that cells overexpressing *RAD5* can recover from radiation-induced DSBs, but struggle to tolerate chronic replication stress, indicating *RAD5*^{OE} likely impacts DNA replication.

2.2.5 *RAD5*^{OE} promotes recombination

While *rad5*Δ and *RAD5*^{OE} are both sensitive to DNA replication inhibitors, and share requirements for HDR, *RAD5*^{OE} uniquely requires crossover resolution proteins, suggesting that it actively promotes recombination. To test this notion, we measured direct-repeat recombination within the ribosomal DNA (rDNA) array (Roy and Runge 2000) – see locus diagram in **Figure 2.5F**. *RAD5*^{OE} causes an approximate 9-fold increase in marker loss, which is not observed in *rad5*Δ cells (**Figure 2.4D**). Interestingly, the *RAD5*^{OE}-induced increase in rDNA recombination is not dependent on Fob1, the programmed DNA replication fork blocking protein known to stimulate rDNA recombination (Kobayashi and Horiuchi 1996). While *fob1*Δ cells do show a 7-fold decrease in rDNA recombination as expected, overexpressing *RAD5* in these cells causes an 18-fold increase in rDNA recombination to levels approximately 3-fold higher than wildtype (**Figure 2.5G**). Thus, *RAD5*^{OE} stimulates rDNA recombination independently of the programmed rDNA replication fork block.

2.2.6 Other PRR genes are not essential for *RAD5*^{OE} induced recombination

Rad5 is an E3 ubiquitin ligase whose RING domain is essential for establishing a physical relationship with its E2 partner – the Mms2-Ubc13 heterodimer (Ulrich and Jentsch 2000). Together, Mms2-Ubc13 & Rad5 can polyubiquitinate Pol30 (PCNA) at lysine residues 127 and 164 (Hoege et al. 2002). This polyubiquitination occurs during PRR downstream of initiating monoubiquitination events mediated by the Rad6 & Rad18 E2/E3 ubiquitin ligase complex. Given the well-studied role of Rad5 in regulating PRR pathway choice, we suspected that the phenotype of *RAD5*^{OE} would depend on upstream PCNA monoubiquitination and its Mms2-Ubc13 E2 partner. First, we examined cisplatin sensitivity of *RAD5*^{OE} in cells that cannot ubiquitinate Pol30 at the critical lysine residues 127 and 164 using the *pol30*^{K127R,K164R} (*pol30*^{RR}) and *rad18*Δ mutant backgrounds. We observed overlapping contributions of *RAD5*^{OE} and *pol30*^{RR}/*rad18*Δ with respect to cisplatin sensitivity (**Figure 2.7A**). In fact, *RAD5*^{OE} mildly suppressed the cisplatin sensitivity of *rad18*Δ. Additionally, in *pol30*^{RR} and *rad18*Δ backgrounds, the effect of *RAD5*^{OE} on rDNA recombination was reduced but not eliminated (**Figure 2.7B**). Thus, *RAD5*^{OE} at least partially bypasses upstream PRR signaling events.

Importantly, neither *pol30*^{RR} nor *rad18*Δ alleviated the *RAD5*^{OE} requirement for the crossover-resolution gene, *SGS1* (**Figure 2.6A**). Next, we tested the requirement for the Rad5 E2 partner by examining the *RAD5*^{OE} phenotype in an *mms2*Δ *ubc13*Δ background. Strikingly, absence of its E2 partner did not suppress the requirement for *SGS1*, cisplatin sensitivity, or increased rDNA recombination caused by *RAD5*^{OE} (**Figure 2.6**). Clearly, the Rad5 E2 partner, Mms2-Ubc13, which is essential for PCNA polyubiquitination during PRR, is not required for the *RAD5*^{OE} phenotype.

2.2.7 The conserved HIRAN domain is essential for *RAD5*^{OE} genetic interactions

To gain mechanistic insight into the consequences of *RAD5* overexpression, we sought to determine which domains within Rad5 were necessary to convey the *RAD5*^{OE} phenotype. The *RAD5* gene encodes a protein containing three primary domains, (1) a *SNF2* helicase domain that is essential for Rad5 to reverse model replication forks *in vitro*, (2) a RING domain that mediates a physical interaction with Ubc13 and is essential for PCNA polyubiquitination, and (3) a HIRAN domain, which is a conserved N-terminal region that, in *HLTF*, forms an OB-fold which can bind the 3′ end of DNA (**Figure 2.8A**). The helicase and RING domains each have well characterized separation of function alleles. Two such alleles are the ATPase defective *rad5*^{DE681,682AA} (*AA*) mutation in the *SNF2* Walker B motif, and the ubiquitination defective *rad5*^{I916A} (*RING*) mutation (Gangavarapu et al. 2006; Blastyák et al. 2007; Chen et al. 2005; Ulrich 2003).

To dissect the relative contributions of *RAD5* domains to the *RAD5*^{OE} phenotype, we generated overexpression constructs for the *AA* and *RING* mutations, as well as three N-terminal deletions designed to separate the contribution of a poorly conserved N-terminal region from the more conserved HIRAN domain (*N164*Δ, *173-281*Δ and *N293*Δ – see **Figure 2.8A**). We crossed each overexpression construct into a mini-array of genetic backgrounds representing 25 of the strongest *RAD5*^{OE} SDI mutant strains. The resulting genetic interaction scores were used to calculate SDI profile distances that are visualized in the clustering dendrogram of **Figure 2.8C**. As expected from our experiments in the *mms2*Δ *ubc13*Δ background, overexpressing the *RING* domain mutant resulted in an interaction profile that was nearly indistinguishable from *RAD5*^{OE} (**Figure 2.8C - *RAD5* cluster**). The AA ATPase mutant partially alleviated many of the *RAD5*^{OE} genetic requirements (e.g. reduced requirement for *SGS1*, **Figure 2.8D**), and cluster along with several *RAD5* mutant alleles that have a weak effect on function (**Figure 2.8C - Intermediate cluster**). Among this group is the rad5^{G535R} – a mutation adjacent to the *SNF2* Walker A motif that is present in the original W303 laboratory strain (Fan et al. 1996; Elserafy et al. 2018). Strikingly, when we truncated the *RAD5* N-terminus up to the HIRAN domain (*N164*Δ) we observed no effect on the interaction profile, but when we further truncated

beyond the HIRAN domain (*N293*Δ), or deleted just the HIRAN domain (*173-281*Δ), we saw near complete elimination of all 25 *RAD5*^{OE} genetic interactions (**Figure 2.8C - Vector cluster**). We conclude that the HIRAN domain is necessary for the *RAD5*^{OE} interactions.

2.2.8 The conserved HIRAN residue, R187, is integral to the *RAD5*^{OE} phenotype

While the HIRAN domain has been characterized for human *HLTF*, less is known about the role of this domain in yeast. To predict structural features of the Rad5 HIRAN domain, we used SWISS-MODEL to perform an unsupervised homology search using the Rad5 HIRAN domain amino-acid sequence (Guex et al. 2009; Bienert et al. 2017; Waterhouse et al. 2018). The top hits included three independent crystal structures of the *HLTF* HIRAN domain, which were then used to construct a structural homology model of the Rad5 HIRAN domain (**Figure 2.8B**) (PDB IDs: 5BNH, 4XZF & 4S0N from Neculai et al. (2015), Hishiki et al. (2015) & Kile et al. (2015)). A visual comparison of an HLTF HIRAN domain structure and the Rad5 HIRAN homology model is shown in **Figure 2.9A**. The predicted yeast structure has a much larger loop between the predicted β2 and β3 strands, which forms part of the putative 3'OH ssDNA binding pocket. A sequence alignment among yeast species showed considerable variability in sequence and length for this region, hence we refer to it as the "divergent loop" (**Figure 2.8B & 2.9A,B**). As expected, given the low sequence conservation and variable length of this region, no alanine substitution tested in the divergent loop had any effect on the *RAD5*^{OE} interaction profile, and deleting the portion of the loop that does not face the binding pocket (*204-214*Δ) also had no effect. However, deleting the portion of the loop facing the binding pocket (*215-225*Δ) had an intermediate effect, and deleting the whole loop (*204-225*Δ) had a strong effect, eliminating most of the *RAD5*^{OE} genetic interactions (**Figure 2.8C**). Therefore, like HLTF, this loop is important for HIRAN domain function.

Examination of the HIRAN domain sequence alignment reveals four highly conserved positively charged residues, three of which face the putative ssDNA binding pocket (R187, R229 and R241) – see **Figure 2.8B and 2.9A,B**. Remarkably, every alanine and/or acidic substitution tested for these three residues in SDI profiling had at least an intermediate effect (R187A, R229A, R229E and R241E) and the R187E substitution alone had a strong effect (**Figure 2.8C**). R187E resembles HLTF^{R71E} in the homology model, which is the HIRAN mutation that had the strongest biochemical effect on ssDNA binding and *in vitro* fork regression observed by Kile et al. (2015). Importantly, the R187E substitution does not cause major changes in protein expression (**Figure 2.9C**).

We next compared *R187E*, *AA*, and *RING* mutants to determine their impact on other important features of

the *RAD5*^{OE} phenotype. With respect to the *SGS1* requirement, *R187E*^{OE} showed the largest improvement in colony fitness, supporting a role for the HIRAN domain in promoting recombination (**Figure 2.8D**). We then integrated the *R187E*, *AA*, and *RING* alleles at the endogenous locus to compare cisplatin sensitivity at normal expression levels. *AA* and *R187E* showed equivalent cisplatin sensitivity (**Figure 2.8E left**), demonstrating that the HIRAN domain is important for damage tolerance at endogenous expression levels. However, cisplatin sensitivity of *R187E* can be dosage suppressed (**Figure 2.8E right**), which is consistent with a proposed model of the HIRAN domain physically interacting with 3'-ends at stalled replication forks (Kile et al. 2015; Shin et al. 2018). Lastly, when we measure rDNA recombination rates, we observe a pattern similar to that observed from genetic interaction profiling. *RAD5*^{OE} and *RING*^{OE} have identical rDNA recombination frequencies, whereas *AA*^{OE} partially, and *R187E*^{OE} strongly suppresses rDNA recombination (**Figure 2.8F**). Taken together, these results clearly indicate that the conserved HIRAN residue R187, and thus its DNA-binding activity, plays a major role in *RAD5*^{OE} genetic interactions, cisplatin tolerance and recombination.

2.2.9 *RAD5*^{OE} drives X-shaped DNA accumulation during replication

The above genetic observations, and the known requirement for Rad5 in PRR template-switching, support a mechanism in which excess Rad5 is capable of driving recombination during DNA replication, resulting in intermediates that are resolved by *SGS1*. Additionally, the analysis of Rad5 separation of function mutants predicts that this mechanism would depend on the 3′-end ssDNA binding HIRAN domain. Given this model, we used two-dimensional (2D) gel electrophoresis to physically analyze the accumulation of recombination intermediates that occur in cells misexpressing *RAD5* during DNA replication. Wildtype or *sgs1*Δ mutant yeast strains containing an empty vector control or *RAD5*^{OE} vectors were cell cycle synchronized by arrest in G1 phase followed by release into S phase. Expression was induced during the course of alpha pheromone treatment so that Rad5 would accumulate before synchronous release (**Figure 2.11A**). Hydroxyurea (HU) was added to cultures at the time of release to deplete cellular dNTP pools and slow DNA replication resulting in cells with less than a G2 DNA content even at 120 minutes after release into S phase (**Figure 2.11B,C**). Digested DNA was separated on a 2D gel and probed to detect a 5.1 Kbp DNA fragment containing the early firing replication origin *ARS305* in the middle of the fragment (**Figure 2.11D**). 2D gel electrophoresis separates DNA by mass in the first dimension and by shape in the second, resulting in characteristic migration patterns depending on the orientation of replication through the probed DNA fragment (**Figure 2.10A**). The centrally located *ARS305* initiates bidirectional replication resulting in a bubble arc

signal in wildtype cells containing the vector control at 30 and 60 minute timepoints which dissipates at later times as replication proceeds beyond the limits of the DNA fragment (**Figure 2.10B**). Slowing replication by release into HU induces template switch recombination in PRR resulting in **X**-shaped molecules in the replicating DNA observed as diagonal spikes extending from the 2N spot (**X**-spike) – see **Figure 2.10A,B**. Wildtype cells with the *RAD5*^{OE} vector show an increase in the **X**-spike signal at 60 minutes compared to the vector control. The **X**-shaped DNA structures are largely resolved by the 90-minute time point and are not present at 120 minutes. Mutant *sgs1*Δ strains have a reduced ability to resolve **X**-shaped replication intermediates (Liberi et al. 2005). Consistent with this model, *sgs1*Δ cells with *RAD5*^{OE} accumulate **X**-shaped DNA to higher levels than wildtype and fail to resolve these structures even 120 minutes after release from arrest (**Figure 2.10C**). Quantification of **X**-spike signal difference between *RAD5*^{OE} and vector control shows that *RAD5*^{OE} induces formation of **X**-shaped DNA and that this structure accumulates in the absence of Sgs1 (**Figure 2.10E, top**).

In addition to an increase in **X**-shaped DNA, we noted an accumulation of **Y**-arc DNA in samples with *RAD5*^{OE} (**Figure 2.10B & C**). The **Y**-arc DNA was increased to a similar extent in both wildtype and *sgs1*Δ mutant cells at the 30 and 60 minute timepoints indicating that the effect is not dependent on *SGS1*. We also note that the excess **Y**-arc DNA is largely absent in later samples. An increased **Y**-arc signal suggests that replication from *ARS305* becomes asymmetric in cells with *RAD5*^{OE} giving a transition from a bubble to a **Y**-arc signal and is evidence for random pausing of replication forks in the *RAD5*^{OE} samples. Quantification of **Y**-arc signal difference from multiple experiments is shown in **Figure 2.10E, bottom**.

We next tested the effect of the *R187E* HIRAN domain mutation on replicating DNA. This mutant was overexpressed in *sgs1*Δ cells and compared to the vector control and *RAD5*^{OE} (**Figure 2.10C**). *sgs1*Δ samples from 60-minutes post release show the largest accumulation of **X**-shaped DNA with *RAD5*^{OE} but do not show this increase with *R187E*^{OE} (**Figure 2.10D & 2.11E,F**). Likewise, **Y**-arc signal in the *sgs1*Δ cells is maximal at 60 minutes, but no significant accumulation is observed in the *R187E*^{OE} samples. Comparison of **X**-shape and **Y**-arc signals to vector control samples is shown in **Figure 2.10E, right**. Thus, the *R187E* mutation suppresses the effect of *RAD5*^{OE} on replicating DNA. Given the conserved position of this residue in the HIRAN domain, these results underscore the importance of DNA 3′-end binding for the *in vivo* mechanism of template switch or fork reversal during DNA replication.

2.3 Discussion

Post replication repair (PRR) is an essential pathway for DNA damage tolerance during replication in mitotic cells. As such, activity and dysregulation of this pathway may have important implications for actively dividing tumor cells challenged with DNA replication-inhibiting chemotherapies. Here we highlight a frequent copy number amplification event in ovarian, esophageal, and squamous cell carcinoma that results in overexpression of HLTF – a large multifunctional protein implicated in multiple aspects of PRR (**Figure 2.1A**). *HLTF* is a well-conserved gene that shares many common features with *S. cerevisiae RAD5*, including PCNA polyubiquitination activity, ATP dependent DNA translocase activity, physical interactions with a Rad18 counterpart, and an ancient region of conserved sequence known as the HIRAN domain, important for binding ssDNA 3' ends. Given this striking conservation of function, we have modeled the potential genomic repercussions of *HLTF* overexpression through a systematic genetic characterization of *RAD5* overexpression ($RAD5^{OE}$) in budding yeast.

Genetic interaction profiling and landscape enrichment analysis of $RAD5^{OE}$ revealed a clear requirement for DNA replication, homology directed repair, and crossover resolution factors (**Figure 2.1B,C & 2.3**). Many of these genetic requirements are shared between $RAD5^{OE}$ and *rad5*Δ (**Figure 2.4A**). Accordingly, both $RAD5^{OE}$ and *rad5*Δ cause cisplatin sensitivity, underscoring the importance of Rad5 regulation in maintaining DNA damage tolerance (**Figure 2.4C**). From the comparison of $RAD5^{OE}$ and *rad5*Δ genetic interactions, we identified the requirement for crossover resolution genes as specific to $RAD5^{OE}$ (**Figure 2.4A,B**). In line with this observation, $RAD5^{OE}$, but not *rad5*Δ, results in nearly an order of magnitude increase in rDNA recombination demonstrating that excess Rad5 promotes genome instability (**Figure 2.4D**).

The $RAD5^{OE}$ genetic interaction and damaging agent sensitivity profiles support a model of $RAD5^{OE}$ acting during replication – an observation consistent with the known role of Rad5 in PRR. While we expected the $RAD5^{OE}$ phenotype to require other members of PRR, this requirement was not observed. For example, the Rad5 E2 partner (Mms2-Ubc13) is completely dispensable for the *SGS1* genetic requirement, cisplatin sensitivity, and rDNA recombination (**Figure 2.6**). Similarly, a well-characterized RING domain mutation in Rad5 fails to suppress the $RAD5^{OE}$ phenotype (**Figure 2.8**). Additionally, in *rad18*Δ, and $pol30^{RR}$ backgrounds, which lack canonical PRR associated PCNA modification, $RAD5^{OE}$ still causes a requirement for *SGS1*, and increases rDNA recombination (**Figure 2.63A & 2.7B**). Thus, we reason that cells misexpressing *RAD5* bypass PRR signaling events usually required to mediate template switching.

We next investigated other functional aspects of the Rad5 protein and identified R187E as a single residue

substitution that nearly eliminates the *RAD5*^{OE} phenotype (**Figure 2.8****4**). Notably, based on primary sequence alignment and structural homology modeling, this mutation closely resembles HLTF^{R71E}, which has a strong effect on 3'OH dependent ssDNA affinity and *in vitro* fork regression activity (**Figure 2.9A,B**) (Kile et al. 2015). Analysis of replication intermediates by 2D gel electrophoresis confirmed that *RAD5*^{OE} drives X-shaped DNA formation during replication near an early firing origin, and that these structures accumulate in an *sgs1*Δ background (**Figure 2.10**). Importantly, X-shaped DNA accumulation during replication is severely reduced, if not eliminated, when overexpressing the *R187E* mutation, emphasizing the importance of the 3'-end binding HIRAN domain for generating crossovers during replication (**Figure 2.10D,E**).

2.3.1 Model for Rad5 action

Any model for the genomic outcomes of *RAD5*^{OE} must account for recombination at a replication fork. One possible model is that Rad5 binding at a replication fork induces collapse leading to dsDNA breaks (DSBs) followed by repair via break-induced replication (BIR) (**Figure 2.12 model 1**). BIR involves a single-ended DNA break invading homologous sequence to establish a replication fork. This process has been implicated in alternative maintenance of telomeres and may drive complex genome rearrangements (Llorente et al. 2008; Sakofsky and Malkova 2017). However, we do not see evidence for DSBs with *RAD5*^{OE}. In **Figure 2.10**, we observe replication proceeding as a replication bubble with some conversion to Y-arc by *RAD5*^{OE}. Breakage occurring in a replication bubble should produce smaller asymmetric Y-arcs that would resolve below the full-length Y-arc signal, and these are not observed (Noguchi et al. 2003). Furthermore, the genetic requirements for BIR do not match the genetic interaction profile of *RAD5*^{OE}. BIR that is induced from single-ended dsDNA breaks outside of the context of DNA replication requires the Pif1 helicase and the DNA polymerase delta subunit Pol32 (Saini et al. 2013; Wilson et al. 2013). In contrast, BIR induced at a replication fork does not require Pif1 or Pol32 (Mayle et al. 2015). *RAD5*^{OE} requires Pol32, but not Pif1 (**Figure 2.13**), which does not fit either model of BIR. We therefore think that fork breakage is unlikely to account for effects that we see with *RAD5*^{OE}.

A second possible model, which involves template switching is shown in **Figure 2.12** (**model 2**). In this model, Rad5 binding at a stalled replication fork results in reprimed leading strand synthesis leaving a single-strand gap behind the replication fork (Marians 2018). The ssDNA gap would then be processed by HDR proteins to generate crossovers. This model accounts for the accumulation of X-spikes in replicating DNA and the genetic requirement for HDR and crossover proteins in *RAD5*^{OE} cells. However, this model does not account for the accumulation of Y-arc DNA as repriming would allow continuation of the replication

fork. In addition, this model does not involve Rad5 translocase activity, which is responsible for at least half of the increased recombination we observe in rDNA. Thus, recombination in ssDNA gaps cannot fully account for our observations.

Therefore, we favor the fork capture and reversal model because it accounts for every aspect of the *RAD5*^{OE} phenotype (**Figure 2.12 model 3**). In this model, the Rad5 HIRAN domain captures a replication fork. The DNA is reorganized such that the two nascent strands can anneal (fork capture) and, when coupled with a translocase activity, leads to extrusion to form a 4-way junction (fork reversal). Both Rad5 and HLTF exhibit this activity *in vitro* on fork-like oligonucleotide substrates, and both ATPase activity and the ssDNA-binding HIRAN domain are important for this function (Blastyák et al. 2007; Kile et al. 2015; Shin et al. 2018). HIRAN domain-mediated capture of a 3' DNA end has been shown to form an initial 3-way junction (Chavez et al. 2018; Shin et al. 2018), and stabilization of this junction would explain the accumulation of Y-arc DNA in 2D gels (**Figure 2.10**). This Y-arc DNA accumulation is absent when overexpressing the *R187E* HIRAN mutant. The 3-way junction would be converted to a 4-way junction aided either by Rad5 translocase activity, or other branch migration mechanisms, which explains the partial requirement for Rad5 translocase activity. This reversed fork structure is likely transient and is quickly resolved by HDR and crossover resolution to re-establish the replication fork, explaining the requirement for HDR and crossover resolution genes.

While we favor the third model, it is possible that multiple pathways are invoked upon Rad5 interaction with a replication fork. For instance, the Mus81-Mms4 structure-specific nuclease complex contributes to cell viability in the context of *RAD5*^{OE} and a simple model is that it functions redundantly with Sgs1. On the other hand, different resolvase activities may be required in different genomic contexts such as repetitive DNA or near telomeres.

In conclusion, *RAD5* overexpression causes hyperrecombination during DNA replication, which can result in genome instability. This observation suggests that human cells with amplified and overexpressed *HLTF* may be prone to similar genome instability – a hallmark of carcinogenesis. Importantly, we also show that *RAD5* dysregulation results in impaired cisplatin tolerance. Our studies suggest that key recombination and repair pathways may become important in cancers harboring *HLTF* copy number alterations. In view of the widespread use of platinum-based chemotherapy in the treatment of cancer, our observations encourage further study into the effect of dysregulated *HLTF* on DNA damage tolerance.

2.4 Methods

2.4.1 Strains and plasmids

All strains are *RAD5* derivatives of W303 (Thomas and Rothstein 1989), unless otherwise stated. Library strains used in our genetic screens are derivatives of S288C. Strains and plasmids constructed for this study, or used in validation experiments, were verified by Sanger sequencing. All genotypes for strains and plasmids used in this study and primers used for construction and validation are listed in **Tables 2.7 & 2.8**.

2.4.2 Cancer genomics

Copy number and mRNA expression data for primary tumor samples originate from The Cancer Genome Atlas (https://cancergenome.nih.gov). These data were accessed via the Memorial Sloan Kettering computational biology portal R package (http://www.cbioportal.org) which provides access to TCGA data processed using the Broad institute GDAC firehose pipeline (https://gdac.broadinstitute.org) (Cancer Genome Atlas Research Network 2008; Jacobsen and Questions 2018; Cerami et al. 2012; Gao et al. 2013). Gene level threshold GISTIC scores were used to classify the predicted ploidy of genes in a given tumor sample with a score of 2 representing gene copy number amplification. Median expression Z-scores were used to assess increases in mRNA transcript abundance for tumors with a given gene amplified (GISTIC = 2) as compared to diploid tumors (GISTIC = 0). An expression Z-score of 0 indicates no change in transcript abundance relative to diploid tumors, whereas an expression Z-score of 2 indicates two standard deviation higher mRNA transcript abundance. Summaries for each PRR gene considered in this study are provided in **Table 2.1**.

2.4.3 Genome-wide identification of SDL interactions

Selective ploidy ablation was used as described by Reid et al. (2011) to transfer galactose-inducible overexpression plasmids into a collection of (1) 4,999 *MAT***a** KanMX gene disruption strains representing individual deletions of 4,899 genes, and (2) a collection of 1,920 *MAT***a** temperature-sensitive strains representing 476 essential genes. Plates were scanned on a Mikrotek ScanMaker 9800XL plus flatbed scanner with transparency cover. Background subtracted colony densities were measured using the screenmill software package available on GitHub (https://github.com/ericedwardbryant/screenmill). Plates were manually reviewed to exclude regions of pinning failure resulting in 4,876 and 363 genes from the non-essential gene deletion

and essential temperature-sensitive collections respectively. For the *MAT***a** screen, colony measurements were normalized to the plate median. For the temperature-sensitive screen, colony measurements were normalized to the mean of healthy strains on each plate at a given temperature. Normalized colony measurements were then compared via growth ratio to an empty vector control screen performed in parallel to each *RAD5*^{OE} screen. Scores for all strains considered in these screens are provided in **Tables 2.2,2.3 & 2.4**.

2.4.4 Landscape enrichment analysis

Landscape enrichment analysis was performed using genetic interaction profile correlations from The Cell Map (Usaj et al. 2017; Costanzo et al. 2016). These correlations were filtered to include only genes that were considered in our SDL screens. An additional filter was applied to remove genes with more than 30% missing values. Remaining missing values were imputed as 0 (i.e. no correlation assumed). Dimension reduction was then performed using t-SNE at the default perplexity of 30 (Krijthe 2015; van der Maaten and Hinton 2008). The resulting two-dimensional matrix was then used to visualize the genetic landscape of *S. cerevisiae* where local clustering represents groups of genes with similar patterns of genetic profile correlations. *MAT***a** gene disruption SDL hits (Z-score < -2) or the top 50 temperature-sensitive SDL hits were then visualized on the landscape to identify clusters. Clusters were assessed for significance by permutation testing. This test involved repeatedly sampling the same number of SDL and TS strains from the landscape to determine the empirical probability of observing a given local density. To encourage further exploration of the landscape, all coordinates are included in **Table 2.5**.

2.4.5 Colony growth and fitness measurements

Overnight liquid cultures were grown to saturation in appropriate selective media. Saturated cultures were equilibrated to an OD_{600} of 1.0, arrayed in a 96 well microtiter plate, and pinned 3 times at 384 colony density (i.e. 4 technical replicates for each position of the microtiter plate) using a Singer RoToR robot. Strains were grown on synthetic complete (SC) solid agar media with 2% galactose and appropriate nutrient drop-out to maintain plasmid selection. Solubilized drugs (e.g. cisplatin in 0.9% NaCl), were added immediately before pouring plates. Plates with pinned colonies were incubated at 30°C for 65 hours. For growth curves, colonies were scanned hourly. Colony density was measured using screenmill. Expected fitness of a double mutant, ab (E^{ab}) was calculated as the multiplicative fitness contributions of each single mutant (F^{a}, and F^{b}) scaled

to fitness of wildtype ($F^{wt} \equiv 1$).

$$E^{ab} = F^a \cdot F^b \pm \varepsilon^{ab}$$

Error in expected fitness (ε^{ab}) was computed by propagating error from estimates of F^a, and F^b using the equation below. Care was taken to minimize systematic bias in experiments (e.g. by distributing strains evenly throughout the 96 well plate to minimize position and neighboring strain effects).

$$\varepsilon^{ab} = E^{ab} \cdot \sqrt{\left(\frac{\varepsilon^a}{F^a}\right)^2 + \left(\frac{\varepsilon^b}{F^b}\right)^2}$$

Dose response models for cisplatin experiments were fit to a logistic curve using nonlinear least squares regression.

2.4.6 Statistics

All error bars are 95% confidence intervals based on a t-distribution with $\alpha = 0.05$ and $n - 1$ degrees of freedom where n is the number of independent biological replicates with a sample standard deviation s.

$$\varepsilon = t^{\alpha}_{n-1} \cdot \frac{s}{\sqrt{n}}$$

A significant difference between samples at $\alpha = 0.05$ is determined when neither interval embraces the other sample's point estimate (i.e. p < 0.05 for a two-sided t-test for independent samples – see Pfister and Janczyk (2013)). For 2D gel quantification, measurements were paired within a batch of experiments by comparing X-spike and Y-arc signal to that of a wildtype strain harboring an empty vector. In this case, a significant change in signal at $\alpha = 0.05$ is determined when the confidence interval does not embrace zero (i.e. p < 0.05 for a two-sided paired-sample t-test – see Pfister and Janczyk (2013)).

2.4.7 Ribosomal DNA recombination assay

Ribosomal DNA (rDNA) recombination assays, initially described by Roy and Runge (2000), were performed by inoculating 3 mL SC -Leu -Ade with a single colony of an rDNA recombination assay strain (*rDNA::ADE2-CAN1*) transformed with a *LEU2*-containing plasmid. The following morning, 30 uL were inoculated into 3

mL SC raffinose -Leu -Ade and grown overnight to saturation. The following morning, approximately 30 uL were inoculated into 3 mL SC Galactose -Leu to a final OD600 of 0.015. Recombination was allowed to take place for 24 hours before concentrating cells to an OD600 of 5 and then plating 100 uL of a 10^{-5} dilution onto fresh SC plates, and 250 uL of a 10^{-4} dilution onto fresh SC -Arg +Canavanine plates. Cells were spread on plates using glass beads. Colonies were counted after 3 to 4 days growth. Individual cultures, representing biological replicates, were plated as three technical replicates with the average colony count from SC plates being used to estimate the number of viable cells present in the culture, and the average colony count on SC -Arg +Canavanine plates being used to estimate the rDNA recombination frequency. Nearly all colonies turned pink indicating simultaneous loss of *ADE2* and *CAN1* in the rDNA array. All incubation steps took place at 30°C.

2.4.8 Galactose induction

Galactose inductions for protein blots and two-dimensional gel analysis was performed by growing cells in SC 3% glycerol 2% lactate media with appropriate nutrient drop-out to maintain selection of a plasmid when present. Cells in log-phase growth were induced by adding galactose to a final concentration of 2%.

2.4.9 Protein blots

Pelleted yeast cultures were resuspended in 5% trichloroacetic acid to precipitate proteins. Protein pellets were then washed in acetone and allowed to dry. Dry pellets were solubilized by bead-beating in TE buffer (50 mM Tris-HCl, 1 mM EDTA, pH 7.5) followed by addition of 3x SDS sample buffer. Samples were boiled for 5 minutes and microcentrifuged at max speed for 5 minutes before SDS-PAGE separation using 4-15% gradient mini-PROTEAN® TGX™ gels (Biorad) and transferred to nitrocellulose membranes with a Trans-Blot® Turbo transfer system (Biorad). Membranes were blotted overnight at 4°C with either goat α-Rad5 polyclonal antibody (SC15548 Santa Cruz Biotechnology), mouse α-FLAG® M2 monoclonal antibody (F3165 Sigma), mouse α-V5 monoclonal antibody (R96025 Invitrogen), or mouse α-Pgk1 monoclonal antibody (22C5D8 Thermo-Fisher). Protein was then visualized using secondary antibodies for α-mouse-IgG or α-goat-IgG conjugated to horseradish peroxidase with light emission catalyzed by SuperSignal® West Pico Chemiluminescent Substrate (Thermo). Light emission was captured using an Amersham™ Imager 600 (GE Healthcare).

2.4.10 Structural homology and sequence alignment

Structural homology modeling was performed using SWISS-MODEL (https://swissmodel.expasy.org) (Guex et al. 2009; Bienert et al. 2017; Waterhouse et al. 2018). HIRAN domain sequences were accessed from Pfam (https://pfam.xfam.org) and sequence alignment was performed using Clustal Omega (Bodenhofer et al. 2015; Sievers and Higgins 2014; Finn et al. 2014).

2.4.11 Analysis of replication intermediates

All strains used for analysis of replication intermediates are *bar1*Δ *RAD5* derivatives of W303 containing an empty vector control plasmid or *RAD5* alleles under the control of the galactose-inducible *GAL1* promoter (**Table 2.8**). Cells were pre-grown overnight in SC medium lacking leucine using 3% glycerol and 2% lactate as carbon sources to relieve glucose repression of the *GAL1* promoter. Cultures in log phase were induced by addition of 2% galactose. At the same time alpha-mating pheromone (Zymo Research, Irvine, CA) was added to 1 μM final concentration. Cultures were incubated with shaking at 30˚C for 3.5 hours and monitored for G1 cell cycle arrest by microscope. Alpha pheromone was removed and cells were released into S phase by pelleting the culture, resuspending in sterile water, re-pelleting and suspending in the original volume of SC -leucine medium containing 2% galactose. *Streptomyces griseus* protease (P8811; Millipore-Sigma) was added to a final concentration of 70 μg/ml to inactivate residual alpha pheromone and hydroxyurea (H8627; Millipore-Sigma) was added to 30mM final concentration. Cultures were incubated with shaking at 30˚C and samples were collected and arrested by addition of sodium azide to 5mM final concentration then stored on ice or at 4˚C until further processing.

Cell cycle progression of samples collected for replication analysis was monitored by staining DNA with SYBR Green I (Thermo Fisher) as previously described (Fortuna et al. 2001) and analyzed using an LSRII analyzer (Becton Dickinson) at the Columbia University Medical Center Department of Microbiology & Immunology Flow Cytometry Core facility. FACS data was analyzed using the flowCore package for R (Ellis et al. 2017). Stabilization and isolation of replicating DNA molecules in the presence of CTAB was performed as described (Neelsen et al. 2014). Approximately 2 μg of each DNA sample was digested with NcoI before electrophoresis.

Neutral-neutral two-dimensional agarose gel electrophoresis was adapted from the procedure of Brewer and Fangman (1987). The first dimension was 0.4% agarose and was run for 16 hr at 1 V/cm in TBE buffer. The second dimension was 0.8% agarose run for 6 hr at 5 V/cm in TBE at 4˚C with 0.4 μg/ml ethidium bromide.

DNA was transferred to Hybond(TM)-XL nylon membranes (GE Healthcare) using an alkaline buffer transfer procedure (Brown 2001). Eight PCR products of 300-500 bp were used as hybridization probes to cover the 5.1 Kb NcoI fragment containing *ARS305*. The amplified probes were pooled in equimolar concentrations and labeled with ^{32}P-CTP via random priming. Primers for probe PCRs are listed in **Table 2.7**. Hybridized blots were exposed to storage phosphor screens (GE Healthcare) and imaged on a Typhoon 7100 phosphorimager (Amersham). Measurements of RI signal intensity was performed by selecting specific RI features using ImageJ 2.0 software (Rueden et al. 2017). A mean background signal was subtracted and each measurement was normalized to the 1N spot. All measurements were performed using images in which the 1N spot signal was below saturation of the phosphor screen. RI feature signal intensity differences between empty vector and *RAD5*^{OE} were calculated for samples collected within the same experimental set.

2.4.12 Software and data analysis

Data analysis was performed using the R project for statistical computing version 3.4.4 (https://www.r-project.org) and RStudio version 1.1.456 (https://www.rstudio.com), making extensive use of packages from Bioconductor version 3.5 (http://bioconductor.org) and a snapshot of CRAN from 2018-03-01 (https://cran.microsoft.com/snapshot/2018-03-01). We would like to specifically acknowledge maintainers of R, RStudio and the tidyverse, knitr, rmarkdown, cgdsr, EBImage, Rtsne, flowCore, and msa packages (R Core Team 2018; RStudio Team 2016; Wickham 2017; Xie 2015; Allaire et al. 2018; Jacobsen and Questions 2018; Krijthe 2015; Pau et al. 2010; Ellis et al. 2017; Bodenhofer et al. 2015).

2.5 Figures and tables

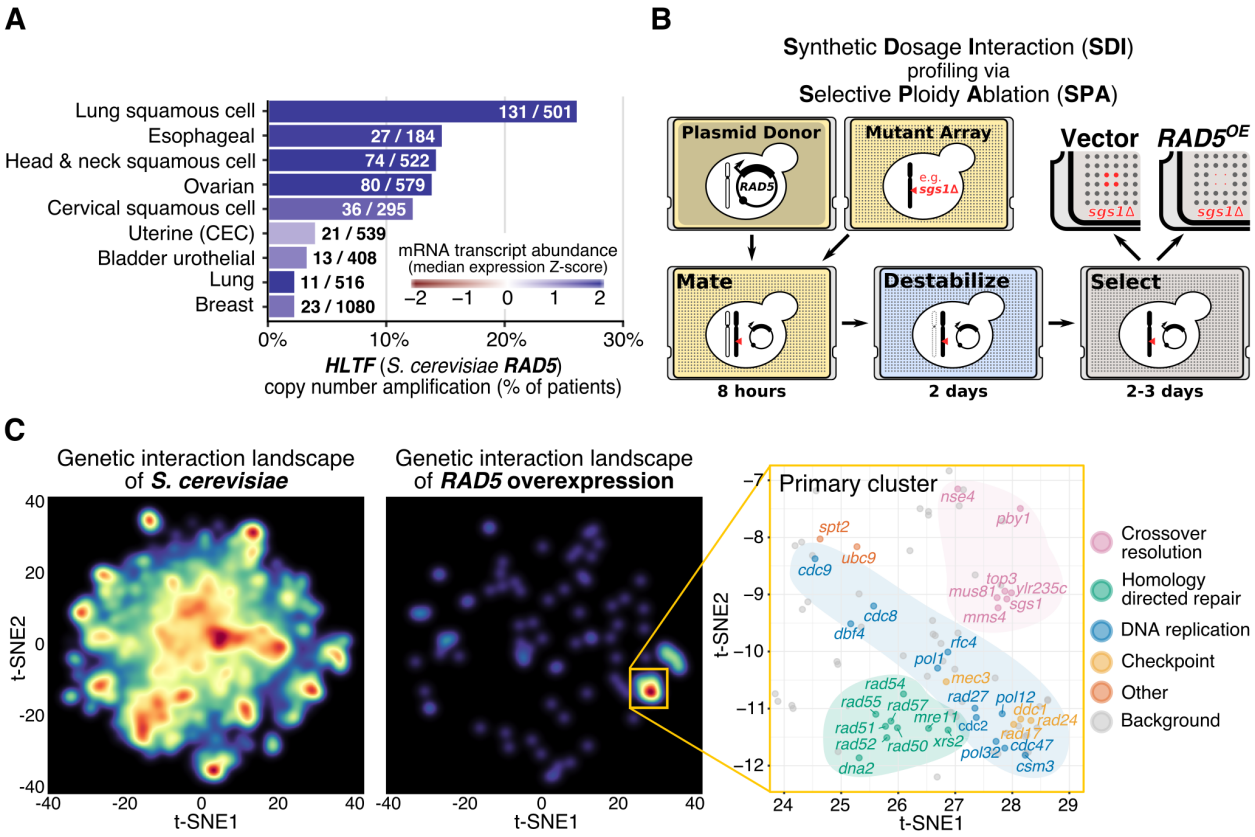

Figure 2.1 – *RAD5*^{OE} causes a requirement for DNA replication and homology directed repair genes. (**A**) The *RAD5* human ortholog, *HLTF*, is overexpressed in squamous cell carcinomas. The percent of TCGA tumor samples with amplified *HLTF* copy number are shown (see Methods). Bar color corresponds to the median of mRNA expression Z-scores within each cohort. Annotated on each bar are the number of samples with *HLTF* amplified and the total number of samples for each cohort. (**B**) Schematic for Synthetic Dosage Interaction (SDI) profiling via Selective Ploidy Ablation (SPA) (Reid et al. 2011) – see **Methods** 2.4.3. (**C**) Genetic interaction landscape enrichment for *RAD5*^{OE} highlights a requirement for DNA replication and homology directed recombination (HDR) genes. On the left, genes represented in the *MAT***a** non-essential gene disruption and temperature-sensitive strain collections were clustered with t-SNE in two dimensions using genetic interaction profile correlation data from The Cell Map (see **Methods** 2.4.4) (Usaj et al. 2017; Costanzo et al. 2016). Dark red spots represent high-density clusters of multiple genes with similar genetic interaction profiles. The locations of *RAD5*^{OE} SDIs in the landscape are shown in the middle. Two significantly enriched clusters were identified using a permutation test ($p < 0.001$). On the right, the region of the landscape representing the primary enriched cluster is shown. *RAD5*^{OE} SDIs are labeled and colored by function. Genes in this region of the landscape that were not identified as having a *RAD5*^{OE} genetic interaction are labeled background and shown as gray dots. Genetic interaction screen was performed by Robert Reid. Related to **Sections** 2.2.1 **&** 2.2.2.

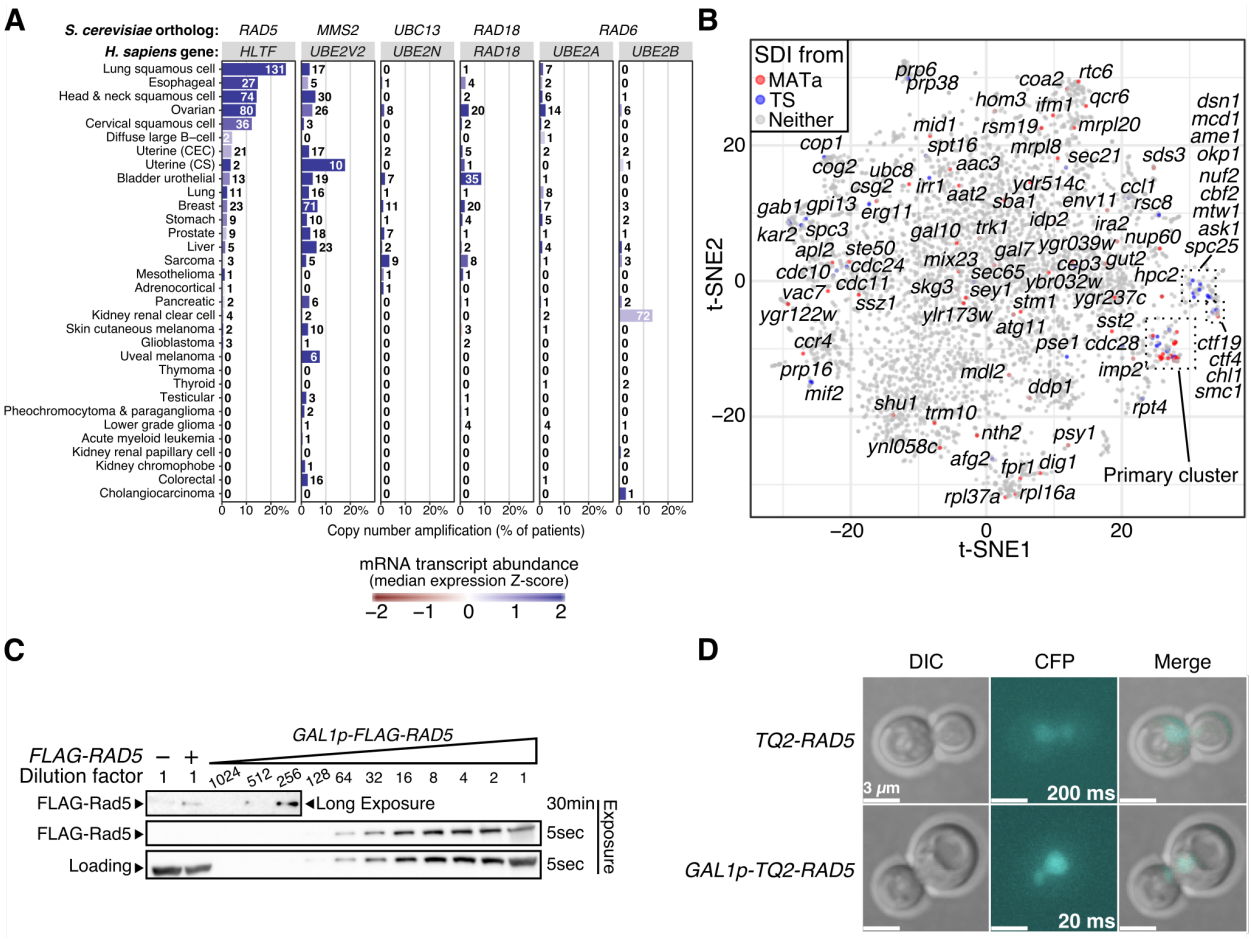

Figure 2.2 – Supplement to Figure 2.1 – Extended TCGA CNA analysis, all SGI hits on the genetic landscape and *RAD5*^{OE} expression characteristics. (**A**) TCGA copy number amplification of PRR genes as a percent of patients in a given cohort (see methods). Numbers by each bar are the number of patients with the gene amplified in the cohort. Color of bars represent the median transcript abundance of amplified samples (blue = increased abundance, red = decreased abundance, white = no change from diploid samples). *S. cerevisiae* ortholog of corresponding *H. sapiens* gene is annotated at the top of the panel. (**B**) The genetic landscape of *S. cerevisiae* from **Figure 2.1C** with all genes shown as gray points, hits from the *MAT***a** KanMX disruption screen shown as blue points and hits from the temperature-sensitive allele screen shown as red points. Gene names of hits are annotated near their location in the landscape. Clusters of hits are outlined with boxes. Genes in the primary cluster are annotated in **Figure 2.1C - right**. (**C**) Endogenously expressed *RAD5* versus galactose-induced *RAD5* expression. For assessing wildtype levels of Rad5 protein expression, a FLAG epitope-tagged version of *RAD5* was incorporated into the *RAD5* locus by marker-free pop-in/pop-out chromosome integration. Only a faint signal of endogenous FLAG-Rad5 was observed after a long exposure (30 minutes – lane 2). For comparison, protein was prepared after four hours of induction of *GAL1p-FLAG-RAD5* and serially diluted to determine the approximate dilution factor necessary to achieve similar signal to endogenous *FLAG-RAD5*. (**D**) Fluorescently-tagged Rad5 (*TQ2-RAD5*) at the endogenous locus was visualized to assess approximate abundance and localization. The same construct under control of the galactose promoter (*GAL1p-TQ2-RAD5*) was induced for four hours and visualized for comparison. Long exposures of the CFP channel (200 milliseconds) were necessary to see endogenous levels of Rad5, whereas short exposures (20 milliseconds) were sufficient to observe galactose-induced Rad5. Related to **Section 2.2.2**.

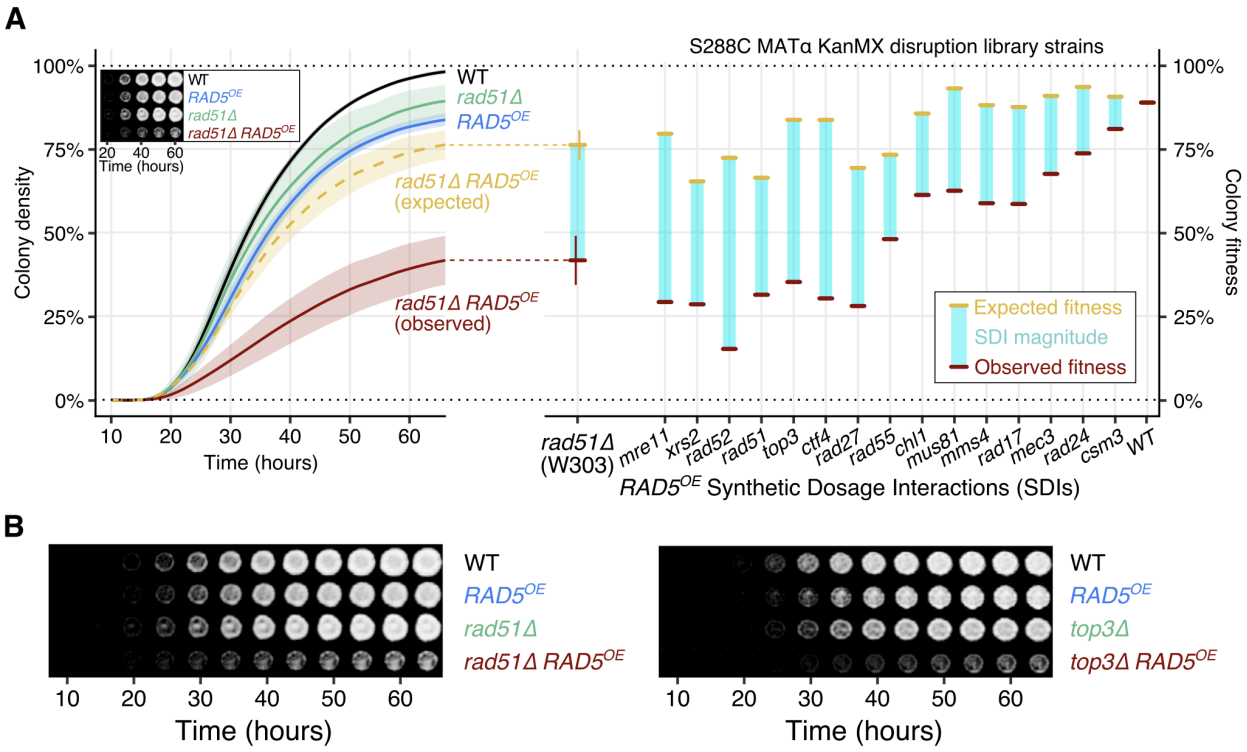

Figure 2.3 – HDR mutants are the most sensitive to *RAD5^{OE}*. (**A**) On the left, growth curves used to quantify the *rad51*Δ *RAD5^{OE}* SDI are shown with representative images of colony growth over time shown as an inset. Shaded regions represent the 95% confidence interval (CI) of at least four independent experiments. Expected growth of *rad51*Δ *RAD5^{OE}* was modeled by multiplying the independent contributions of *rad51*Δ and *RAD5^{OE}* to colony fitness (see **Methods** 2.4.5). On the right, differences in observed from expected colony fitness at 65 hours of growth are summarized using strains from the *MATα* gene disruption collection for genes selected from the clusters identified by landscape enrichment analysis. (**B**) Representative time series of colony growth demonstrating the phenotype measured for the *RAD5^{OE}* genetic interactions with *rad51*Δ and *top3*Δ strains from the *MAT***a** KanMX disruption collection. Colony images were programmatically stitched together and normalized such that no-signal is black and highest colony density is white. Related to **Section** 2.2.2.

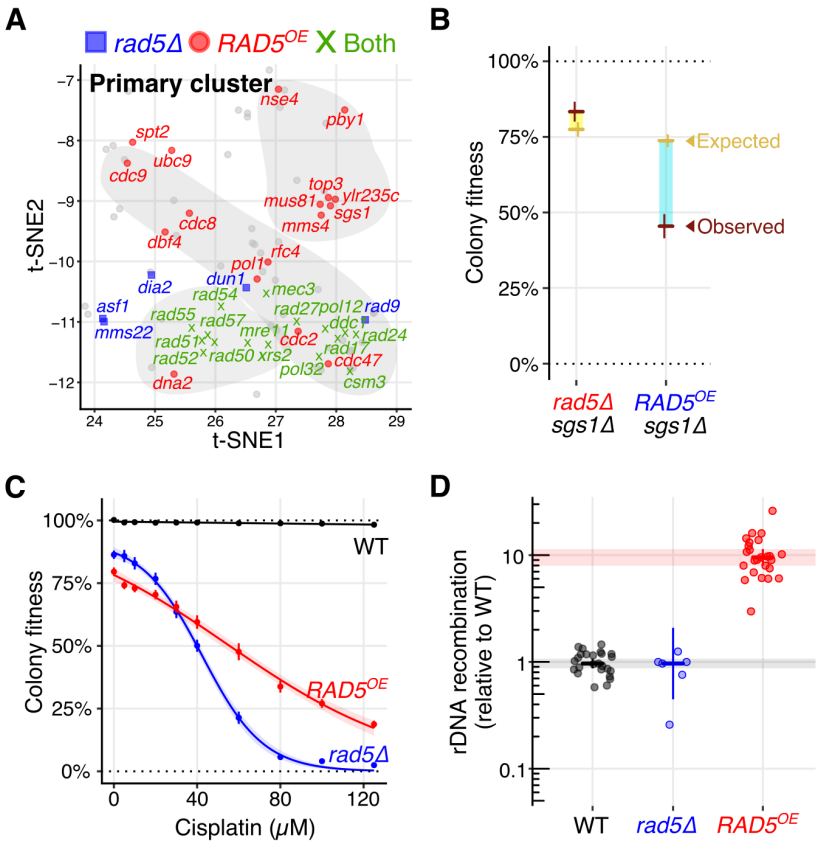

Figure 2.4 – *RAD5^{OE}* causes sensitivity to replication inhibitors and increased recombination. (**A**) Panel shows the primary cluster from Figure 2.1C highlighting genetic interactions specific to *RAD5^{OE}* (red circles), *rad5*Δ (blue squares), or shared between *RAD5^{OE}* and *rad5*Δ (green Xs). Interaction data for *rad5*Δ was acquired from BioGRID (Stark et al. 2006). (**B**) Quantified colony fitness confirms the absence of an *sgs1*Δ *rad5*Δ genetic interaction, and presence of an *sgs1*Δ *RAD5^{OE}* interaction. Horizontal and vertical bars represent the mean and 95% CI of at least four independent experiments for observed (red) and expected (yellow) fitness. Turquoise and yellow rectangles represent negative and positive interaction magnitudes respectively. (**C**) Both *RAD5^{OE}* and *rad5*Δ cause cisplatin sensitivity. Cisplatin dosage response curves for wildtype (WT), *RAD5^{OE}*, and *rad5*Δ are shown. Colony fitness was scored relative to untreated WT. Points and vertical bars represent the mean and 95% CI of four independent experiments. Shaded regions represent the 95% confidence band of the modeled dose response curves. (**D**) *RAD5^{OE}*, but not *rad5*Δ, promotes tandem repeat recombination in ribosomal DNA (rDNA). rDNA recombination frequency was measured by loss of_ADE2_ and *CAN1* markers incorporated within the rDNA tandem repeat array. Frequencies are normalized to the median recombination frequency observed for a WT assay strain with an empty vector. Horizontal and vertical bars indicate the median and 95% CI for at least 6 independent experiments shown as points. rDNA recombination experiments were performed by Ivana Šunjevarić. Related to **Sections 2.2.3, 2.2.4 & 2.2.5**.

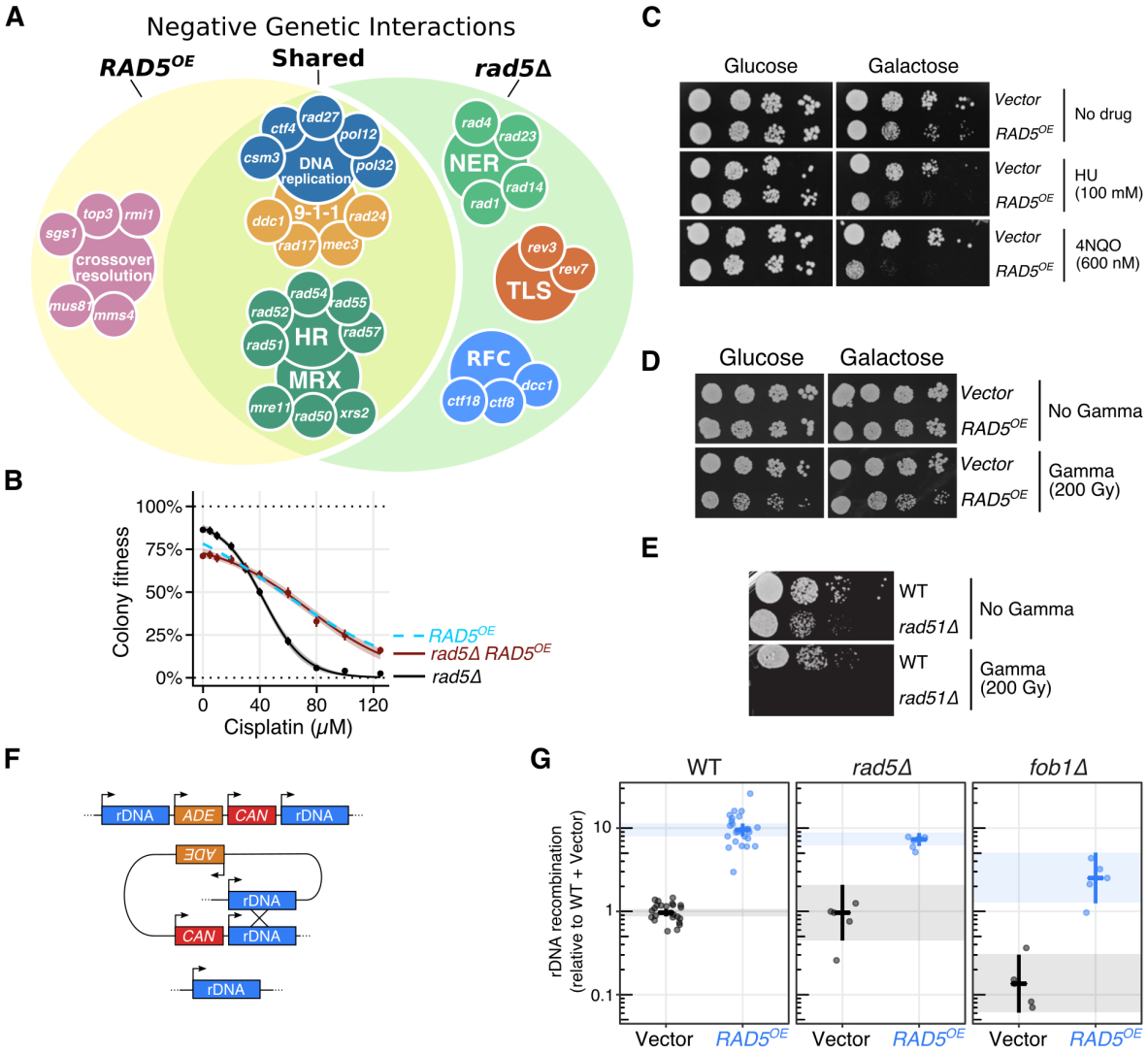

Figure 2.5 – Supplement to Figure 2.4 – Overlapping interaction profiles of *RAD5*^{OE} & *rad5*Δ, suppression of *rad5*Δ cisplatin sensitivity by *RAD5*^{OE}, other damaging agent sensitivities, and *fob1*Δ independent rDNA recombination. (**A**) Comparison of *RAD5*^{OE} and *rad5*Δ genetic interactions. *rad5*Δ genetic interactions were derived from BioGRID data. A complete list of such genes is included in **Table 2.6**. (**B**) Cisplatin sensitivity of *RAD5*^{OE} in a *rad5*Δ background. Colony fitness was scored relative to untreated wildtype. Blue dashed line indicates cisplatin sensitivity of *RAD5*^{OE} in a wildtype background for reference. Points and vertical bars represent the mean and 95% CI of four independent experiments. Shaded regions represent the 95% confidence band of the modeled dose response curves. (**C,D**) *RAD5*^{OE} response to 100 mM HU, 600 nM 4NQO, and 200 Gy gamma radiation. Spots represent 10-fold serial dilutions of the designated plasmid and drug combination. Galactose media is used to induce *RAD5*^{OE}. Gamma treated cells were irradiated with a Cobalt-60 source. (**E**) 10-fold serial dilution drop assay showing sensitivity of *rad51*Δ to 200 Gy gamma radiation. (**F**) rDNA assay schematic. Loss of both *CAN1* and *ADE2* via recombination between rDNA repeats results in canavanine resistance and red colonies. (**G**) *RAD5*^{OE} driven rDNA recombination in wildtype, *rad5*Δ and *fob1*Δ backgrounds. Frequencies are normalized to the median recombination frequency observed for a wildtype assay strain with an empty vector. Horizontal and vertical bars indicate the median and 95% CI for at least 5 independent experiments shown as points. rDNA recombination experiments were performed by Ivana Šunjevarić. Serial dilution drop assays were performed by Ivana Šunjevarić (**C** & **D**), and Michael Smith (**E**). Related to **Sections 2.2.3, 2.2.4 & 2.2.5**.

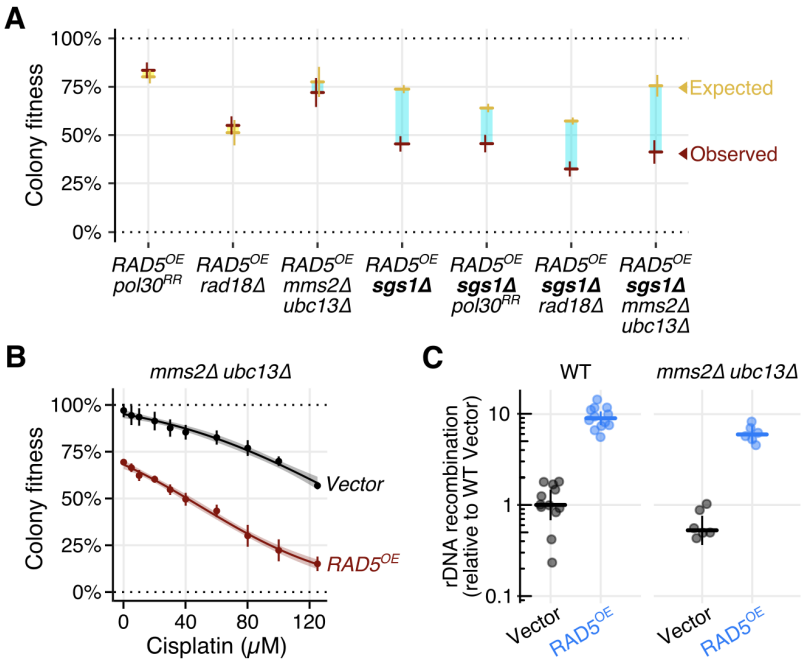

Figure 2.6 – Other PRR members are not essential for *RAD5^{OE}* driven recombination. (**A**) Quantified colony fitness reveals *RAD5^{OE}* requires *SGS1* even in the absence of PRR members *RAD18*, *MMS2*, *UBC13*, and in the absence of canonical PRR PCNA ubiquitination sites (i.e. *pol30^{RR}*). Horizontal and vertical bars represent the mean and 95% CI of at least four independent experiments for observed (red) and expected (yellow) fitness. Turquoise and yellow rectangles represent negative and positive interaction magnitudes respectively. (**B**) Quantified colony growth reveals cisplatin sensitivity of *RAD5^{OE}* does not depend on the Rad5 E2 partner – Mms2-UBC13. Colony fitness was scored relative to untreated wildtype. Points and vertical bars represent the mean and 95% CI of four independent experiments. Shaded regions represent the 95% confidence band of the modeled dose response curves. (**C**) *RAD5^{OE}* drives rDNA recombination even in the absence of Mms2-Ubc13. Frequencies are normalized to the median recombination frequency observed for a WT assay strain with an empty vector. Horizontal and vertical bars indicate the median and 95% CI for at least 6 independent experiments shown as points. rDNA recombination experiments were performed by Ivana Šunjevarić and myself. Related to **Section 2.2.6**.

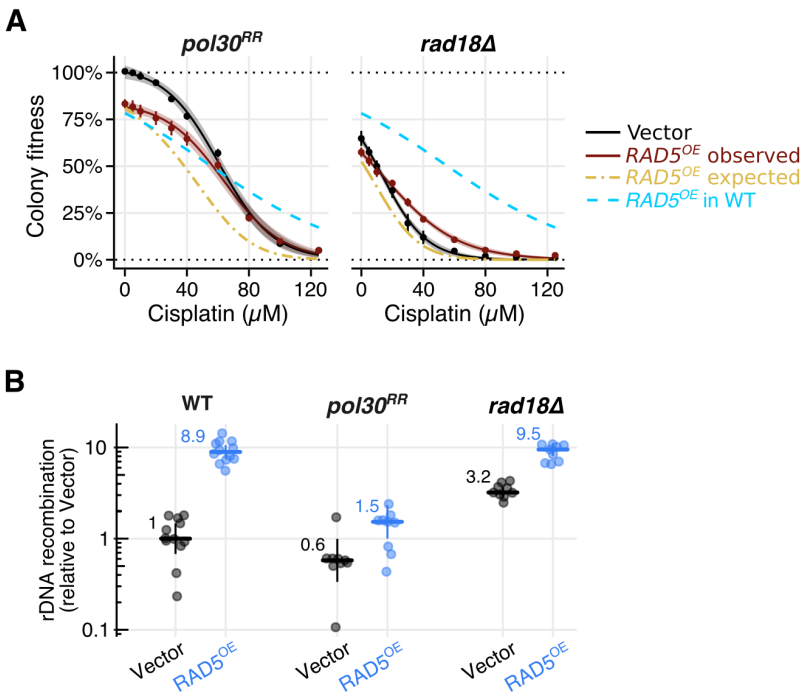

Figure 2.7 – Supplement to Figure 2.6 – *pol30*^{RR} and *rad18*Δ contributions to *RAD5*^{OE} cisplatin sensitivity and rDNA recombination. (**A**) Cisplatin sensitivity of *RAD5*^{OE} in *pol30*^{RR} and *rad18*Δ backgrounds. Colony fitness was scored relative to untreated wildtype. Blue dashed lines indicate cisplatin sensitivity of *RAD5*^{OE} in a wildtype background for reference. Yellow dashed line indicates expected fitness if the *RAD5*^{OE} contribution to fitness were independent from either *pol30*^{RR}, or *rad18*Δ. Points and vertical bars represent the mean and 95% CI of four independent experiments. Shaded regions represent the 95% confidence band of the modeled dose response curves. (**B**) *RAD5*^{OE} driven rDNA recombination in wildtype, *pol30*^{RR} and *rad18*Δ backgrounds. Frequencies are normalized to the median recombination frequency observed for a wildtype assay strain with an empty vector. Horizontal and vertical bars indicate the median and 95% CI for at least 9 independent experiments shown as points. Value of horizontal bars is annotated to the left of each condition. rDNA recombination experiments were performed by Ivana Šunjevarić and myself. Related to **Section 2.2.6**.

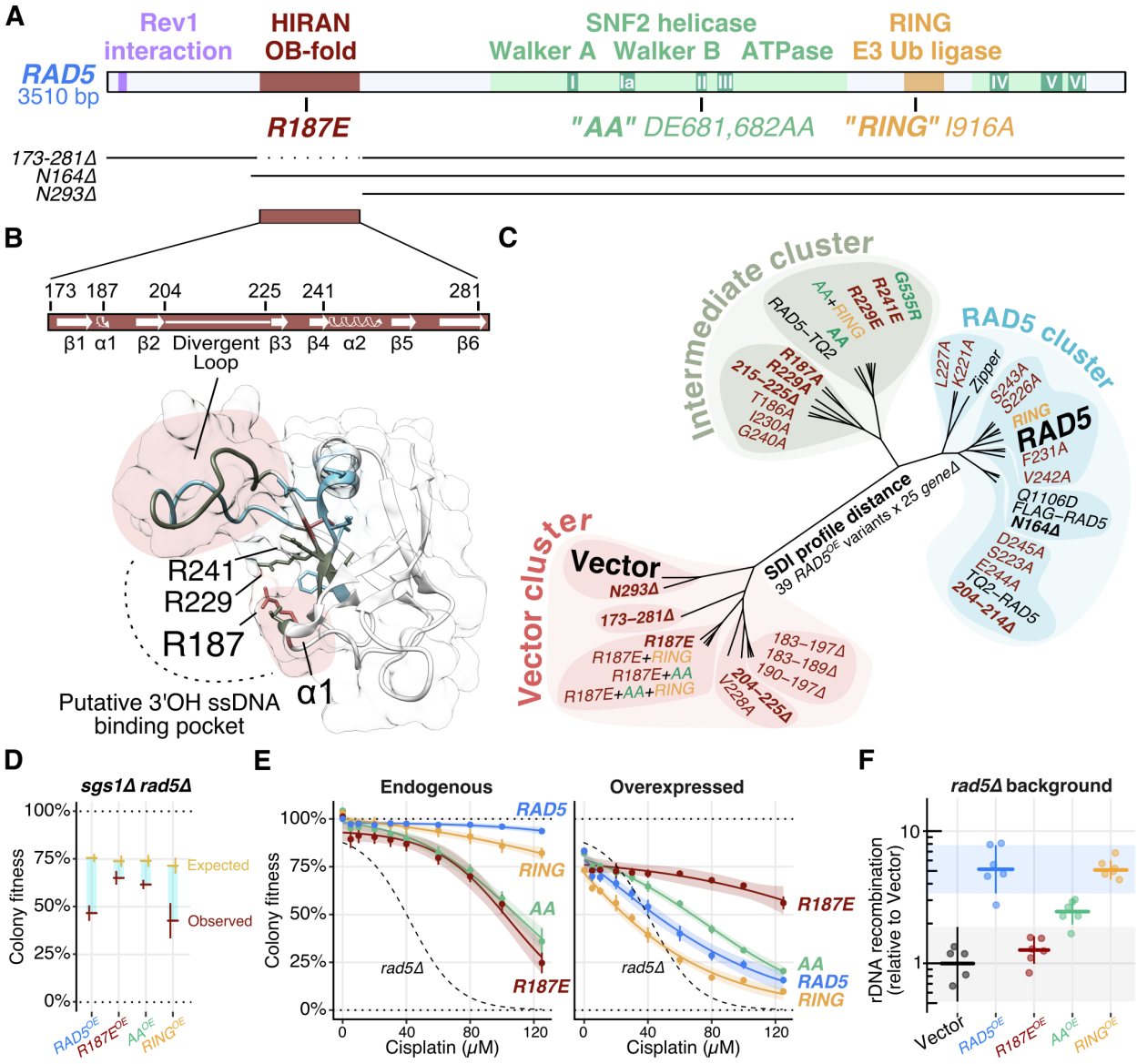

Figure 2.8 – The HIRAN domain is essential for the *RAD5*^{OE} phenotype. (**A**) *RAD5* domain map with N-terminal truncations (*N164*Δ & *N293*Δ) and HIRAN domain deletion (*173-281*Δ) variants indicated below. (**B**) Predicted secondary and tertiary structures from the *RAD5* HIRAN domain *HLTF* homology model. The corresponding putative 3'OH ssDNA binding pocket, described in *HLTF* by Kile et al. (2015), is indicated with a dotted line. Three highly conserved positively charged residues are annotated. Colors indicate vector cluster (red), intermediate cluster (green), and *RAD5* cluster (blue) for variants described in panel **C**. Red shaded regions highlight loops critical for *RAD5*^{OE} genetic interactions. (**C**) Unrooted dendrogram visualizing the SDI profile distance between 39 overexpressed *RAD5* variants scored for interaction against 25 *RAD5*^{OE} SDIs using selective ploidy ablation (Reid et al. 2011). Variants having no effect on the interaction profile fall in the *RAD5* cluster (blue). Variants having a partial effect fall in the intermediate cluster (green). Variants that nearly eliminate *RAD5*^{OE} genetic interactions fall in the vector cluster (red). Bold *RAD5* variants are discussed in the text. (**D,E,F**) Experiments, as described for previous figures, comparing *sgs1*Δ interaction, cisplatin sensitivity, and rDNA recombination for *RAD5*, *rad5*^{R187E} (*R187E*), *rad5*^{DE681,682AA} (*AA*) and *rad5*^{I916A} (*RING*) variants (see also Methods). (**E**) Endogenous refers to variants expressed from the native *RAD5* promoter in the chromosome. Overexpressed refers to variants expressed from the plasmid-born *GAL1* promoter in the corresponding *RAD5* variant background. rDNA recombination experiments were performed by Ivana Šunjevarić. Related to **Sections 2.2.7 & 2.2.8**.

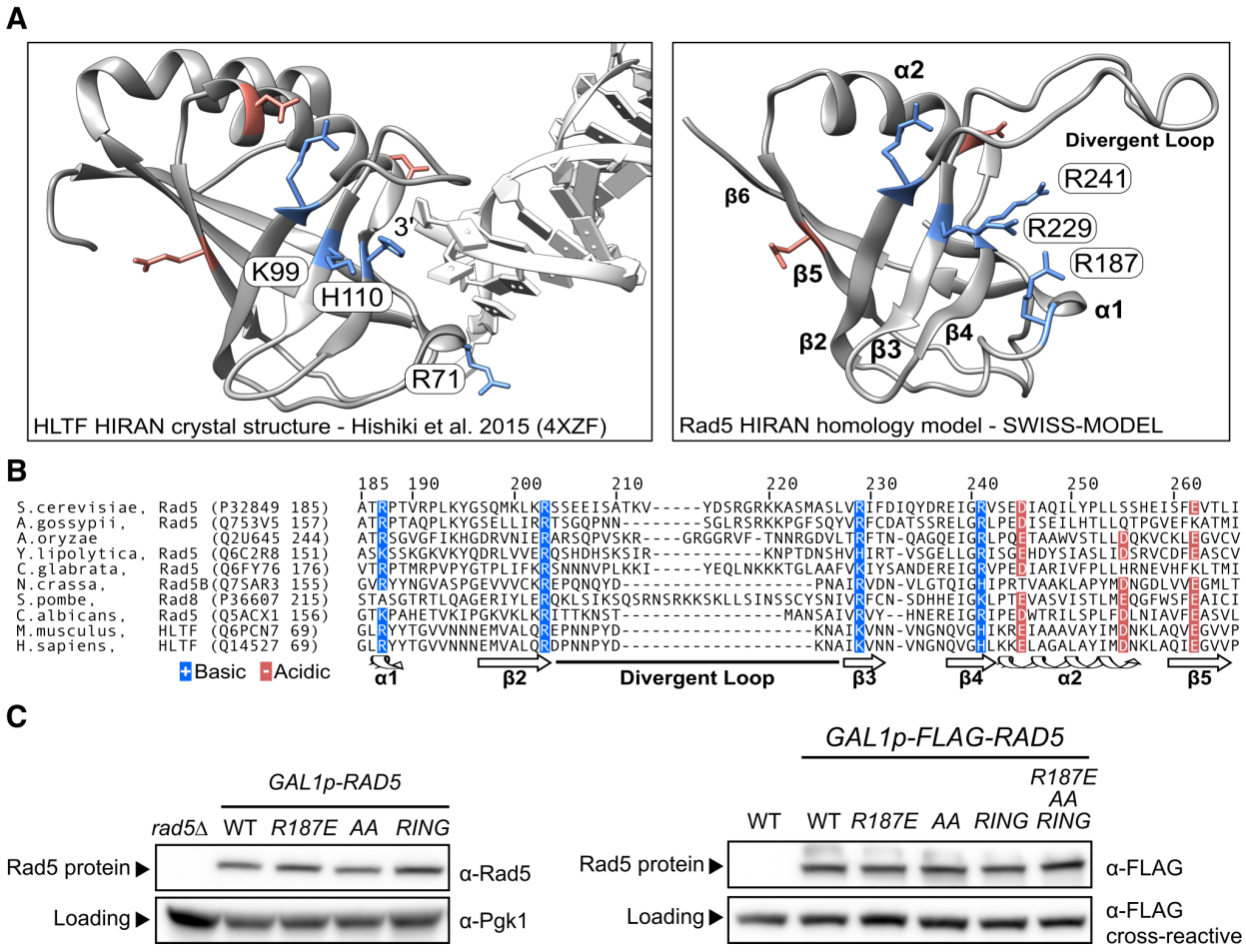

Figure 2.9 – **Supplement to Figure 2.8 – Rad5 HIRAN domain conservation, and variant protein stability.** (**A**) On the left is a crystal structure of the HLTF HIRAN domain in complex with the 3' end of one strand from a short DNA duplex from Hishiki et al. (2015). On the right is the Rad5 HIRAN domain homology model. Conserved basic and acidic residues are colored blue and red respectively. Conserved basic residues facing the ssDNA binding pocket are annotated. (**B**) Alignments of HIRAN domains across diverse yeast species. Protein sequences were downloaded from Pfam and aligned with Clustal Omega. Numbers at the top indicate residue number for *S. cerevisiae* Rad5. Conserved basic and acidic residues are colored blue and red respectively. Approximate locations of secondary structure from the homology model are annotated below the alignment. (**C**) Protein blots of overexpressed Rad5 and FLAG epitope tagged Rad5. Variants in the overexpressed Rad5 protein are indicated above their corresponding lane. Related to **Section 2.2.8**.

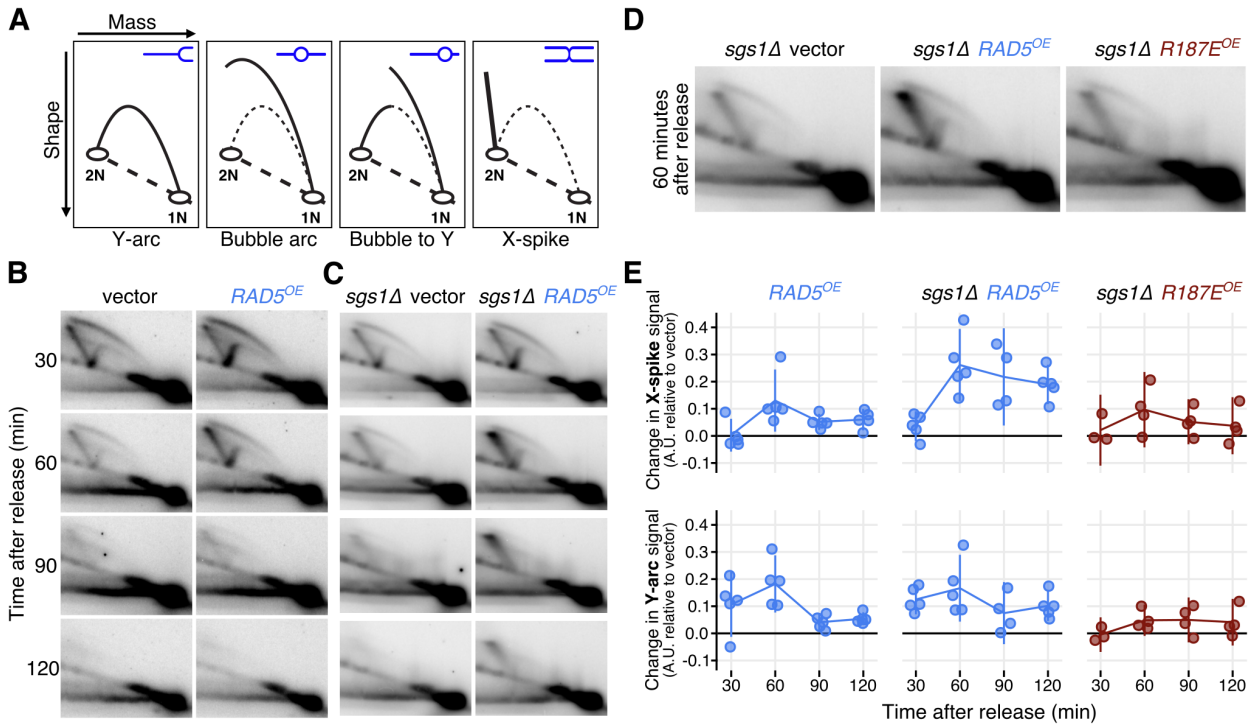

Figure 2.10 – ***RAD5^{OE}* promotes aberrant template switching during DNA replication.** (**A**) Representation of replication intermediate shapes detectable by 2D gel analysis. Replicating DNA samples are separated by electrophoresis in one dimension based on mass (top arrow on leftmost diagram) and in a second dimension by shape (left arrow). Each panel depicts a mode of replication through a restriction fragment. A representative DNA molecule shape (blue) in the inset and their migration are represented by the black arc in the diagram. Non-replicating linear molecules migrate to a single spot (1N) and the dashed diagonal line represents the line of linear DNA molecules. Thin dashed arcs representing Y-arcs are included in other panels for reference. (**B**) The 2D gel electrophoresis analysis of wildtype cells. Replicating DNA was isolated after arresting cells in G1 and releasing them into medium containing 30mM HU with either a control plasmid (vector) or a *RAD5* expression plasmid (*RAD5^{OE}*). Expression was induced during arrest by addition of galactose. Samples were collected at the indicated times after release and replicating DNA was visualized by probing a 5.1 Kbp restriction fragment from chromosome III containing an origin of replication at its midpoint. (**C**) Replicating DNA from an *sgs1*Δ mutant strain was collected and analyzed as in **B**. (**D**) 2D gel analysis of an *sgs1*Δ strain with vector control, *RAD5^{OE}* or the mutant allele *R187E^{OE}*. Samples shown are from the same experiment 60 minutes after release. Vector and *RAD5^{OE}* samples are the same as those shown in panel C. (**E**) Hybridization signal intensities for the X-spike and the Y-arc were measured in multiple experiments and normalized to the 1N spot signal. Normalized measurements from vector control samples were subtracted from the measurements in *RAD5^{OE}* or *R187E^{OE}* cells to obtain the change in signal from vector (points) for each experiment. Vertical bars indicate the 95% CI. 2D-gel experiments were performed by Robert Reid with assistance from Ivana Šunjevarić. Related to **Section 2.2.9**.

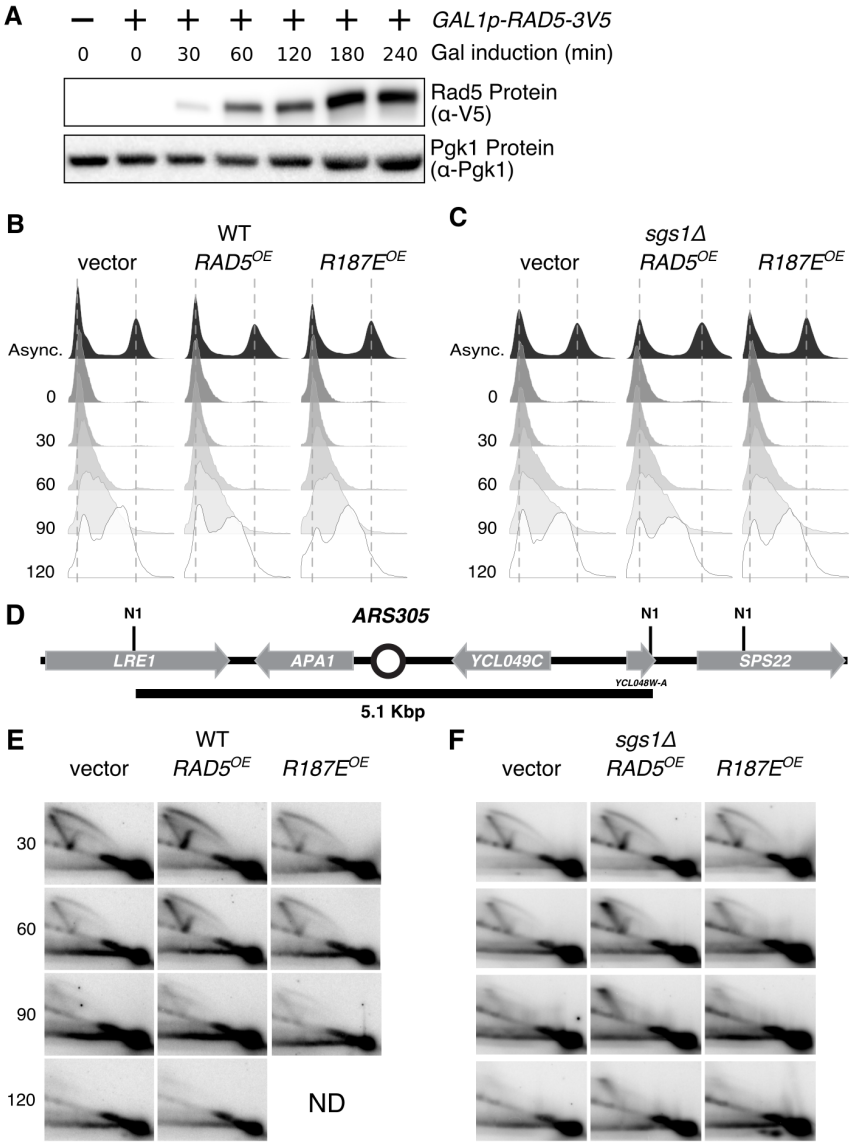

Figure 2.11 – Supplement to Figure 2.10 – Replication fork analysis of *ARS305* by two-dimensional gel electrophoresis. (**A**) Induction of *RAD5* expression from the *GAL1* promoter. A plasmid with N-terminally tagged Rad5 (3xV5 epitope) (**+**) was transformed into a wildtype strain, grown to log phase in SC-Leu medium containing 2% lactate and 3% glycerol as a carbon source. Then *RAD5* expression was induced by addition of 2% galactose. A wildtype strain with an empty vector control (**-**) was also included. Protein samples were collected at the indicated time points after addition of galactose and processed as described in (Methods). Rad5 was detected with the anti-V5 (α-V5) antibody and anti-PGK1 (α-Pgk1) was used as a loading control. Expression of Rad5 peaks by 3 hours of galactose induction. (**B**) FACS analysis of WT strains containing expression vectors for *RAD5* (*RAD5*^{OE}) or the rad5-R187E HIRAN domain mutant *R187E*^{OE}. Asynchronous samples (Async.) were collected before arrest and samples taken after release into medium containing hydroxyurea are indicated by time (minutes). (**C**) FACS analysis of sgs1 samples as in **B**. (**D**) Map of *Saccharomyces* chromosome III proximal to the replication origin *ARS305* (circle). Arrows indicate open reading frames. Vertical lines labeled "N1" indicate NcoI restriction sites, and the 5.1 Kbp NcoI fragment indicated by the thick horizontal line was detected on blots using radiolabeled probes. (**E**) 2D electrophoresis showing *ARS305* replication intermediates at specific timepoints (minutes) after release into hydroxyurea. ND indicates not done. (**F.**) 2D gel analysis of the *sgs1*Δ strain as in **D**. 2D-gel and FACS experiments were performed by Robert Reid and Ivana Šunjevarić. Related to **Section 2.2.9**.

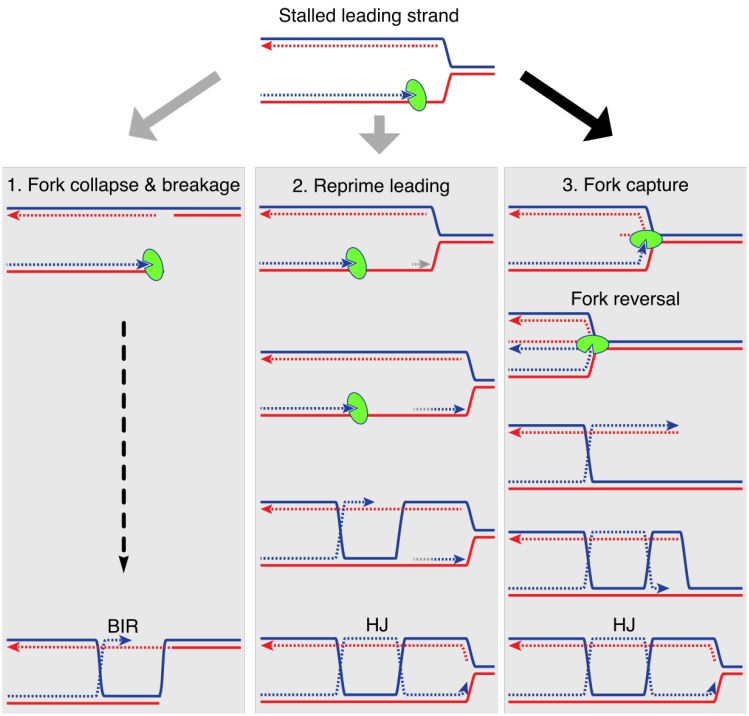

Figure 2.12 – **Models for excess Rad5 at a replication fork.** A replication fork is shown with blue and red solid lines representing the parental strands. Nascent DNA is represented as dashed lines and arrowheads indicate the direction of DNA polymerization. Rad5 (green oval) is shown binding to the 3′ end of the leading strand via the HIRAN domain (cleft in the oval). Figure graphic created by Robert Reid. See text of Discussion section for description of models (**Section 2.3.1**).

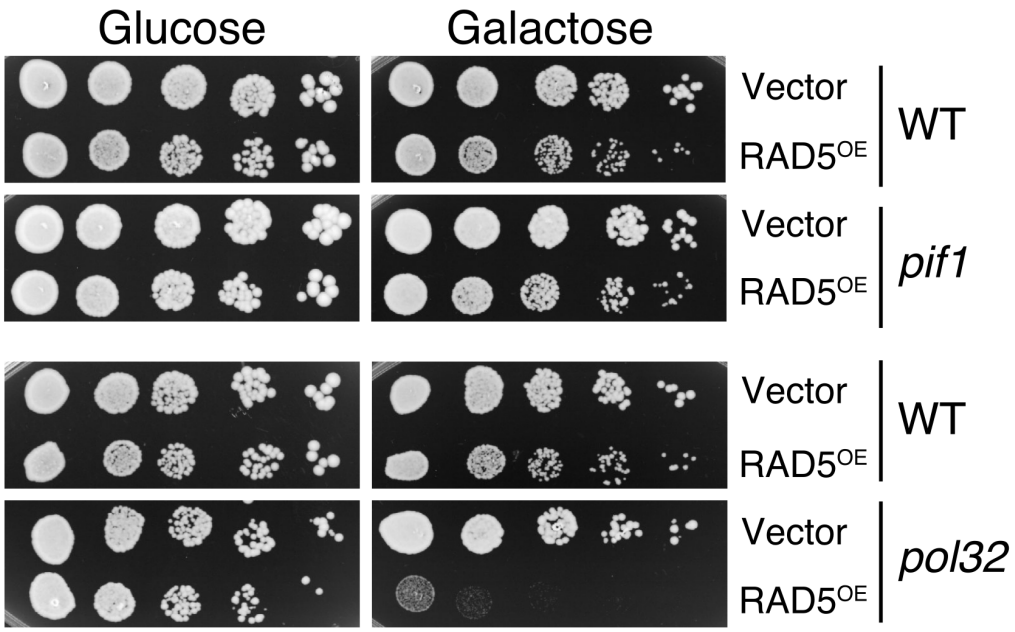

Figure 2.13 – **Supplement to Figure 2.12 – *RAD5^{OE}* cells require Pol32 but do not require Pif1.** Spots represent 5-fold serial dilutions of the designated plasmid and genetic background combinations. All strains are derived from W303. *pol32* strain contains a null allele. *pif1* strain contains a mutation in the second ATG codon in the open reading frame so the polypeptide is imported into mitochondria, but does not function in the nucleus (Zhou et al. 2000). Galactose media is used to induce *RAD5^{OE}*. Serial dilution drop assays were performed by Ivana Šunjevarić. Related to **Section 2.3.1**.

Table 2.1 – PRR gene copy number alterations in cancer

Header	Description
`study_id`	cBioportal study identifier.
`study_name`	cBioportal study name.
`gene`	HGNC gene name for PRR genes *HLTF*, *RAD18*, *UBE2A*, *UBE2B*, *UBE2N* and *UBE2V2*.
`med_amp_exp`	Median expression Z-score for amplified samples.
`n`	Number of samples in cohort *.
`n_amp`	* with two or more copy gain of gene (amplified).
`n_gain`	* with single copy gain of gene.
`n_dipl`	* with diploid copy number of gene.
`n_loss`	* with single copy deletion of gene.
`n_del`	* with two copy deletion of gene.
`freq_amp`	`n_amp` divided by `n`.
`freq_gain`	`n_gain` divided by `n`.
`freq_dipl`	`n_dipl` divided by `n`.
`freq_loss`	`n_loss` divided by `n`.
`freq_del`	`n_del` divided by `n`.

Related to **Figure 2.1**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-01.csv.

Table 2.2 – *RAD5*^{OE} SDL interaction scores – gene disruptions

Header	Description
`strain_id`	Yeast gene deletion consortium record numbers.
`strain_name`	Name of array strain (i.e. name of deleted gene).
`size`	Size of strain with *RAD5*^{OE} normalized to plate median. Colony density measurements are from the synthetic dosage interaction screen of *RAD5*^{OE} (pWJ220) paired with the *MAT***a** KanMX gene deletion strain collection.
`control_size`	Size with vector normalized to plate median.
`log2_ratio`	Log base 2 of `size` divided by `control_size`.
`z`	`log2_ratio` centered at 0 and scaled by standard deviation.

Related to **Figure 2.1**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-02.csv.

Table 2.3 – *RAD5*^{OE} SDL interaction scores – TS alleles

Header	Description
`strain_id`	Yeast gene deletion consortium record numbers.
`strain_name`	Name of array strain (i.e. name of TS allele).
`temperature`	Temperature of incubation (˚C).
`size`	Size of strain with *RAD5*^{OE} normalized to plate median. Colony density measurements for the synthetic dosage interaction screen of *RAD5*^{OE} (pWJ220) paired with the *MAT***a** temperature-sensitive allele strain collection.
`control_size`	Size with vector normalized to plate median.
`log2_ratio`	Log base 2 of `size` divided by `control_size`.
`z`	`log2_ratio` centered at 0 and scaled by standard deviation.

Related to **Figure 2.1**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-03.csv.

Table 2.4 – *RAD5*^{OE} SDL interaction scores – aggregated TS alleles

Header	Description
`gene`	Gene name.
`temperature`	Temperature of incubation (˚C).
`n`	Number of alleles scored for `gene`.
`score_combined`	Scores combined using Stouffer's method ($\sum Z/\sqrt{n}$) (Stouffer et al. 1949).
`max_combined`	Maximum combined score across temperatures
`min_combined`	Minimum combined score across temperatures

Related to **Figure 2.1**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-04.csv.

Table 2.5 – Landscape coordinates

Header	Description
`orf`	ORF ID for gene
`gene`	gene name
`alleles`	alleles of `gene`
`MATa_hit`	did gene have `z` ≤ -2 from Table 2.3?
`TS_hit`	top 50 TS hit with respect to `min_combined` from Table 2.5?
`tSNE1`	t-SNE dimension 1
`tSNE2`	t-SNE dimension 2

Related to **Figure 2.1C**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-05.csv.

Table 2.6 – Aggregated interaction scores for *rad5*Δ from BioGRID

Header	Description
`gene`	Gene name.
`sum_neg`	Sum of BioGRID interactions involving *RAD5* and `gene` assigned to "Negative Genetic", "Synthetic Growth Defect", or "Synthetic Lethality".
`sum_pos`	Sum of BioGRID interactions involving *RAD5* and `gene` assigned to "Synthetic Rescue", or "Positive Genetic".
`score`	Positive minus negative interactions (i.e. `sum_pos` - `sum_neg`). Aggregate scores used to identify synthetic genetic interactions with *rad5*Δ. Genes with aggregate scores of ≤ -2 were considered suitable evidence for *rad5*Δ synthetic genetic interaction.

Related to **Figure 2.2A**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-06.csv.

Table 2.7 – Oligonucleotides

Header	Description
`id`	Rothstein lab primer ID.
`name`	The name of the primer.
`sequence`	The sequence of the primer.
`description`	A brief description of the primer.

Related to **Chapter 2**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/RAD5-07.csv.

Table 2.8 – Strains and plasmids

Name	Genotype	Notes
W9100-17D	*MAT***a** *ADE2 can1-100 his3-11,15 leu2-3,112 lys2Δ MET15 RAD5 TRP1 ura3-1*	Herrero and Thorpe (2016)
W9100-2D	*MATα*	See W9100-17D.
W11353-1A	*MATα rad5::KanMX4*	This study.
W11345-10D	*MATα rad5^{R187E}*	This study.
W11331-7A	*MATα rad5^{DE681,682AA}*	This study.
W11332-7D	*MATα rad5^{I916A}*	This study.
W11362-5B	*MATα rad51::KanMX4*	This study.
W11366-3A	*MATα sgs1::KanMX4*	This study.
W11391-8C	*MATα pol30^{KK127,164RR}-NatMX4*	This study.
W11409-1D	*MATα rad18::HphMX4*	This study.
W11410-4B	*MATα mms2::NatMX4 ubc13::KanMX4*	This study.
W11365-5A	*MATα sgs1::KanMX4 rad5::KanMX4*	This study.
W11391-2C	*MATα sgs1::KanMX4 pol30^{KK127,164RR}-NatMX4*	This study.
W11409-4D	*MATα sgs1::KanMX4 rad18::HphMX4*	This study.
W11410-13D	*MATα sgs1::KanMX4 mms2::NatMX4 ubc13::KanMX4*	This study.
W11248-5B	*MAT***a** *LYS2 bar1::NatMX4*	This study.
W11260-7A	*MAT***a** *bar1::NatMX4 sgs1::KanMX4*	This study.
W4314-2C	*MATα ade2-1 LYS2 trp1-1 rDNA::ADE2-CAN1*	This study.
W5270-1A	*MATα ade2-1 LYS2 trp1-1 rDNA::ADE1-CAN1 fob1::KanMX4*	This study.
W11349-7D	*MAT***a** *ade2-1 LYS2 trp1-1 rDNA::ADE2-CAN1 rad5::KanMX4*	This study.
W11262-22C	*MATα ade2-1 LYS2 trp1-1 rDNA::ADE1-CAN1 pol30^{KK127,164RR}-NatMX4*	This study.
W11421-44B	*MATα ade2-1 LYS2 trp1-1 rDNA::ADE1-CAN1 rad18::HphMX4*	This study.
W11422-30C	*MAT***a** *ade2-1 LYS2 trp1-1 rDNA::ADE1-CAN1 mms2::NatMX4 ubc13::KanMX4*	This study.
pWJ1781	*CEN LEU2 Amp^{r} GAL1p-HpaI-ADH2t*	Reid et al. (2016)
pWJ2496	*GAL1p-HpaI-3V5-ZraI-ADH2t*	This study. †
pWJ2509	*GAL1p-PmeI- mTurquoise2-ZraI-ADH2t*	This study. †
pWJ2200	*GAL1p-RAD5*	This study. †
pWJ2244	*GAL1p-rad5^{T186A}*	This study. †
pWJ2555	*GAL1p-rad5^{R187E}*	This study. †
pWJ2239	*GAL1p-rad5^{V228A}*	This study. †
pWJ2240	*GAL1p-rad5^{R229A}*	This study. †
pWJ2557	*GAL1p-rad5^{R229E}*	This study. †
pWJ2241	*GAL1p-rad5^{I230A}*	This study. †
pWJ2243	*GAL1p-rad5^{G240A}*	This study. †
pWJ2559	*GAL1p-rad5^{R241E}*	This study. †
pWJ2532	*GAL1p-rad5$^{DEELES328-333AAAAAA}$* “Zipper”	This study. †
pWJ2124	*GAL1p-rad5^{G535R}*	This study. †
pWJ2230	*GAL1p-rad5^{DE681,682AA}*	This study. †
pWJ2231	*GAL1p-rad5^{I916A}*	This study. †
pWJ2238	*GAL1p-rad5^{Q1106D}*	This study. †
pWJ2233	*GAL1p-rad5$^{p.1-164Δ}$*	This study. †
pWJ2234	*GAL1p-rad5$^{p.1-293Δ}$*	This study. †
pWJ2447	*GAL1p-rad5$^{p.173-281Δ}$*	This study. †

Name	Genotype	Notes
pWJ2448	*GAL1p-rad5*$^{p.183\text{-}189\Delta}$	This study. †
pWJ2449	*GAL1p-rad5*$^{p.183\text{-}197\Delta}$	This study. †
pWJ2450	*GAL1p-rad5*$^{p.190\text{-}197\Delta}$	This study. †
pWJ2451	*GAL1p-rad5*$^{p.204\text{-}214\Delta}$	This study. †
pWJ2452	*GAL1p-rad5*$^{p.204\text{-}225\Delta}$	This study. †
pWJ2453	*GAL1p-rad5*$^{p.215\text{-}225\Delta}$	This study. †
pWJ2221	*GAL1p-FLAG-RAD5*	This study. †
pWJ2586	*GAL1p-FLAG-rad5*^{R187E}	This study. †
pWJ2584	*GAL1p-FLAG-rad5*^{DE681,682AA}	This study. †
pWJ2473	*GAL1p-FLAG-rad5*^{I916A}	This study. †
pWJ2583	*GAL1p-FLAG-rad5*^{RDEI187,681,682,916EAAA}	This study. †
pWJ2512	*GAL1p-RAD5-3V5*	This study. ‡
pWJ2511	*GAL1p-TQ2-RAD5*	This study. ‖
pWJ2513	*GAL1p-RAD5-TQ2*	This study. ‖

Related to **Chapter 2** – Full genotype is shown for W9100-17D – a derivative of W303 with *rad5*^{G535R} corrected to *RAD5* (Thomas and Rothstein 1989; Herrero and Thorpe 2016). For all other strains, only mating type and differences from W9100-17D are shown. Deletion consortium and temperature-sensitive strain collections are described by Winzeler et al. (1999) and Li et al. (2011). All gene disruptions were constructed by PCR amplification of the corresponding gene disruption from the S288C gene disruption library, using primers that include 300-400 bp of homology flanking each side of the dominant selectable marker. Gene disruption alleles containing the NatMX4 or HphMX4 marker were first constructed by transferring the KanMX4-marked disruption into the new strain, then marker swapping the dominant selectable marker by transformation with a PCR fragment containing the new marker amplified from plasmid pAG36 or pAG32, respectively (Addgene plasmids #35126 and #35122 from Goldstein and McCusker (1999)). All plasmids in this study were constructed using recombination in budding yeast by digesting an empty vector and integrating PCR product with at least 20 bp of homology on either side of the cut site. Parent plasmids are indicated with "†", "‡" and "‖" which correspond to pWJ1781, pWJ2496 and pWJ2509 respectively.

Chapter 3

CRISPR-mediated nonsense

Standard CRISPR-mediated gene disruption strategies rely on Cas9-induced DNA double-strand breaks (DSBs). Here, we show that CRISPR-dependent base editing efficiently inactivates genes by precisely converting four codons (CAA, CAG, CGA, and TGG) into STOP codons without DSB formation. To facilitate gene inactivation by induction of STOP codons (iSTOP), we provide access to a database of over 3.4 million single guide RNAs (sgRNAs) for iSTOP (sgSTOPs) targeting 97%–99% of genes in eight eukaryotic species, and we describe a restriction fragment length polymorphism (RFLP) assay that allows the rapid detection of iSTOP-mediated editing in cell populations and clones. To simplify the selection of sgSTOPs, our resource includes annotations for off-target propensity, percentage of isoforms targeted, prediction of nonsense-mediated decay, and restriction enzymes for RFLP analysis. Additionally, our database includes sgSTOPs that could be employed to precisely model over 32,000 cancer-associated nonsense mutations. Altogether, this work provides a comprehensive resource for DSB-free gene disruption by iSTOP.

3.1 Introduction

CRISPR-Cas9 technology allows the precise modification of genomic sequences and interrogation of gene function at unprecedented speed (Barrangou and Doudna 2016; Komor et al. 2017). This technology relies on the ability of CRISPR single guide RNAs (sgRNAs) to target the Cas9 endonuclease to precise genomic locations, where Cas9 introduces DNA double-strand breaks (DSBs) (Hsu et al. 2014). Current gene disruption strategies depend on the repair of Cas9-induced DSBs by non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Jasin and Haber 2016). During DSB repair, NHEJ occasionally introduces nucleotide insertions and deletions (indels) that result in frameshift mutations, thus disrupting gene open reading frames (ORFs), while HDR can lead to the integration of gene disruption cassettes at targeted loci (Ran et al. 2013). Gene disruption by NHEJ is efficient, although it does result in the generation of mosaic knockout (KO) alleles due to the variable number of nucleotides inserted or deleted prior to DSB end joining (van Overbeek et al. 2016). On the other hand, gene disruption by HDR is an accurate but inefficient process that requires exogenous DNA donor sequences as a template for DSB repair.

One common limitation of both NHEJ- and HDR-dependent gene editing approaches is their reliance on the formation of DSBs, which are toxic DNA lesions that can cause genomic rearrangements and translocations, activate DNA damage checkpoints, and induce cell death (Aguirre et al. 2016; Choi and Meyerson 2014; Frock et al. 2015; Ghezraoui et al. 2014; Roukos and Misteli 2014; Torres et al. 2014). Furthermore, while NHEJ and HDR have both been exploited to modify precise genomic sequences with CRISPR-Cas9 technology (Cong et al. 2013; Mali et al. 2013), these gene editing approaches can occasionally alter non-targeted genomic loci due to Cas9-induced DSB formation at off-target sites (Hsu et al. 2013).

As an alternative to NHEJ- and HDR-dependent genome editing, CRISPR-dependent editing strategies that entail direct modification of DNA bases have recently been developed (Hess et al. 2016; Komor et al. 2016; Ma et al. 2016; Nishida et al. 2016; Plosky 2016; Yang et al. 2016). Distinct from standard CRISPR-Cas9-dependent genome editing, CRISPR-mediated base editing avoids DSB formation and displays reduced genome-wide off-targeting (Kim et al. 2017a). CRISPR-dependent base editors consist of a catalytically inactive form of Cas9 or a Cas9 nickase mutant fused to cytidine deaminases, such as APOBEC1 or AID. For example, the CRISPR-dependent base editor BE3 is a fusion of rat APOBEC1 (rAPOBEC1), a uracil glycosylase inhibitor (UGI) and the Cas9-D10A nickase mutant (Komor et al. 2016). Following BE3 binding to target sites mediated by sgRNAs, rAPOBEC1 converts a targeted cytosine (C) into uracil (U) and UGI inhibits U removal by DNA glycosylases (Komor et al. 2016). The resulting G:U mismatch is then converted into an A:T base pair upon Cas9-mediated nicking of the G-containing DNA strand followed by DNA synthesis. This process generates permanent modifications of DNA bases within a window of high BE3 activity (13–17 nucleotides from the Cas9 protospacer adjacent motif or PAM) (Komor et al. 2016). Although recent studies have narrowed the activity window of BE3 variants from 5 to 1–2 nucleotides (Kim et al. 2017c), the presence of more than one cytosine within the BE3 activity window can result in the modification of multiple bases, causing potentially undesirable base substitutions (Kim et al. 2017c; Komor et al. 2016).

Base substitutions in the codons of ORFs can potentially cause amino acid substitutions (missense mutations) or generate premature STOP codons (nonsense mutations). Nonsense mutations can lead to the synthesis of truncated proteins or the degradation of mRNA transcripts by nonsense-mediated decay (NMD) and are often associated with human disease (Lykke-Andersen and Jensen 2015). In particular, inherited nonsense mutations in tumor suppressor genes are observed in 10%–30% of patients that suffer from hereditary cancer syndromes (e.g., familial breast and ovarian cancer, hereditary non-polyposis colorectal cancer, familial adenomatous polyposis) (Bordeira-Carriço et al. 2012). In addition, nonsense mutations account for 12% of sporadic non-silent mutations in tumor suppressors, as reported by the catalog of somatic mutations in cancer (COSMIC) (Forbes et al. 2017). Despite their importance in the pathogenesis of human

cancer, strategies to accurately and efficiently model tumor-associated nonsense mutations remain to be developed.

In this manuscript, we develop an application of CRISPR-mediated base editing for gene disruption studies. We show that CRISPR base editors efficiently inactivate human genes through the induction of STOP codons (iSTOP) in gene ORFs. This system relies on the ability of CRISPR-Cas9-dependent base editors to precisely convert four codons (CAA, CAG, CGA, and TGG) into STOP codons (TAG, TAA, or TGA) (**Figure 3.1**). The occurrence of iSTOP-mediated editing in sgRNA-transfected cellular populations or single clones can be rapidly monitored by restriction fragment length polymorphism (RFLP) assay using restriction enzymes that recognize bases targeted by iSTOP (**Figure 3.3B,D**). To facilitate the use of iSTOP, we have generated an online database to design nonsense generating sgRNAs (sgSTOPs) in eight eukaryotic species (available at http://www.ciccialab-database.com/istop). Notably, 94%–99% of genes in the analyzed eukaryotic genomes can be targeted with multiple sgSTOPs, highlighting the genome-wide potential of iSTOP. Furthermore, iSTOP can be employed to model over 32,000 nonsense mutations observed in human cancer. Altogether, our studies establish iSTOP as a gene disruption technology to study gene function and model human disease.

3.2 Results

3.2.1 C to T editors can create nonsense at four codons

To study the mutagenic properties of cytidine deaminase-dependent CRISPR base editors, we investigated the repertoire of mutations that can be generated by these enzymes from the conversion of cytosine to thymine (C > T) on the coding or non-coding strand (**Figures 3.2A,B & Table 3.1**). This analysis revealed that 87.5% (56/64) of the codons can be mutated by CRISPR-dependent base editors in either the coding or non-coding strand and 27% (20/74) of the changes introduced by these gene editing enzymes generate silent mutations (**Figure 3.2C & Table 3.1**). Despite the fact that CRISPR-dependent base editors can only generate 7.6% (32/420) of all the possible 420 non-silent substitutions (**Figures 3.1A,B & 3.2D**), these enzymes can convert four different codons (CAA, CAG, CGA, and TGG) encoding for 3 amino acids (glutamine [Gln], arginine [Arg], tryptophan [Trp]) exclusively into the TAA, TAG, and TGA STOP codons (**Figures 3.1B,C & 3.2D**). CAA, CAG, and CGA triplets can be converted into STOP codons when modified by CRISPR-dependent base editors on the coding strand, while TGG can be converted into STOP codons

if targeted on the non-coding strand (**Figures 3.1C & 3.2A,B**).

3.2.2 CRISPR-mediated base editing efficiently disrupts genes through iSTOP

To determine whether CRISPR-mediated base editing could be effectively used in human cells to convert CAA, CAG, CGA, and TGG into STOP codons, we designed sgSTOPs to target these four codons in the *SPRTN*, *FANCM*, *TIMELESS*, and *CHEK2* genes within the window of high activity (13–17 nucleotides from the PAM) of the CRISPR-dependent base editor BE3 (Komor et al. 2016) see **Figure 3.1C**. To simplify the detection of base editing, we targeted codons containing cytosines within the recognition sequences of restriction enzymes, thus allowing us to monitor base mutations that rendered restriction sites refractory to restriction digestion (**Figure 3.3A**). In these experiments, the sgSTOPs were transfected into HEK293T cells together with a DNA plasmid encoding the BE3 enzyme (**Figure 3.3A**). Three days after transfection, the genomic DNA of the transfected cell populations was extracted and the targeted genomic loci of *SPRTN*, *FANCM*, *TIMELESS*, and *CHEK2* were PCR-amplified. The amplicons were subsequently digested with restriction enzymes recognizing the targeted sites (**Figures 3.3A,B & 3.4A,B**). Remarkably, restriction fragment length polymorphism (RFLP) analysis revealed that 21%–35% of the PCR products specifically amplified from sgSTOP-transfected cell populations were insensitive to restriction enzyme digestion (**Figure 3.3B**). Complete digestion of PCR amplicons was observed upon incubation with restriction enzymes that did not recognize the targeted sites, confirming the specificity of this RFLP assay for the detection of sgSTOP-mediated editing (**Figures 3.4A–C**). Furthermore, a time course experiment showed that the incomplete digestion of PCR amplicons was not due to a limited incubation time of the reaction (**Figure 3.6A**). Cloning and sequencing of individual *SPRTN*, *FANCM*, *TIMELESS*, and *CHEK2* PCR products refractory to restriction enzyme digestion confirmed the presence of STOP-inducing C > T transitions in each of the four targeted codons (**Figure 3.3B**). Sequencing of a *TIMELESS* amplicon resistant to restriction enzyme digestion also revealed editing of another cytosine located within the window of BE3 activity, as previously observed in other BE3-edited loci (Kim et al. 2017c; Komor et al. 2016) see **Figure 3.3B**.

Besides causing restriction site loss, iSTOP-mediated editing can create new restriction sites, resulting in the generation of new DNA fragments upon digestion (**Figure 3.3A**). To test whether restriction site gain could be observed in iSTOP-edited cells, HEK293T were transfected with the *SPRTN* sgSTOP indicated above, which causes the simultaneous loss of a PvuII site and gain of an NheI site. Notably, a fraction of *SPRTN* PCR amplicons from *SPRTN* sgSTOP-transfected cells was digested by NheI, while being refractory to PvuII digestion, revealing that restriction site gain could be employed to monitor iSTOP-mediated

editing (**Figure 3.3C**). As shown in **Figure 3.3C**, a lower percentage of editing was observed by NheI digestion compared to PvuII digestion (30.7% versus 36.2%). This observation could depend on the fact that BE3-induced deamination of a second cytosine within the NheI restriction site (**Figure 3.3C**) could result in the disruption of NheI-mediated digestion. Collectively, these experiments demonstrate the feasibility of introducing nonsense mutations into the human genome by CRISPR-mediated base editing of CAA, CAG, CGA, and TGG codons and define a rapid RFLP assay to monitor the occurrence of base editing within populations of modified cells.

3.2.3 iSTOP allows the efficient generation of knockout human cell lines

Next, we determined whether our iSTOP approach is an efficient strategy to create human knockout cell lines. To this end, we targeted *SMARCAL1*, a gene mutated in Schimke immuno-osseous dysplasia that has been implicated in the DNA damage response (Bansbach et al. 2009; Ciccia et al. 2009; Postow et al. 2009; Yuan et al. 2009; Yusufzai et al. 2009). To disrupt the *SMARCAL1* gene, we utilized an sgSTOP that efficiently converts a CAG triplet located within the first exon of the *SMARCAL1* locus into a STOP codon (**Figure 3.5A**). To further increase iSTOP efficiency, we determined whether cells edited by iSTOP in the *SMARCAL1* locus could be enriched by using a marker-free co-selection strategy previously employed to select cells that have undergone genome editing at Cas9- or Cpf1-induced DSBs (Agudelo et al. 2017). This strategy relies on the simultaneous targeting of a gene of interest and the sodium-potassium channel *ATP1A1*, which when mutated in its first extracellular loop renders cells resistant to ouabain treatment (Agudelo et al. 2017). Using this approach, HEK293T cells were transfected with BE3 in combination with sgSTOPs targeting the first exon of *SMARCAL1* and the extracellular loop sequence of *ATP1A1* and then cultured in the presence or absence of ouabain (**Figure 3.5B**). Interestingly, the number of cells edited by iSTOP in the *SMARCAL1* locus increased from 24% to 38.8% upon ouabain selection, as determined by RFLP analysis (**Figure 3.5C**). To isolate *SMARCAL1* KO clones, ouabain-resistant cell populations transfected with *SMARCAL1* and *ATP1A1* sgSTOPs were seeded at single-cell density and isolated cell clones were screened by RFLP analysis (**Figures 3.5D & 3.6E**). From this analysis, we retrieved 10/19 (52.6%) clones with at least one edited SMARCAL1 allele, with 2/19 (10.5%) clones (#16 and #17) being homozygously edited (**Figure 3.6E**). DNA sequencing and protein blot analysis of the two *SMARCAL1* homozygously edited clones confirmed the presence of the desired STOP codon within the *SMARCAL1* gene and the absence of SMARCAL1 protein (**Figures 3.5E & 3.6F**). Sanger sequencing also confirmed the genotype of the heterozygous clones that we observed using the RFLP assay (**Figure 3.6G**). Although ouabain treat-

ment causes the accumulation of indels in the *ATP1A1* locus (**Figure 3.6D**) that can result in the generation of gain-of-function *ATP1A1* alleles (Agudelo et al. 2017), we only observed 1/11 modified clones with indels in the *SMARCAL1* locus (**Figure 3.6G**, see clone #18). This observation could be explained by the fact that indel formation depends on the DNA nickase activity of BE3, which is significantly less efficient than the cytidine deaminase activity of BE3 (Komor et al. 2016), thus rendering the probability of ouabain to select for an infrequent event at two independent sites remarkably rare. As further validation of the above co-selection strategy, we also observed enrichment in iSTOP-edited cells upon co-targeting of the *SPRTN* and *ATP1A1* genes, as determined by both restriction site loss and gain analysis and DNA sequencing of *SPRTN* amplicons from ouabain-resistant cell populations (**Figures 3.6B,C**). Taken together, these findings establish iSTOP as an efficient CRISPR-mediated base editing approach to disrupt human genes and demonstrate that iSTOP is compatible with co-selection enrichment methods.

3.2.4 iSTOP enables the disruption of human genes on a genome scale

As mentioned above, the use of iSTOP is restricted to four codons located at a specific distance from PAM sequences (**Figure 3.1C**). To determine whether, despite this limitation, iSTOP could be utilized to disrupt genes on a genome-wide scale, we identified sgSTOPs for all 69,180 ORFs reported in the human reference genome GRCh38 (**Methods section 3.4.11**). Using CDS coordinates from the UCSC genome browser, we first identified all CAA, CAG, CGA, and TGG codons and PAMs for all validated BE3 variants located at the correct distance from the targeted base (Kim et al. 2017c; Komor et al. 2016) (**Figure 3.7A**). This genome-wide analysis revealed that 62.5% of the CAA, CAG, CGA, and TGG codons in the human reference genome can be targeted by iSTOP, thus allowing the possibility to precisely convert 523,340 codons into STOP codons (**Figures 3.7C & 3.8C**). Remarkably, 98.6% of human ORFs, corresponding to 99.7% of human genes, can be targeted by one or more sgSTOPs, with over 80% of ORFs targetable within the first 20% of their sequence, 99.2% of ORFs targetable within the first 100 codons, and 94.7% of ORFs targetable at a position predicted to cause nonsense-mediated decay (Popp and Maquat 2016) (**Figures 3.7C,D & 3.8E**). An sgSTOP can be designed for approximately one out of every 26 codons such that the typical gene of 600 amino-acid residues can potentially be targeted at 23 different codons (**Figures 3.7B & 3.8D**). From this analysis, we observed that only 68 (0.35%) human genes cannot be targeted by iSTOP due to unavailable CAA, CAG, CGA, or TGG codons or PAM sequences (**Figures 3.8A & 3.8B**). However, 24/68 of the non-targetable human genes have eukaryotic orthologs that can be targeted by iSTOP, thus allowing the study of these genes in model organisms (**Figure 3.8A & Table 3.2**). To determine the off-target

propensity of iSTOP guides, we performed a genome-wide search for sequences similar to each sgSTOP. When allowing up to 2 mismatches outside of the guide's seed sequence, we observe that 74% of NGG PAM-based sgSTOPs map uniquely to the genome (**Figure 3.8F**). To aid in selecting sgSTOPs with low off-target propensity, we identified putative off-target sites in the human genome for each sgSTOP. This analysis demonstrates that almost all human ORFs can be targeted by iSTOP, thus highlighting iSTOP as a promising application of CRISPR-mediated base editing for genome-wide gene disruption studies in human cells.

To determine whether the RFLP assay described above could be generalized to monitor the efficacy of multiple sgSTOPs, we identified restriction sites of enzymes that recognize DNA bases targetable by iSTOP in the human genome (**Table 3.4**). This analysis revealed that over 70% of targetable sites, corresponding to 98% of human genes, can be monitored by either restriction site loss or gain using enzymes that do not cut within 50 nucleotides from the targeted base (**Figure 3.7E**). Furthermore, this analysis predicted BsrI and BfaI as two of the most common restriction enzymes for detecting iSTOP-edited targets (>15,000 sgSTOPs) in the human genome by restriction site loss or gain, respectively (**Figure 3.7F**). We validated the use of BsrI in the RFLP assay by targeting the PARP4 gene with an sgSTOP that edits a BsrI restriction site (**Figure 3.4C**). These studies indicate that the RFLP assay described here allows rapid assessment of base editing activity for the majority of sgSTOPs, thus greatly simplifying the use of iSTOP to interrogate gene functions in human cells.

3.2.5 iSTOP is compatible with genome-wide gene disruption studies

To determine whether iSTOP could be utilized for genome-wide studies in other eukaryotic species, we identified sgSTOPs for all ORFs of 7 other eukaryotic model organisms, ranging from *S. cerevisiae* to *M. musculus*. An additional 2,490,479 codons can be targeted in these species, with the possibility to design sgSTOPs approximately every 26 codons in each organism, with the exception of *A. thaliana* and *S. cerevisiae*, which can be targeted approximately once every 32 codons (**Figure 3.8C**). Moreover, we found that the percentage of ORFs that can be targeted by iSTOP is between 97% (*S. cerevisiae*) and 99.7% (*M. musculus*), demonstrating the robustness of this method to target almost any ORF of various models (**Figure 3.7G**). To facilitate the use of iSTOP, we provide a full list of sgSTOP sequences, genomic coordinates of each targeted base and targeted codons for any gene in the examined eukaryotic species (**Methods section 3.4.11**). In addition, this database contains information on the number and percentage of alternatively spliced isoforms targeted by each sgSTOP, a prediction of nonsense-mediated decay, counts of putative off-

target sites for each sgSTOP and a list of restriction enzymes that can be used to monitor iSTOP-mediated editing. This database contains a total of 3,483,549 targetable gene coordinates and is available online at http://www.ciccialab-database.com/istop.

3.2.6 iSTOP enables the modeling of cancer-associated nonsense mutations

Nonsense mutations are frequently associated with human diseases. In cancer, nonsense mutations account for 4%–5% of the total number of observed mutations in coding sequences. Cytidine deaminases of the APOBEC protein family are frequently overexpressed in cancer and are responsible for up to 68% of all nonsense and missense mutations observed in certain cancer types (Roberts and Gordenin 2014). To determine whether iSTOP could be utilized to model cancer-associated nonsense mutations due to cytidine deamination, we determined the prevalence of C > T and G > A base transitions occurring at CAA, CAG, CGA, and TGG codons observed in COSMIC (**Figures 3.9A–C**). When considering each mutation site once, this analysis revealed that over 50% of nonsense mutation sites across all cancers occur at CAA, CAG, CGA, and TGG codons, with 32,723 (61%) of these nonsense mutation sites being directly reproducible using iSTOP (**Figure 3.9A**). We then determined which genes within each cancer type incurred nonsense mutations in CAA, CAG, CGA, and TGG codons more often than expected – we refer to these genes as frequent iSTOPers (**Figures 3.9B,C & 3.10A**). The top 100 iSTOPers include 37 tumor suppressor genes (TSGs) listed in the Cancer Gene Census (e.g., *TP53*, *PTEN*, *VHL*, *APC*, *ATM*, and *SETD2*) and an additional 13 putative TSGs identified by previously published methods, such as TUSON, MutSigCV, and 20/20+ (Davoli et al. 2013; Lawrence et al. 2013; Tokheim et al. 2016) (**Figure 3.10B**). Of the 120 frequent iSTOPers ($q < 0.001$), 118 have at least one cancer-associated nonsense mutation that can be directly modeled with iSTOP (**Table 3.4**). Using iSTOP, we successfully modeled the most recurrent nonsense mutation (R348*; 26 cases in COSMIC) in the iSTOPer *PIK3R1*, which is known to promote tumorigenesis (Jaiswal et al. 2009) (**Figure 3.10C**). This work establishes iSTOP as a promising approach to model cancer-associated nonsense mutations for functional studies.

3.3 Discussion

In this study, we describe iSTOP as a CRISPR-mediated base editing methodology that converts CAA, CAG, CGA and TGG codons into STOP codons, thus enabling gene disruption and facile generation of KO cells. Moreover, we describe an RFLP assay for easy monitoring of the efficiency of base editing in cell popu-

lations and clones. Additionally, we provide access to an online resource containing 3,483,549 targetable gene coordinates in eight different eukaryotic species. This resource includes useful annotations to aid in searching and selecting sgSTOPs. These annotations include off-target propensity, nonsense-mediated decay prediction, relative position in the gene, percentage of isoforms targeted and a list of restriction enzymes that can be used to monitor iSTOP-mediated editing. Finally, we provide software to design base-editing guides for any annotated genome and any PAM. Collectively, these studies greatly expand the targeting range recently described for iSTOP-based approaches (Kuscu et al. 2017).

3.3.1 Advantages of iSTOP over standard gene disruption strategies

iSTOP exhibits several properties that distinguish it from standard CRISPR-mediated base editing applications for correcting/inserting targeted missense mutations. Given that iSTOP causes the disruption of genes and not the creation/correction of precise missense gene variants, (1) iSTOP is not affected by potentially undesired cytosine deamination events catalyzed by BE3 within its activity window. Furthermore, whereas standard CRISPR-mediated approaches of genome base editing can generate distinct missense mutations from the same targeted codon (**Table 3.1**), (2) iSTOP induces the precise conversion of four codons exclusively into STOP codons. These properties, combined with the fact that over 3 million targetable codons in eight eukaryotic species can be precisely converted into STOP codons by iSTOP to allow targeting of 97%–99.7% of eukaryotic genes, render iSTOP the first CRISPR-mediated base editing application compatible with genome-scale analyses. In addition, iSTOP presents several advantages compared to traditional CRISPR-Cas9-mediated gene disruption technology. Indeed, (1) iSTOP does not induce DSBs, whose formation is particularly deleterious in human cells (Kuscu et al. 2017) and during organismal development (Harrison et al. 2014). By avoiding DSB formation, iSTOP leads to reduced cell death compared to conventional CRISPR-Cas9-based methods (Kuscu et al. 2017), thus potentially facilitating the generation of KO animals, as recently shown for CRISPR-mediated strategies of genome base editing in mouse embryos (Kim et al. 2017b). Furthermore, (2) iSTOP disrupts genes by precisely editing DNA bases, without relying on NHEJ-induced frameshift mutations, which create a mosaic population of distinct KO alleles. Absence of mosaicism has in fact been observed in mice subjected to CRISPR-mediated base editing (Kim et al. 2017b). Moreover, (3) iSTOP does not require the use of synthetic DNA donor molecules employed for HDR-dependent genome editing, thus rendering iSTOP-gene disruption strategies simpler and more efficient than HDR-dependent approaches, particularly for the genetic engineering of plants, for which the use of DNA donor templates is limited by technical and legislative obstacles (Jones 2015). Recent studies have

indeed shown that CRISPR-dependent approaches of genome base editing are highly efficient in rice, wheat, tomatoes, and maize (Shimatani et al. 2017; Zong et al. 2017).

3.3.2 Limitations of iSTOP

At present, one of the main limitations of iSTOP arises from the restrictive rules for designing sgSTOPs. Unlike canonical Cas9/ sgRNAs, which can be programmed to generate DSBs adjacent to any PAM sequence, BE3/sgSTOPs require the presence of CAA, CAG, CGA, and TGG codons and a PAM located 13–17 bps away from the targeted base(s). Although 60% of all available CAA, CAG, CGA, and TGG codons are targetable by iSTOP, the remainder cannot currently be targeted due to the absence of a nearby PAM sequence (**Figures 3.8B,C**). We anticipate that the generation of additional BE3 variants and the conversion of other RNA-guided DNA nickases into base editors that recognize different PAM sequences will increase the number of codons targetable by iSTOP. An additional limitation of iSTOP is due to the inability of BE3 to efficiently edit cytosines with a G on the immediate 5′ side, as recently shown by Komor et al. (2016). As a result, iSTOP is predicted to have reduced editing efficiency at 24% of targetable sites, although future improvements in the catalytic activity of BE3 may mitigate this effect. Similar to other CRISPR-related technologies, the efficacy of iSTOP could be affected by the occurrence of off-target mutations. Notably, initial work that examined the specificity of the cytidine deaminase activity of BE3 using a BE3 variant lacking the UGI domain (BE3ΔUGI) concluded that BE3 causes fewer off-targets than Cas9 (Kim et al. 2017a). Additional studies on BE3-dependent off-targets led to the development of a high-fidelity BE3 enzyme (HF-BE3) with reduced off-target base editing activity (Rees et al. 2017). Despite these important studies, it remains to be determined whether the nickase activity of BE3 might also contribute to the generation of off-targets. To limit the use of sgSTOPs with potentially higher off-target activity, we have estimated the off-target propensity of all our predicted sgSTOPs and have designed filters in our online database that allow users to exclude the most promiscuous sgSTOPs targeting their genes of interest.

3.3.3 Potential iSTOP applications

iSTOP provides possible strategies to address currently challenging biological questions. Indeed, by disrupting gene ORFs without altering gene structures, (1) iSTOP could allow the separation of coding from non-coding functions of genes, such as *ASCC3* (Williamson et al. 2017), that encode for both proteins and long non-coding RNAs (lncRNAs), and lncRNAs that contain putative alternative short ORFs (Andrews

and Rothnagel 2014). In addition, (2) iSTOP could be employed to incorporate modified or non-natural amino acids into proteins using tRNAs that suppress the newly introduced STOP codons (Elsässer et al. 2016). Moreover, by enabling the modeling of nonsense mutations at a genome-wide level, (3) iSTOP could allow genome-wide studies to investigate eukaryotic gene functions and (4) facilitate the identification of pathogenic variants in cancer through large-scale functional studies of cancer-associated mutations. Recent work has indeed confirmed the promising potential of iSTOP-related approaches for pooled screening applications (Kuscu et al. 2017). Altogether, our work establishes iSTOP as a robust and efficient gene disruption technology compatible with genome-wide studies to investigate eukaryotic gene functions and model human diseases.

3.4 Methods

3.4.1 Cell lines

HEK293T cells (ATCC) were cultured in DMEM supplemented with 10% Fetalgro bovine growth serum (BGS, RMBIO) and 1X penicillin-streptomycin (Life Technologies). Cells were grown at 37°C with 5% CO_2. HEK293T cells were tested for mycoplasma.

3.4.2 Transfection

HEK293T cells were seeded at 50%–70% confluency into 24-well plates and reverse transfected by mixing 0.5 µg of sgRNA and 1 µg of BE3 plasmid into 100 µL of DMEM without Fetalgro BGS and antibiotics. 4.5 µL of TransIT-293 (Mirus) was added to the DNA, mixed and incubated for 15 min at room temperature. The DNA-transfection reagent mix was added drop-wise to the cells and incubated at 37°C with 5% CO_2 for 3 days.

3.4.3 Vector construction

The B52 plasmid used to express sgRNAs in human cells is a derivative of the MLM3636 construct (Addgene #43860). B52 contains a second sgRNA-expressing cassette, which was PCR amplified from the plasmid eSpCas9 (1.1) (Addgene #71814) using the primers PB96 and PB97 (**Table 3.5**) and cloned into the MLM3636 plasmid using the restriction enzymes HindIII-HF (NEB #R3104S) and XhoI (NEB #R0146S).

The final B52 plasmid contains a ColE1 origin, an ampicillin resistance gene and two sgRNA expression cassettes with either BsmBI or BbsI sites for sgRNA cloning (see below). The B52 plasmid has been further modified to contain EcoRI and XbaI sites for cloning additional sgRNA-expressing cassettes. This plasmid is available in Addgene.

3.4.4 sgSTOP design and cloning

sgSTOPs were synthesized as oligonucleotide pairs (IDT) compatible with BbsI and BsmBI restriction sites. The oligonucleotide sequences were designed as follows: 5′-ACACCG(N)20G-3′ and 5′-AAAAC(N)20CG-3′ for BsmBI restriction sites; 5′-CACCG(N)20-3′ and 5′-AAAC(N)20C-3′ for BbsI restriction sites, where "(N)20" corresponds to each sgRNA sequence. The following oligonucleotides were used for sgSTOP cloning: *SPRTN* (PB551 and PB552), *SMARCAL1* (PB580 and PB581), *CHEK2* (PB732 and PB733), *PARP4* (PB734 and PB735), *FANCM* (PB736 and PB737), *TIMELESS* (PB738 and PB739), and *PIK3R1* (PB776 and PB777). Oligonucleotide sequences are available in **Table 3.5**. Oligonucleotide pairs were resuspended in TE (100 µM final concentration) and annealed in the following reaction buffer: 6.5 µL water, 2 µL 5X T4 ligase buffer (Life Technologies), 0.5 µL T4 PNK (NEB #M0201L) and 0.5 µL of each oligonucleotide (100 mM; IDT). The reaction was conducted for 1 hr at 37°C followed by incubation for 5 min at 95°C and gradual temperature decrease from 95°C to 15°C. sgSTOP oligonucleotides were cloned into the B52 plasmid, which was digested with either BbsI-HF (NEB #R3539L) or BsmBI (NEB #R0580L) in a 20 µL reaction according to the manufacturer's recommendations. 0.5 µL of recombinant shrimp alkaline phosphatase (NEB #M0371L) were then added to the digestion reaction and incubated for 1 hr at 37°C to dephosphorylate the plasmid. The linearized plasmid was gel purified and incubated with the sgSTOP oligonucleotides for 1 hr at room temperature in the following ligation reaction: 50 ng of digested plasmid, 0.5 µL phosphorylated and annealed sgRNAs (diluted 1/100 in water), 1 µL 5X ligase buffer (Life Technologies) and 0.25 µL of T4 ligase. Ligation products were sequenced to confirm the integration of the sgRNA using the oligonucleotides PB518 or PB573 (Table S5) for BsmBI or BbsI cloning, respectively.

3.4.5 RFLP assay and DNA sequencing

The occurrence of iSTOP-mediated editing was monitored three days after transfection of HEK293T cells with BE3 and sgSTOPs, as indicated above. Cells were recovered by trypsin (Life Technologies) and washed with PBS. The cell pellet was resuspended in the Quick Extract DNA Extraction Solution (Epi-

center) and heated sequentially at 65°C for 5 min and 95°C for 5 min to isolate genomic DNA (gDNA). The isolated gDNA was quantified using Nanodrop, diluted in water and stored at -20°C or directly used in PCR reactions. The following genomic loci were targeted by sgSTOPs for *SPRTN* (chr1+231338561), *SMARCAL1* (chr2+216414804), *CHEK2* (chr22 - 28734431), *TIMELESS* (chr12-56431493) and *FANCM* (chr14+45164491). PCR primers were designed by Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0) using the default parameters with the following changes: Mispriming library = "HUMAN," Primer size "min = 25, Opt = 27, Max = 30," Primer Tm "Min= 57.0, Opt = 60.0, Max = 63.0". PCR reactions were prepared in a 25 µL reaction volume containing: 1 µM forward and reverse primers, 0.1 nM dNTPs (NEB #N0447L), 1X Q5 buffer (NEB), 200-300 ng gDNA, 1 unit Q5 polymerase (NEB) and water. PCR reactions were conducted as follows: 95°C for 1 min, 30 cycles of 95°C for 10 s, melting temperature (Tm) for 10 s and 72°C for 45 s and a final step at 72°C for 1 min. The following oligonucleotides were utilized for this reaction: *SPRTN* (PB571 and PB572, Tm = 64°C), *SMARCAL1* (PB590 and PB591, Tm = 60°C), *ATP1A1* (PB711 and PB712, Tm = 63°C), *CHEK2* (PB740 and PB741, Tm = 62°C), *PARP4* (PB742 and PB743, Tm = 64°C), *FANCM* (PB744 and PB745, Tm = 63°C), *TIMELESS* (PB746 and PB747, Tm = 61°C), and *PIK3R1* (PB782 and PB783, Tm = 60°C). The sequence of each oligonucleotide is available in **Table 3.5**. 4 mL of PCR products were digested until completion with the indicated restriction enzymes (NEB) in a 20 µL reaction. For BsrI digestions, PCR products were purified on columns before digestion. In the case of SfaNI, reactions were heated at 65°C for 30 min to denature the enzyme after digestion. After digestion, 5 µL of 6X loading buffer (NEB) was added to each reaction and 2 µL of the mixture was loaded on a 6% TBE polyacrylamide gel. Gels were run at 160 V in 1X TBE buffer, then washed in 1X TBE and stained for 5 min using SybrGold diluted in 1X TBE buffer. Gels were developed using LI-COR Odyssey. To confirm base editing by Sanger sequencing, 6 to 8 PCR reactions were performed as indicated above, pooled together, purified on columns and digested in 20 µL reaction using the respective restriction enzymes. DNA fragments refractory to digestion were gel purified and cloned using the pCR-BLUNT II-TOPO kit (Life Technologies). The cloned products were sequenced by Eton Bioscience using the universal T7 promoter primer (5′-TAATACGACTCACTATAGGG-3′). Sanger sequencing data were analyzed using Serial Cloner and SnapGene Viewer. The sequencing profiles shown in this manuscript were generated by SnapGene Viewer.

3.4.6 ATP1A1 co-selection strategy

To conduct the marker-free co-selection strategy based on ATP1A1 targeting (Agudelo et al. 2017), we generated the plasmid B270 expressing an sgRNA against *ATP1A1* and an empty sgRNA-expressing cassette.

B270 was derived from B52 upon cloning of *ATP1A1* sgRNA oligonucleotides (PB609 and PB610) into the BsmBI restriction site, as detailed above. sgRNA oligonucleotides for *SMARCAL1* (PB580 and PB581) or *SPRTN* (PB551 and PB552) were introduced into the B270 plasmid using BbsI cloning sites. Transfection of HEK293T cells was conducted as follows: 6-well plates were seeded at approximately 70% cell confluency and HEK293T cells were reverse transfected by mixing 1 µg of sgRNA and 2 µg of BE3 plasmid into 600 µL of DMEM without Fetalgro BGS and antibiotics. Four days after transfection, the cells were split in 3x6-well plates and treated with 1 µM ouabain. Untreated cells were plated in 12-well plates and grown in parallel. 48 hr after treatment, ouabain was removed from the medium and the surviving cells were cultured in regular DMEM medium for 2 days, prior to a new 1 µM ouabain treatment for 2 additional days. The surviving cells were amplified over the course of 2 weeks and untreated cells were grown in parallel as controls.

3.4.7 Western blot

HEK293T cell pellets were resuspended in 30 µL of PBS and 30 µL of 2X lysis buffer (100 mM Tris pH 6.8, 4% SDS and 1.716 M 2-mercaptoethanol) and incubated for 10 min at 95°C. Cell extracts were then diluted in 1X loading buffer (NuPage, Life Technologies) and heated for 5 min at 95°C. Samples were run on an 8% polyacrylamide gel at 160 V in Tris-glycine buffer. Gels were subsequently transferred onto nitrocellulose membranes, which were stained with ponceau solution (Sigma) and then incubated for 30 min with 5% milk (Bio Basic) in TBS 0.1% Tween 20 (TBS-T). Membranes were then incubated with primary antibodies in TBS-T supplemented with 1% milk for 2 hr at room temperature. The following antibodies were used for western blotting: α-SMARCAL1 (1/500) and α-GAPDH (1/1000) diluted in TBS-T with 1% milk (Bio-Rad). Membranes were then washed three times in TBS-T and incubated for 1 hr with α-mouse IgG secondary antibodies coupled with HRP at 1/5000 dilution in TBS-T/1% milk. Membranes were subsequently washed three more times in TBS-T and the HRP signal was detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) and autoradiography films (Southern Labware).

3.4.8 Genome-wide iSTOP analysis

To identify iSTOP targetable sites, coding sequence was constructed from genomes using available CDS coordinates. Each coding sequence was verified to have a start and STOP codon, no internal STOP codons and a sequence length evenly divisible by three. Coding sequence coordinates for CAA, CAG, CGA and TGG codons were then mapped to genome sequence coordinates to then extract the genomic sequence

context of each targeted base. sgSTOPs were designed by searching for each PAM 13 to 17 bps downstream of each targeted C. To count putative off-target sites, all sgSTOPs (including PAMs) were aligned to the genome, allowing up to two mismatches in the first eight bases of the guide sequence. Alignment criteria were chosen as a compromise between computational efficiency and the reported mismatch tolerance of Cas9 (Cho et al. 2014). Nonsense-mediated decay prediction was determined based on whether the targeted base was 55 nucleotides upstream of the final exon-exon junction (Popp and Maquat 2016). All genomes and CDS coordinates are from the UCSC genome browser (https://genome.ucsc.edu), with the exception of the *A. thaliana* genome, which originates from TAIR (https://www.arabidopsis.org), and *A. thaliana* CDS coordinates, which originate from Ensembl BioMart (http://www.biomart.org). All computational analysis was performed in R (https://www.r-project.org) via RStudio (https://www.rstudio.com), making extensive use of Tidyverse (http://www.tidyverse.org), and Bioconductor (https://bioconductor.org) R packages. A step-by-step guide to reproduce the figures and computational analyses found in this paper is available on GitHub (https://github.com/CicciaLab/iSTOP-paper).

3.4.9 Quantification of iSTOP-mediated editing

The percentage of iSTOP-mediated editing (% editing) was estimated by quantifying digested and undigested bands in RFLP assays using ImageJ. iSTOP-mediated editing resulting in the loss of a restriction site was estimated by the following formula: intensity of the undigested band divided by the combined intensity of digested and undigested bands. iSTOP-mediated editing resulting in the gain of a restriction site was quantified as follows: intensity of the digested bands divided by the combined intensity of digested and undigested bands.

3.4.10 Analysis of frequent iSTOPers

Genes observed to incur frequent cancer-associated nonsense mutations at CAA, CAG, CGA and TGG codons (iSTOPers) were scored using a one-tailed binomial test. For this test, the probability of "success" for each gene is estimated based on the number of nonsense mutations observed at CAA, CAG, CGA and TGG codons in a given cancer type and the number of CAA, CAG, CGA and TGG codons present in the gene. The number of "successes" is the number of CAA, CAG, CGA and TGG codons in the gene observed to incur nonsense mutations in a given cancer type. Mutated coordinates in the COSMIC database were only counted once to avoid bias from duplicated entries. We report unadjusted p-values and FDR corrected

q-values to account for multiple tests (Benjamini and Hochberg 1995). An “All cancers” category is included as a summary of the lowest q-value observed for a gene across all cancer types. High scoring iSTOPers are shown in **Figures 3.9B,C**. Genes/cancer type combinations with a q-value less than 0.001 are included in **Table 3.4**.

3.4.11 Data and software availability

All source code necessary to reproduce figures and computational analyses is available on GitHub (https://github.com/CicciaLab/iSTOP-paper). An R package that enables iSTOP detection for any DNA sequence with annotated CDS coordinates and customizable PAM/codon/spacing parameters is available on GitHub (https://github.com/CicciaLab/iSTOP). Detailed installation and usage instructions for both of these projects are available on GitHub. An interface for convenient search of sgSTOPs is available online (http://www.ciccialab-database.com/istop). Comprehensive datasets of pre-designed and annotated sgSTOPs for 8 organisms and 6 PAMs are available on Mendeley Data (http://dx.doi.org/10.17632/xbdtvf6bvj.1). Plasmids created in this study will be made available from Addgene. Raw gel images are available on Mendeley Data (http://dx.doi.org/10.17632/jw5rjmypy2.1).

3.5 Figures and tables

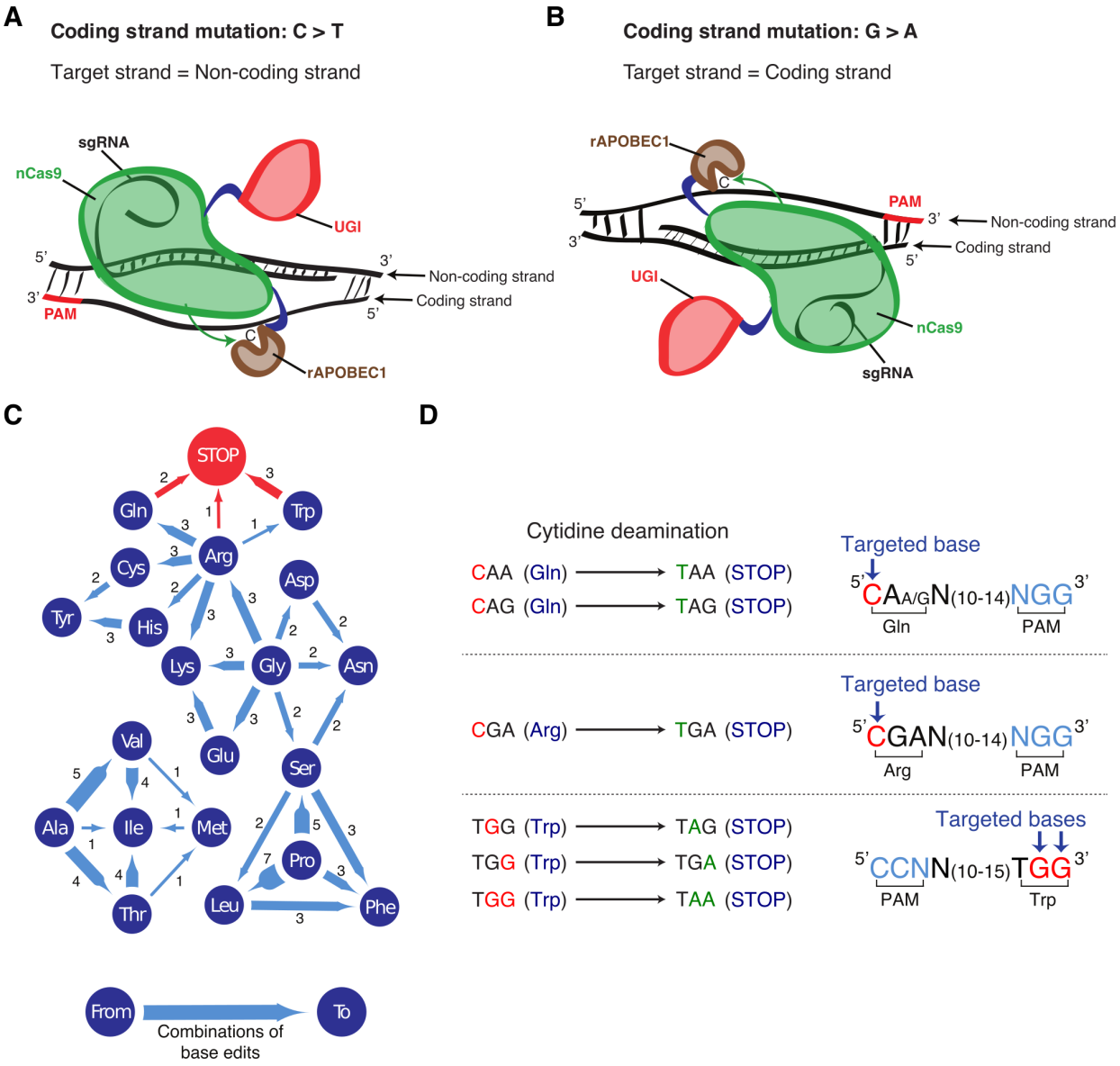

Figure 3.1 – Generation of STOP codons via CRISPR-mediated base editing. (**A**) Representation of amino acid substitutions generated by cytidine deaminase-dependent CRISPR base editors. Dark blue circles indicate amino acids, blue lines show the direction of amino acid substitutions induced by CRISPR-dependent base editing and the number of possible combinations to obtain the indicated substitutions. See also **Figure 3.2C,D & Table 3.1**. (**B**) Representation of the cytidine deamination reactions induced by CRISPR-dependent base editors to generate STOP codons. The CRISPR base editor BE3 converts CAA, CAG, CGA, and TGG codons into STOP codons when the targeted base(s) (red) is at the correct distance (13–17 bps) from a protospacer adjacent motif (PAM, blue). Schematics in panels **A**, **B** and **D** were created by Pierre Billon. Related to **Section 3.2.1**.

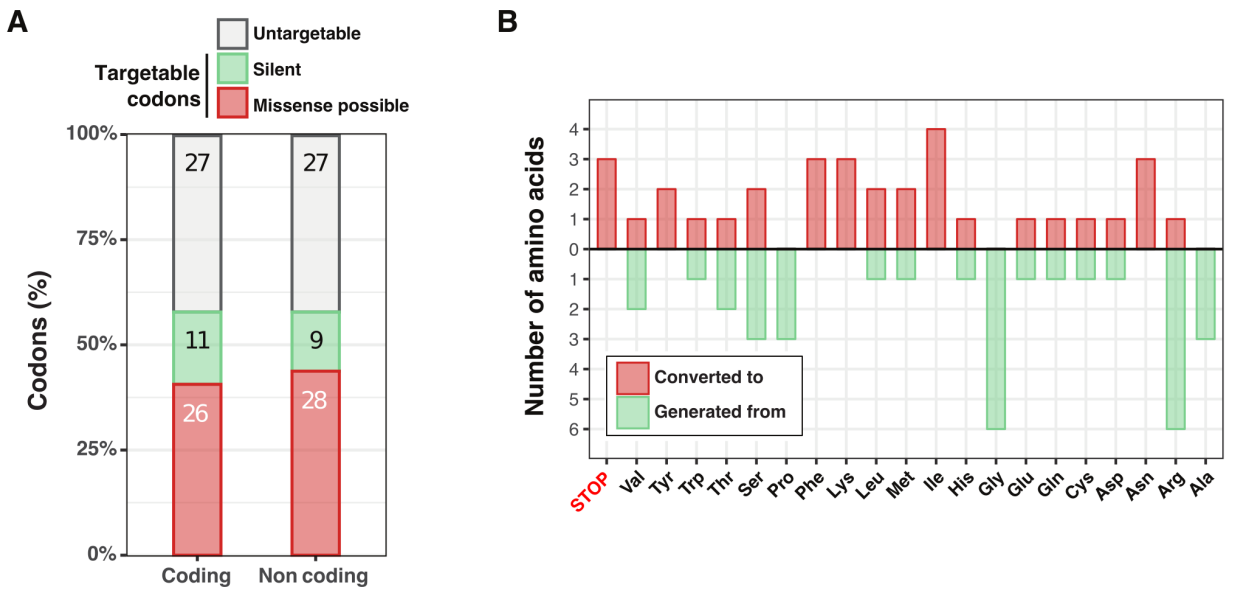

Figure 3.2 – Supplement to Figure 3.1 – Repertoire of amino acid substitutions generated by CRISPR-mediated base editing on coding or non-coding strands. (**A**) Percentage and number of targetable and untargetable codons and type of mutations generated on the 64 codons by targeting cytidine deaminase-dependent CRISPR base editors on either coding or non-coding strands. Codons untargetable by CRISPR base editors are indicated in white. Codons that when targeted cause silent or possible missense mutations are indicated in green and orange, respectively. (**B**) Number of amino acids that can be converted into other amino acids/STOP codons (orange) or generated from different amino acids (green) by cytidine deaminase-dependent CRISPR base editors, as shown in **Figure 3.1C**. Ala = Alanine, Arg = Arginine, Asn = Asparagine, Asp = Aspartic acid, Cys = Cysteine, Gln = Glutamine, Glu = Glutamic acid, Gly = Glycine, His = Histidine, Ile = Isoleucine, Met = Methionine, Leu = Leucine, Lys = Lysine, Phe = Phenylalanine, Pro = Proline, Ser = Serine, Thr = Threonine, Trp = Tryptophan, Tyr = Tyrosine, Val = Valine and STOP = STOP codon. See also **Table 3.1**. Related to **Section 3.2.1**.

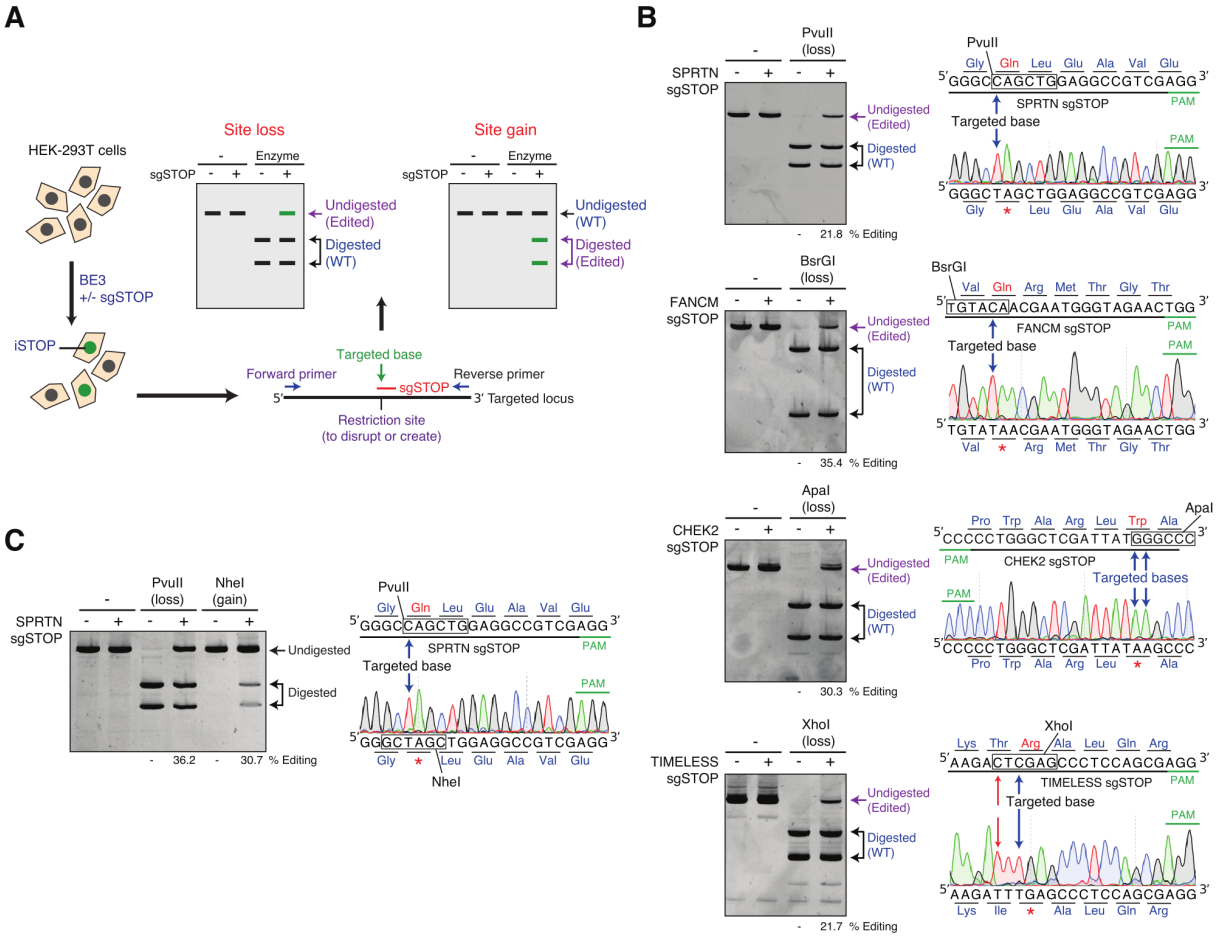

Figure 3.3 – Restriction fragment length polymorphism assay to detect iSTOP-edited cells. (**A**) Schematic representation of the protocol utilized to disrupt genes by iSTOP. HEK293T cells are transfected with BE3 with or without sgSTOP, thus resulting in the generation of cells edited by iSTOP (green). The targeted locus is then amplified by PCR and digested with a restriction enzyme that recognizes a restriction site containing the base targeted by iSTOP or a restriction site generated by iSTOP-mediated base editing. Base editing by iSTOP results in PCR products refractory to restriction digestion (Site loss) or induces the formation of new restriction sites (Site gain). (**B**) Conversion of CAG, CAA, TGG, and CGA codons into STOP codons by iSTOP in human cells. PCR products amplified from four different genomic loci (*SPRTN*, *FANCM*, *CHEK2*, and *TIMELESS*) edited by iSTOP were subjected to restriction digest with enzymes (PvuII, BsrGI, ApaI, and XhoI) that recognize sites containing the targeted bases. Products of the restriction digest reactions were run on polyacrylamide gels (left) and editing efficiency was determined by the percentage of undigested PCR amplicons (purple). Sequencing profiles of undigested PCR products (right) are compared to the wild-type genomic sequence, including the targeted base (blue arrow) and PAM (green). Editing of a cytosine that generates a missense mutation in the *TIMELESS* locus is indicated by red arrows. The sequencing profiles are representative of four to eight sequences for each targeted locus. Each base is colored according to the sequencing peaks (A in green, G in black, T in red, and C in blue). PAM, protospacer adjacent motif. See also **Figure 3.4A,B**. (**C**) Detection of iSTOP-induced events by both loss and gain of restriction sites. The *SPRTN* locus was PCR-amplified and left undigested or digested with either PvuII or NheI (left). Sequencing profiles of PCR amplicons from sgSTOP-treated cells, including the targeted base (blue arrow) and PAM (green), are shown on the right. Editing efficiency was determined by the percentage of PCR amplicons refractory to PvuII digestion, as in (B), or by the percentage of NheI-digested PCR amplicons. Experiments performed by Pierre Billon, Tarun Nambiar, and Samuel Hayward. Analysis performed by Pierre Billon. Related to **Section 3.2.2**.

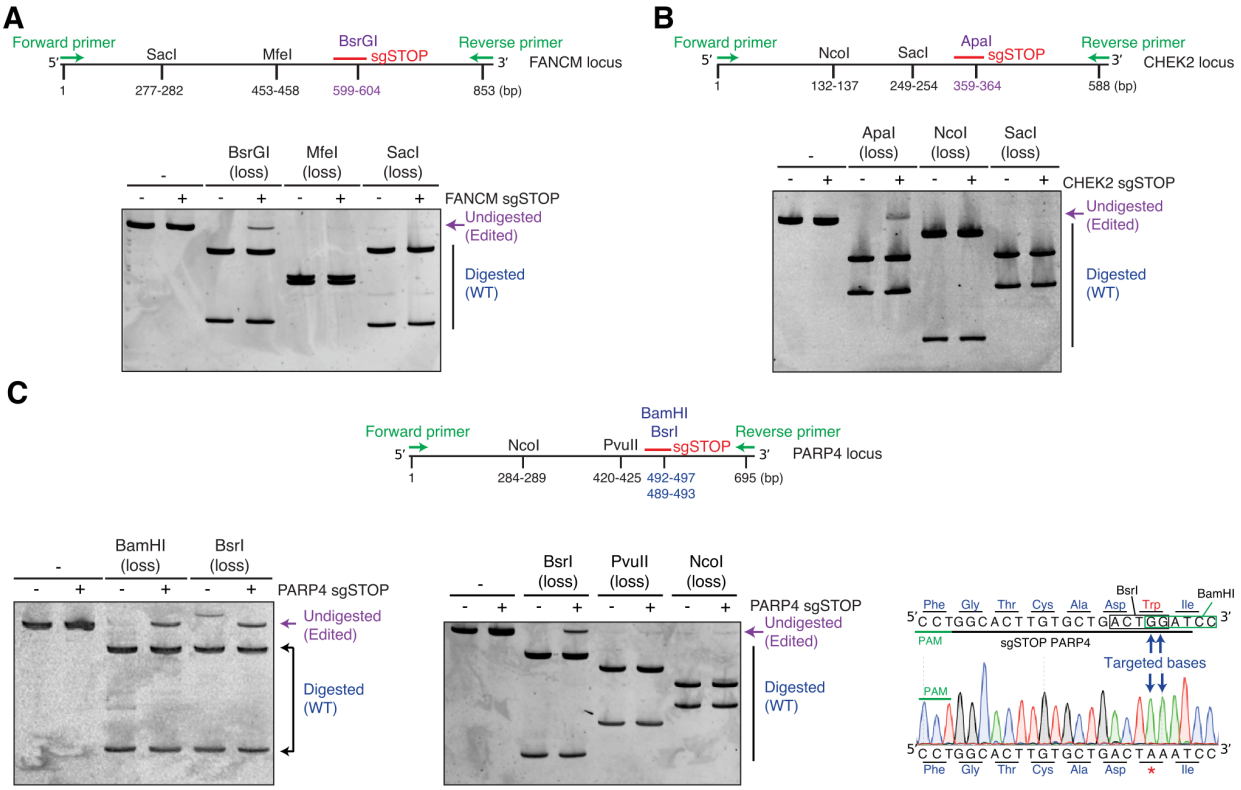

Figure 3.4 – Supplement to Figure 3.1 – Specificity of the RFLP assay utilized to detect iSTOP-mediated editing. (**A**) BsrGI-, MfeI- and SacI-mediated digestion of PCR products of the *FANCM* locus targeted with an sgSTOP that edits a BsrGI restriction site. A schematic map of the *FANCM* locus is indicated above. (**B**) ApaI-, NcoI- and SacI-mediated digestion of PCR products of the *CHEK2* locus targeted with an sgSTOP that edits an ApaI restriction site. A schematic map of the *CHEK2* locus is indicated above. (**C**) BamHI-, BsrI-, PvuII- and NcoI-mediated digestion of PCR products of the *PARP4* locus targeted with an sgSTOP that edits BamHI and BsrI restriction sites. A schematic map of the *PARP4* locus is indicated above. One sequencing profile representative of 4 sequences of *PARP4* amplicons refractory to BsrI digestion is shown on the right inside. Experiments performed by Pierre Billon, Tarun Nambiar, and Samuel Hayward. Analysis performed by Pierre Billon. Related to **Section 3.2.2**.

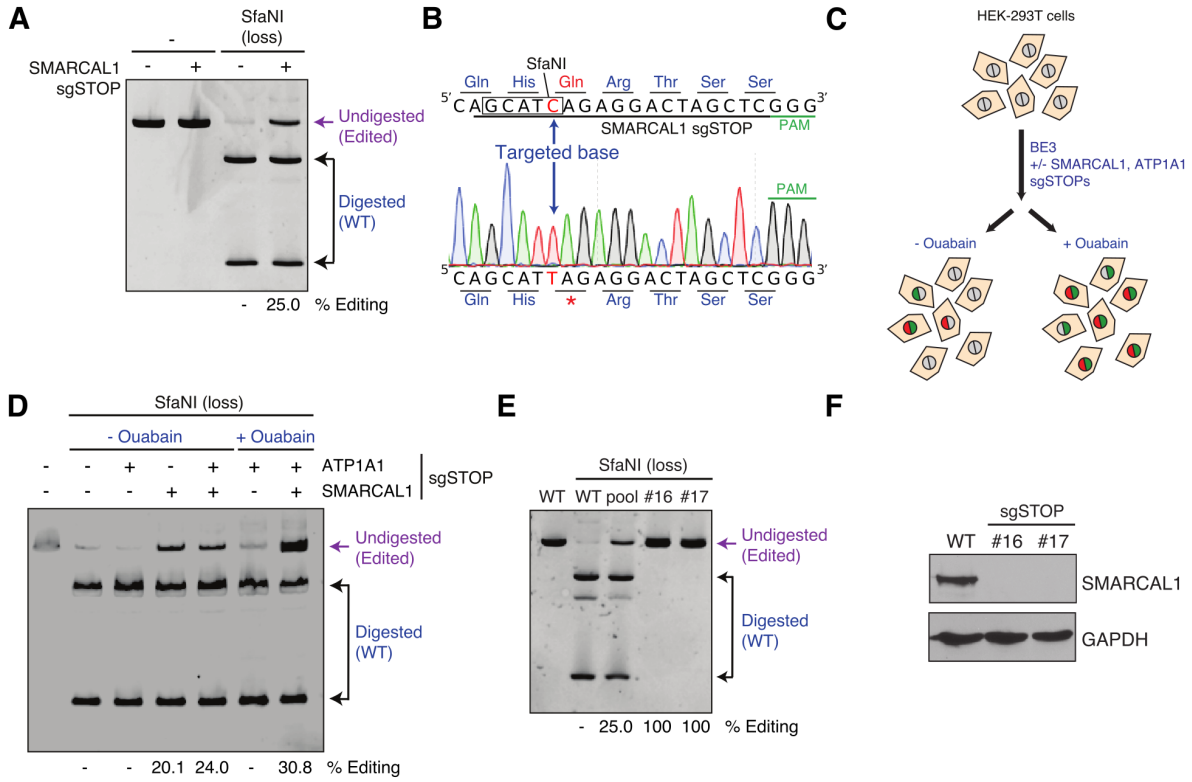

Figure 3.5 – **Generation of knockout human cell lines using iSTOP.** (**A**) Editing of the *SMARCAL1* locus by iSTOP. The *SMARCAL1* locus was PCR-amplified after transfection of BE3 with or without *SMARCAL1* sgSTOP. The *SMARCAL1* amplicon was then digested with SfaNI, which recognizes a restriction site containing the base targeted by iSTOP and the products of the restriction digest reaction were run on a polyacrylamide gel (left). Editing efficiency at the *SMARCAL1* locus was estimated by the percentage of *SMARCAL1* PCR amplicons refractory to SfaNI digestion (purple), as indicated in **Figure 3.3B**. One sequencing profile representative of four sequences of undigested *SMARCAL1* PCR products, including the targeted base, is indicated on the right inside. (**B**) Schematic representation of the *ATP1A1* co-selection strategy. HEK293T cells were transfected with BE3 with or without sgSTOPs targeting *SMARCAL1* and/ or *ATP1A1*. Cell populations were subsequently left untreated or treated with 1 mM ouabain, resulting in the enrichment of cells that had undergone genome editing at the *ATP1A1* and *SMARCAL1* loci. (**C**) Polyacrylamide gel showing the SfaNI digestion products of *SMARCAL1* PCR amplicons from HEK293T transfected with BE3 with or without sgSTOPs targeting *SMARCAL1* and/or *ATP1A1* and subjected to ouabain treatment, as represented in (B). Percentage of editing at the *SMARCAL1* locus with or without ouabain treatment was assessed as indicated in (A). (**D**) RFLP analysis of HEK293T clones edited by iSTOP in the *SMARCAL1* locus. Clones (#16 and #17) were isolated from cells transfected with BE3 and sgSTOPs targeting *SMARCAL1* and *ATP1A1* and subjected to ouabain selection, as described in (B). Amplicons of the *SMARCAL1* locus were digested by SfaNI and analyzed on polyacrylamide gel, as in (A). Restriction digest products of *SMARCAL1* amplicons from wild-type (WT) cells, iSTOP-edited cellular pool and clones #16 and #17 are indicated. See also **Figure 3.6E–G**. (**E**) Western blot analysis of SMARCAL1 protein levels in whole cellular extracts obtained from WT cells and clones #16 and #17, as shown in (D). GAPDH levels are used as loading controls. Experiments performed by Pierre Billon, Tarun Nambiar, and Samuel Hayward. Analysis performed by Pierre Billon. Related to **Section 3.2.3**.

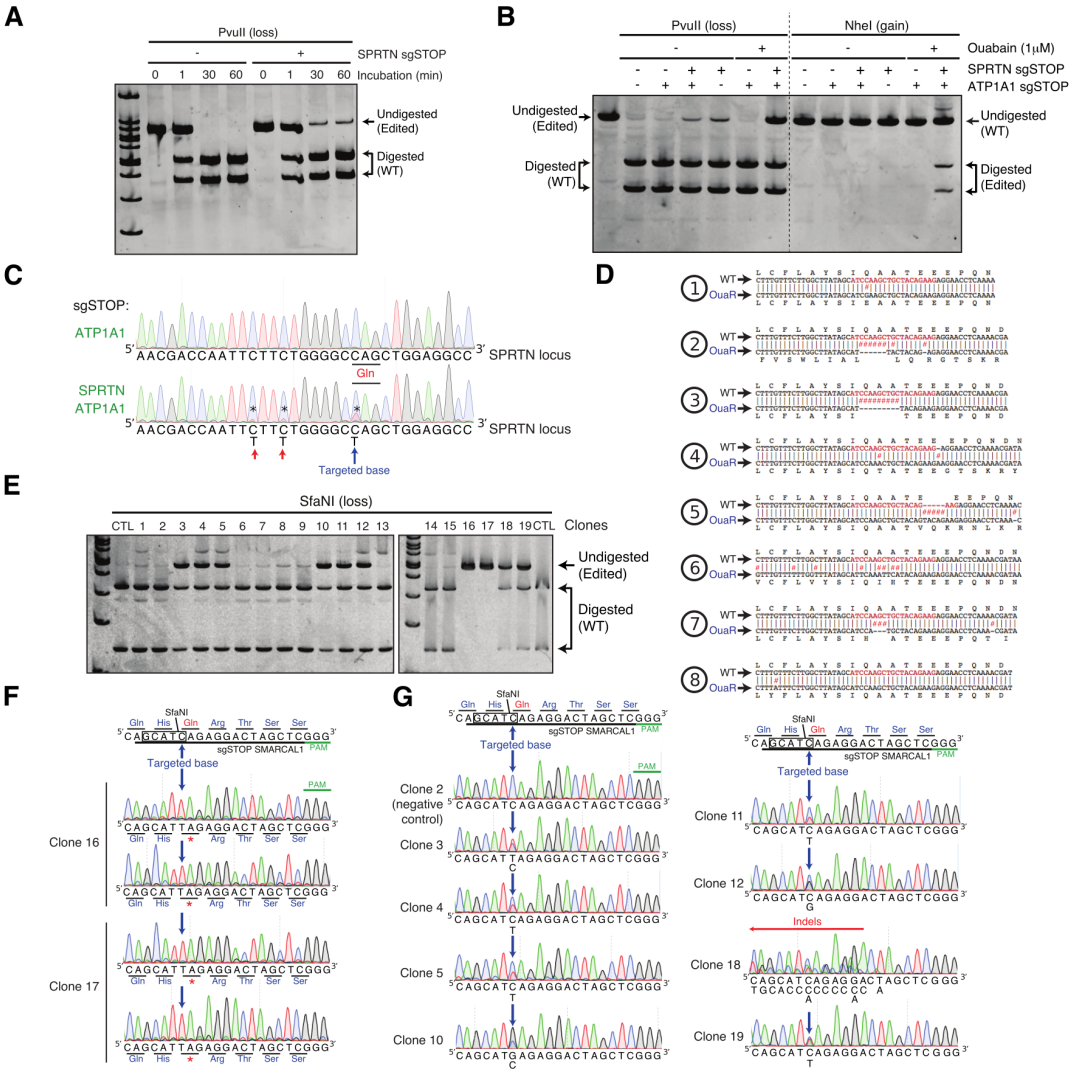

Figure 3.6 – Supplement to Figure 3.5 – Co-selection strategy to enrich for iSTOP- mediated editing and detection by RFLP assay. (**A**) PvuII restriction digest of PCR products of the *SPRTN* locus targeted with an sgSTOP that edits a PvuII restriction site. The reaction was terminated after 1, 30 or 60 min and the products of the reaction were run on a polyacrylamide gel. (**B**) Digestion of SPRTN amplicons from cells transfected with sgSTOPs targeting *SPRTN* and/or *ATP1A1* with or without 1 μM ouabain selection using the restriction enzymes PvuII and NheI. Editing efficiency was monitored by loss of PvuII cutting and gain of NheI cutting, as indicated in **Figure 3.3C**. (**C**) Sequencing profile of the targeted *SPRTN* locus amplified from cell populations transfected with sgSTOPs targeting *SPRTN* and/or *ATP1A1*. The targeted base that creates a STOP codon (blue arrow) and other targeted bases that generate missense mutations (red arrows) in the *SPRTN* locus are indicated by asterisks (*). (**D**) Alignment of sequences of *ATP1A1* alleles from cells targeted with an *ATP1A1* sgSTOP and selected with ouabain. The *ATP1A1* sgSTOP target sequence is represented in red. The symbol "#" indicates a mismatch between the WT genomic sequence (above) and the genomic sequence isolated from ouabain resistant cells (below). (**E**) RFLP analysis of the *SMARCAL1* locus from 19 single cell clones derived from HEK293T transfected with sgSTOPs targeting *SMARCAL1* and/or *ATP1A1* and selected with ouabain. *SMARCAL1* amplicons derived from the above clones were digested with SfaNI, as indicated in **Figure 3.5D**. (**F**) Sequencing profiles of *SMARCAL1* loci from the *SMARCAL1* KO clones #16 and #17 shown in (E). The targeted base is indicated by a blue arrow. (**G**) Sequencing profiles of the *SMARCAL1* locus in *SMARCAL1* WT and heterozygous mutant clones shown in (E). The targeted base is indicated by a blue arrow and indels are indicated by a red arrow. Experiments performed by Pierre Billon, Tarun Nambiar, and Samuel Hayward. Related to **Section 3.2.3**.

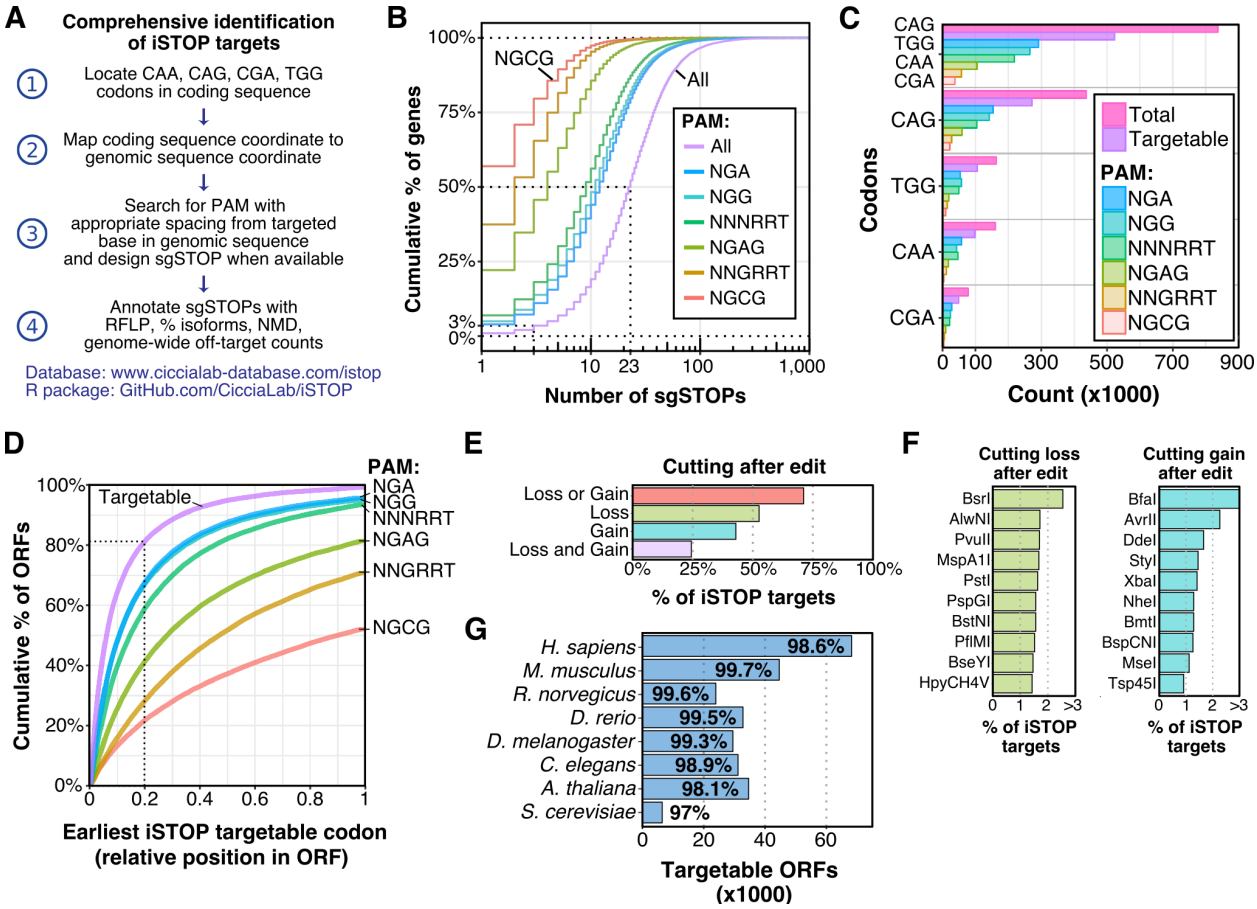

Figure 3.7 – Comprehensive detection of iSTOP targets in eukaryotic genomes. (**A**) Workflow utilized to identify iSTOP targetable sites in all ORFs with CDS coordinates available from the UCSC genome browser (https://genome.ucsc.edu/). Targetable sites were identified by first locating all CAA, CAG, CGA, and TGG codons in each coding sequences, then mapping the coordinates of the targeted base(s) in each codon to a genomic coordinate (steps 1–2). In total, 150 bases of genomic sequence flanking the targeted site were used to search for an appropriately spaced PAM (13–17 nucleotides from a targeted base) for all validated BE3 variants and for unique cutting of restriction enzymes (steps 3–4). Targeted isoforms and NMD predictions were determined as described in **Methods section 3.4.8**. (**B**) Cumulative distribution of the number of sgSTOPs designed per gene. The number of sgSTOPs for an average ORF (50%) in the human genome is indicated by a dotted line. Distributions for distinct PAM specificities (NGA, NGG, NNNRRT, NGAG, NNGRRT, and NGCG) are also shown. (**C**) Number of CAG, TGG, CAA, and CGA codons in the GRCh38 human reference genome targetable by iSTOP using BE3 variants with distinct PAM specificities (NGA, NGG, NNNRRT, NGAG, NNGRRT, and NGCG). (**D**) Relative position of the earliest iSTOP codon targetable in human ORFs (cumulative percentage) by BE3 variants with distinct PAM specificities (NGA, NGG, NNNRRT, NGAG, NNGRRT, and NGCG). The purple line takes into account all iSTOP targetable codons. (**E**) Percentage of human iSTOP sites verifiable by RFLP analysis using restriction enzymes that cut only once within a genomic region of ±50 bps flanking the targeted site. Bars indicate percentage of sites that can be verified by restriction enzyme cutting loss and/or gain. (**F**) Top ten restriction enzymes that can be utilized to validate iSTOP targets in the human genome by loss or gain of cutting within a genomic region of ±150 bps flanking the targeted site. The complete enzyme list is available in **Table 3.3**. (**G**) Number and percentage of iSTOP targetable ORFs in eight different eukaryotic species. See also **Figure 3.8C–E**. Related to **Section 3.2.4**.

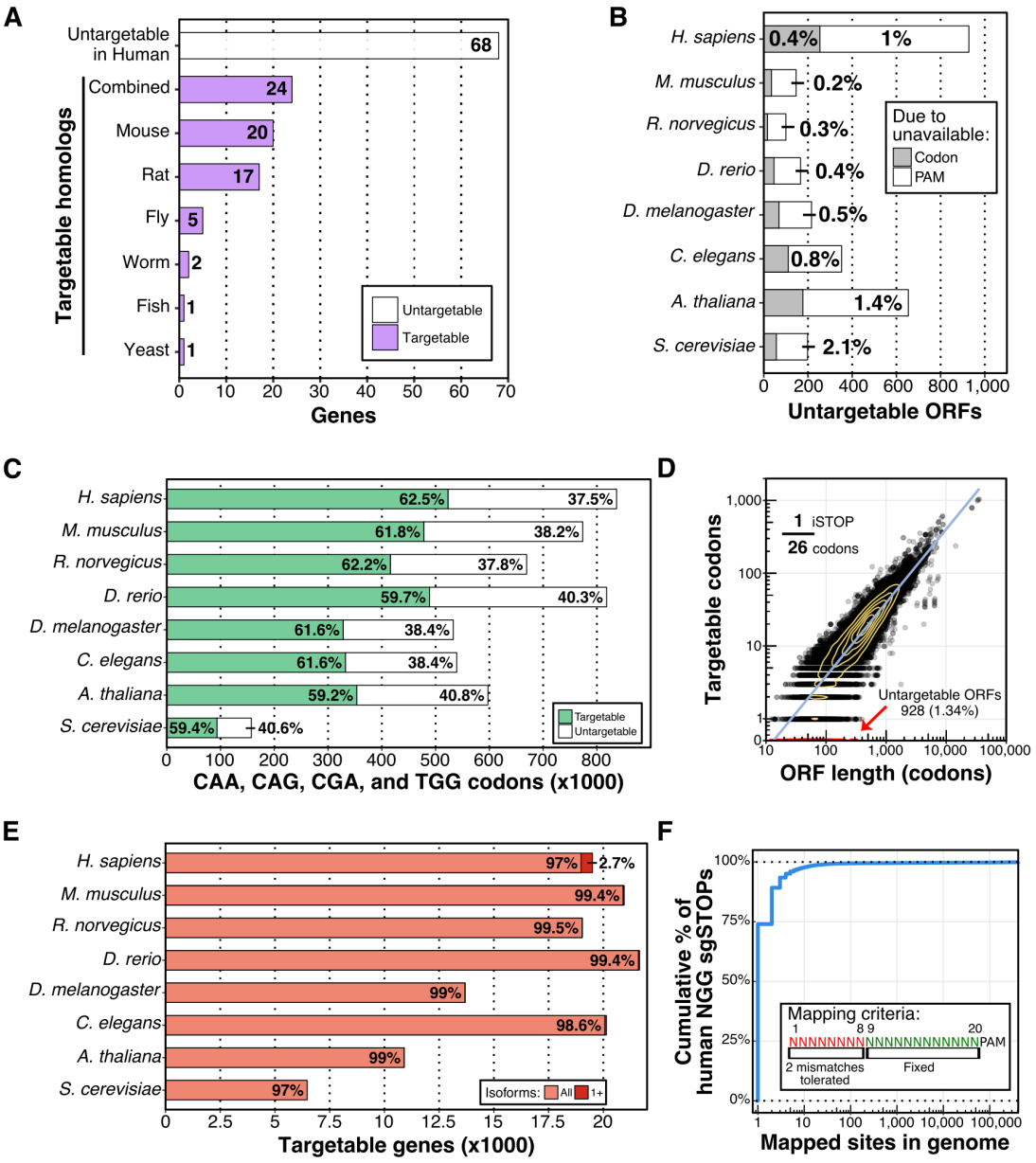

Figure 3.8 – Supplement to Figure 3.7 – Extended genomic analysis of iSTOP targetable sites in the genomes of 8 eukaryotic species. (**A**) Human genes untargetable by iSTOP that have targetable orthologs in other eukaryotic species. All homologs reported by Ensembl (www.ensemble.org) were considered. The complete list of untargetable human genes is available in **Table 3.2**. (**B**) Number of untargetable ORFs in all species considered in this study. Percentage of all ORFs that are untargetable due to an unavailable PAM are indicated in black text. (**C**) Number of all CAA, CAG, CGA and TGG codons in each species, and whether they are targetable with iSTOP. Percentage of each category is annotated on each bar. (**D**) Number of iSTOP targetable codons per ORF length in the human genome. (**E**) Number of genes that are targetable in all or at least one (1+) isoform. Percentage of total number of genes considered is annotated in text on each bar. (**F**) Distribution of the number of mapped sites in the human genome for NGG sgSTOPs. Each NGG sgSTOP was mapped to all matching locations in the genome allowing up to two mismatches outside of the guide's seed sequence (positions 1 through 8). Each guide is expected to map once in the genome. More than 1 mapped site indicates potential for off-target binding. Related to **Section 3.2.4**.

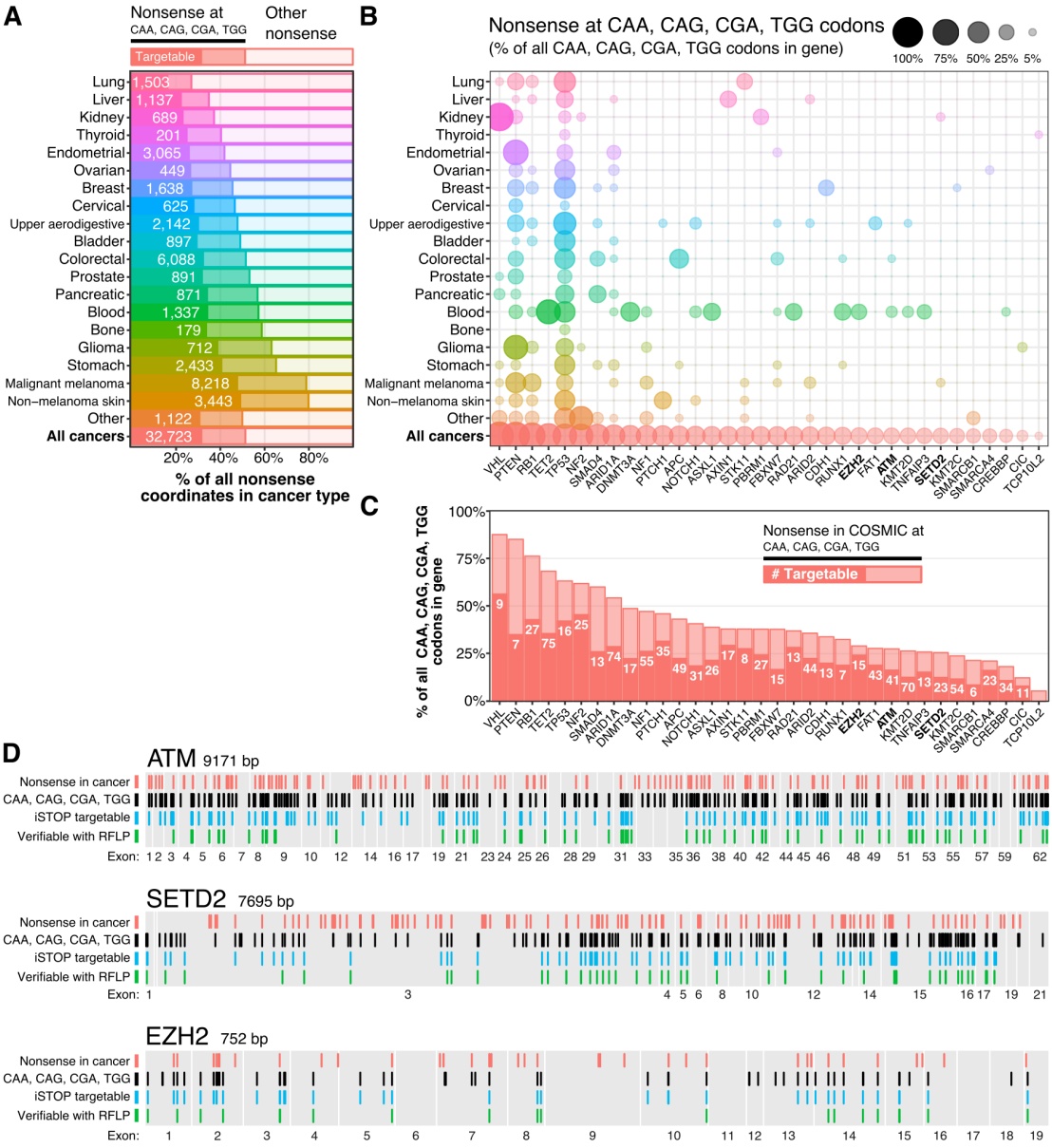

Figure 3.9 – **Modeling of cancer-associated nonsense mutations by iSTOP.** (**A**) Percentage of unique nonsense coordinates in cancer types, as observed in COSMIC. The percentage of base substitutions in CAA, CAG, CGA, or TGG codons that result in nonsense mutations in each cancer type is indicated. The total number of iSTOP targetable sites in each cancer type is annotated in white text. (**B**) Genes with frequently observed nonsense mutations at CAA, CAG, CGA, and TGG codons (iSTOPers) and their prevalence in different cancer types. The size and opacity of each circle represents the percentage of possible CAA, CAG, CGA, and TGG codons in the gene that were observed mutated to nonsense in each cancer type. See also **Figures 3.10A,B**. (**C**) Percentage of CAA, CAG, CGA, and TGG codons observed mutated to nonsense in the genes shown in (B) across all cancers. The total number of iSTOP targetable sites in each gene is annotated in white text. See also **Figure 3.10A,B**. (**D**) Maps of three iSTOPers (*ATM*, *SETD2*, and *EZH2*) indicating locations of (1) nonsense base substitutions in cancer (red tick marks), (2) CAA, CAG, CGA, and TGG codons (black tick marks), (3) iSTOP targetable codons (blue tick marks), and (4) iSTOP targetable codons that are verifiable via RFLP (green tick marks). The largest isoform for each gene is shown with exon numbers indicated below the gene. The size of the maps is not proportional to the length of the ORFs. Related to **Section 3.2.6**.

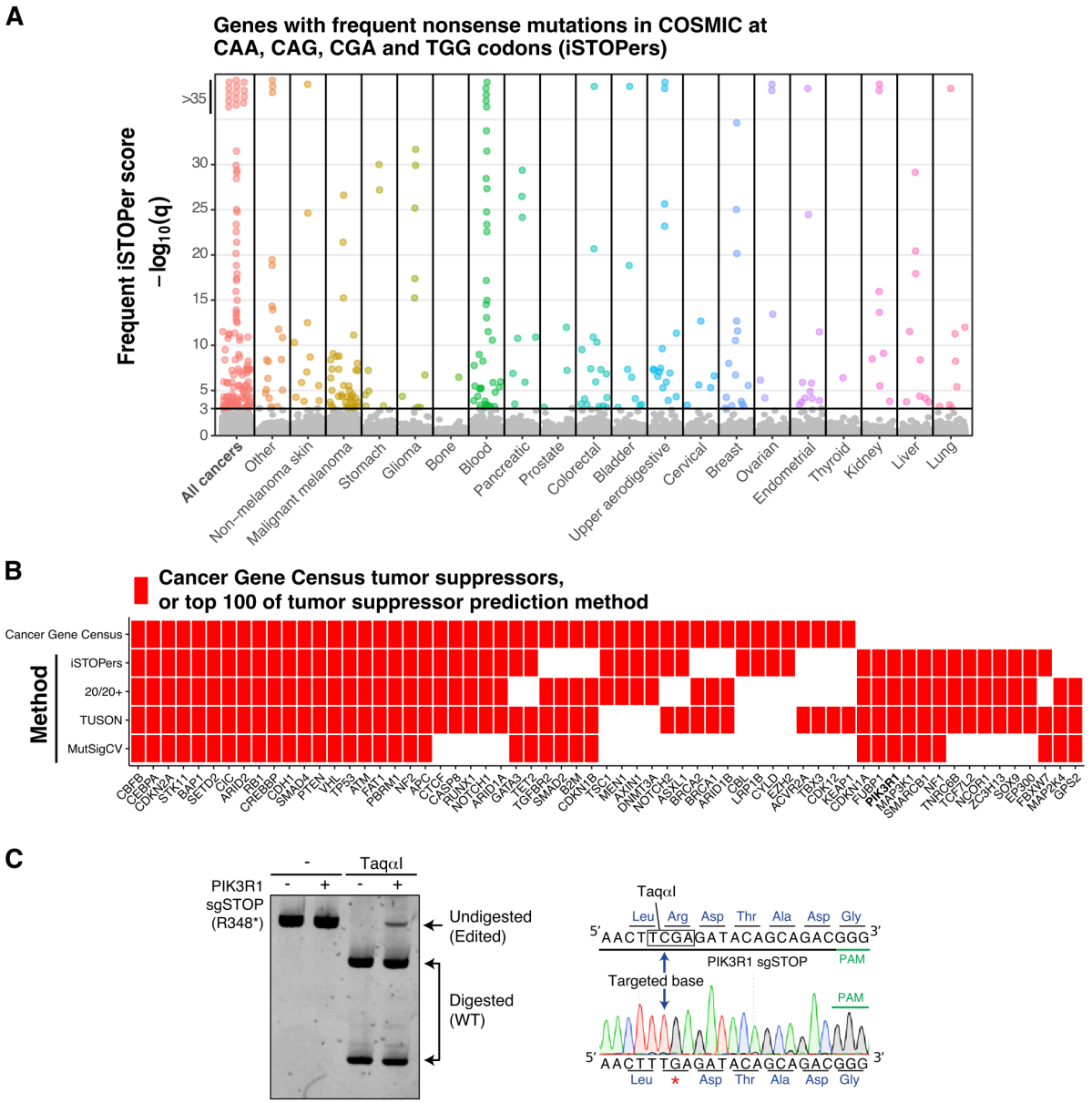

Figure 3.10 – Supplement to Figure 3.9 – Analysis of genes with frequent nonsense mutations at CAA, CAG, CGA and TGG codons in cancer and modeling of a recurrent cancer-associated nonsense mutation by iSTOP. (**A**) Genes with frequent nonsense mutations in COSMIC observed at CAA, CAG, CGA and TGG codons (iSTOPers). The frequent iSTOPer $-\log_{10}(q)$ score is derived from an FDR adjusted one-tailed binomial test (**Methods section 3.4.10**). "All cancers" is the highest score observed across all cancer types. (**B**) Comparison to tumor suppressor annotation and prediction methods. The top 100 iSTOPers from "All cancers" were compared to Tumor Suppressor Genes (TSGs) annotated by the Cancer Gene Census, the top 100 TSGs from the 20/20+ prediction method, the top 100 TSGs from the TUSON prediction method and the top 100 significantly mutated genes from the MutSigCV method. Genes included in this figure were either annotated as a TSG by the Cancer Gene Census, or were considered a TSG by at least 3 of the 4 TSG prediction methods. The complete list of iSTOPers ($q < 0.001$) is available in **Table 3.5**. (**C**) TaqαI-mediated digestion of PCR products of the *PIK3R1* locus targeted with an sgSTOP that edits a TaqαI restriction site to generate the cancer associated nonsense mutation R348*. One sequencing profile representative of 4 sequences of *PIK3R1* amplicons refractory to TaqαI digestion is shown on the right inside. Experiment and analysis presented in panel **C** was performed by Pierre Billon. Related to **Section 3.2.6**.

Table 3.1 – All possible coding sequence transitions using a C to T base editor

Header	Description
Coding Strand	
• Codon	Three letter DNA codon. Targetable base (C) is highlighted in red.
• Amino acid	Three letter amino acid encoded by codon. Three letter amino acid abbreviations – Ala = Alanine, Arg = Arginine, Asn = Asparagine, Asp = Aspartic acid, Cys = Cysteine, Gln = Glutamine, Glu = Glutamic acid, Gly = Glycine, His = Histidine, Ile = Isoleucine, Met = Methionine, Leu = Leucine, Lys = Lysine, Phe = Phenylalanine, Pro = Proline, Ser = Serine, Thr = Threonine, Trp = Tryptophan, Tyr = Tyrosine, Val = Valine and STOP = STOP codon.
• Codon generated	Comma separated codons generated by converting C to T with a C to T base editor targeting the coding strand.
• Amino acid generated	Three letter amino acid encoded by generated codons. Generated amino acids that differ from the reference are highlighted in red.
Non-Coding Strand	
• Codon	Three letter DNA codon. Targetable base (G) is highlighted in red.
• Amino acid	Three letter amino acid encoded by codon.
• Codon generated	Comma separated codons generated by converting G to A with a C to T base editor targeting the non-coding strand.
• Amino acid generated	Three letter amino acid encoded by generated codons. Generated amino acids that differ from the reference are highlighted in red.

Related to **Figure 3.1**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/iSTOP-01.xlsx.

Table 3.2 – Untargetable human genes and targetable homologs

Header	Description
`Human_gene`	Name of untargetable human gene (UCSC gene name).
`Mouse_gene`	Name of targetable *M. musculus* gene from Ensembl (UCSC ID) – http://www.ensembl.org. Multiple homologs within a species are separated by "|" marks. NA — Not Available.
`Rat_gene`	Name of targetable *R. norvegicus* gene from Ensembl (UCSC gene name)
`Fish_gene`	Name of targetable *D. rerio* gene from Ensembl (UCSC gene name).
`Worm_gene`	Name of targetable *C. elegans* gene from Ensembl (UCSC gene name).
`Fly_gene`	Name of targetable *D. melanogaster* gene from Ensembl (UCSC gene name).
`Yeast_gene`	Name of targetable *S. cerevisiae* gene from Ensembl (UCSC gene name).

Related to **Figure 3.8**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/iSTOP-02.xlsx.

Table 3.3 – Restriction enzymes considered for RFLP validation

Header	Description
`Enzyme`	Name of the restriction enzyme.
`Loss of cutting (N)`	Number of targetable sites in the human reference genome (GRCh38) that can be validated by loss of cutting with the corresponding enzyme (unique cutting +/- 150 bps).
`Loss of cutting (%)`	Percentage of all targetable sites that can be validated by loss of cutting with the corresponding enzyme.
`Gain of cutting (N)`	Number of sites that can be validated by gain of cutting after base editing (unique cutting +/- 150 bps).
`Gain of cutting (%)`	Percentage of all targetable sites that can be validated by gain of cutting with the corresponding enzyme.

Related to **Figure 3.7E,F**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/iSTOP-03.xlsx.

Table 3.4 – Frequent iSTOPers

Header	Description
Gene	Human gene name (UCSC gene). Genes observed to incur frequent cancer-associated nonsense mutations at CAA,CAG, CGA and TGG codons (i.e., iSTOPers).
Cancer_type	Type of cancer. An “All cancers” category is included as a summary of the lowest p-value observed for a gene across all cancer types.
N_nonsense_in_gene_in_cancer	Number of nonsense base substitutions in the given gene and cancer type.
N_iSTOP_sites_in_gene_in_cancer	Number of nonsense base substitutions observed in CAA, CAG, CGA and TGG codons in the given gene and cancer type (irrespective of PAM).
N_iSTOP_sites_in_gene_in_cancer_with_PAM	Number that are targetable with known PAM.
N_iSTOP_sites_in_gene	Number of iSTOP nonsense coordinates in gene (irrespective of PAM).
N_nonsense_in_cancer	Number of nonsense base substitutions observed in cancer type.
N_iSTOP_sites_in_cancer	Number of nonsense base substitutions observed in CAA, CAG, CGA and TGG codons in the given cancer type (irrespective of PAM).
N_iSTOP_sites_in_cancer_with_PAM	Number that are targetable with known PAM.
p-value	One-tailed binomial test p-value. See Methods section 3.4.10
Min_p-value	Smallest p-value across all cancer types for the given gene.
q-value	FDR adjusted p-value. Genes included in this table have a q-value < 0.001.
-LOG10(q)	Negative log transformed q-value.
Rank	q-value rank within the given cancer type.

Related to **Figure 3.10**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/iSTOP-04.xlsx.

Table 3.5 – Oligonucleotides

Header	Description
`Oligo name`	Name of oligonucleotide.
`Sequence (5 ->3 )`	Sequence of oligonucleotide in 5′ to 3′ orientation.

Related to **Chapter 3**.
Download – https://github.com/EricEdwardBryant/dissertation-data/raw/master/iSTOP-05.xlsx.

Table 3.6 – Key resources

Reagent or Resource	Source	Identifier
Antibodies		
• Mouse monoclonal α-SMARCAL1	Santa Cruz	sc-376377;AB_10987841
• Mouse monoclonal α-GAPDH	Novus	NB300-221;AB_10077627
• α-Mouse IgG, HRP-linked	GE Healthcare	NA931-1ML;AB_772210
Bacterial and Virus Strains		
• Subcloning efficiency DH5α	Life Technologies	18265-017
Reagents		
• Transfection reagent: TransIT-293	Mirus	MIR 2700
• Carbenicillin	Goldbio	C-103-25
• Ouabain octahydrate	Sigma	03125-1G
• Q5 High-Fidelity DNA polymerase	NEB	M0491L
Critical Commercial Assays		
• Quick Extract DNA Extraction solution	Epicenter	QE09050
• Zero BLUNT II TOPO PCR Cloning	Life Technologies	450245
• SuperSignal West Pico Chemiluminescent Substrate	Thermo Fisher	34087
Deposited Data		
• Unprocessed images of gels and blots	This paper; Mendeley Data	http://dx.doi.org/10.17632/jw5rjmypy2.1
• Annotated sgSTOPs for 8 organisms and 6 PAMs	This paper; Mendeley Data	http://www.ciccialab-database.com/istop; http://dx.doi.org/10.17632/xbdtvf6bvj.1
Cell Lines		
• Human: HEK293T	HTCC	CRL-11268
Oligonucleotides		
• Primers for PCR	This paper	**Table 3.5**
• sgRNA cloning	This paper	**Table 3.5**
Recombinant DNA		
• Plasmid: eSpCas9(1.1)	Addgene	71814
• Plasmid: MLM3636	Addgene	438860
• Plasmid: B52 (containing 2 empty sgRNAs-expressing cassettes)	This paper	Available in Addgene
• Plasmid: B270 (containing sgRNA targeting ATP1A1 + empty sgRNA-expressing cassette)	This paper	Available in Addgene
• Plasmid: pCMV-BE3	Addgene	73021
Software and Algorithms		
• iSTOP R package 0.1.0	This paper	https://github.com/CicciaLab/iSTOP/tree/0.1.0
• iSTOP reproducible analysis 0.1.1	This paper	https://github.com/CicciaLab/iSTOP-paper/tree/0.1.1
• R 3.4.1	The R project for statistical computing	https://www.r-project.org
• RStudio Desktop IDE 1.0.143	RStudio	https://www.rstudio.com
• Bioconductor 3.5	Bioconductor	http://bioconductor.org

Reagent or Resource	Source	Identifier
• Tidyverse 1.1.1	CRAN	https://cran.r-project.org/web/packages/tidyverse/index.html
• ImageJ	NIH	https://imagej.nih.gov/ij
• SnapGene Viewer	SnapGene	https://www.snapgene.com/products/snapgene_viewer
• Li-COR Odyssey	N/A	https://www.licor.com/bio/products/imaging_systems/odyssey
Other		
• Catalogue of Somatic Mutations in Cancer v80; Cancer gene census tumor suppressor annotations April 2017	Forbes et al. (2015)	http://cancer.sanger.ac.uk/cosmic
• Tumor suppressor predictions	Tokheim et al. (2016)	http://www.pnas.org/content/suppl/2016/11/22/1616440113.DCSupplemental/pnas.1616440113.sd04.xlsx

Related to **Chapter 3**.

Chapter 4

Discussion

Genome integrity is essential for organisms to faithfully pass genetic information to future generations. While genomic DNA is constantly subject to damage, organisms have evolved numerous mechanisms to maintain sequence information and allow replication of their genomes despite this damage. Defects in DNA damage tolerance pathways, either inherited from birth, or acquired during the life of an organism, can result in numerous diseases such as Fanconi anemia, xeroderma pigmentosum, ataxia telangiectasia, Nijmegen breakage, Cockayne, Werner & Bloom syndromes (Knoch et al. 2012). In addition to these diseases, defects in DNA damage tolerance often arise during the development of cancer. Perhaps the best known examples of this are mutations in the breast cancer susceptibility genes, *BRCA1* and *BRCA2*, which are both thought to be involved in homology directed repair (Roy et al. 2011; Zhang 2013). By leveraging knowledge of genetic interactions, new drugs have been developed to selectively kill cancers harboring mutations in *BRCA* genes by exploiting synthetic lethality between *PARP* and *BRCA* (Lord et al. 2015; Ashworth and Lord 2018). Long before it became clear that many cancers harbor mutations in DNA repair pathways, oncologists have treated patients with DNA damage through radiation and chemotherapy. As DNA sequencing costs have dropped, ever more tumors are being sequenced to identify common mutations and changes in gene expression, which has led to an increased appreciation that the status of DNA damage tolerance pathways in tumors can have a profound effect on a tumor's response to our most common cancer therapies (Alexandrov et al. 2013; Dietlein et al. 2014). As insight into the DNA damage tolerance network develops, strategies for selectively killing repair deficient tumors with drug combinations designed to target redundancies in the repair network are beginning to emerge (Cheng et al. 2013; Nickoloff et al. 2017). While some tumors respond well to radiation and DNA damaging chemotherapies, many tumors either do not respond, or acquire resistance to such treatments. This phenomenon underscores the importance of determining how the genetic landscape of a tumor impacts the overall damage tolerance network. The emerging field of systems genetics (Baliga et al. 2017), and specifically the study of genetic interaction networks, offers an important perspective for researchers and clinicians trying to understand how the DNA damage tolerance network enables tumor evolution while maintaining cancer cell viability. The structure of this network can enable precision medicine by providing critical insights into pathway organization that can

reveal tumor specific genetic weaknesses (Hartwell et al. 1997). Below, I will discuss some lessons and insights from my work on the systems genetics of DNA damage tolerance – a stepping stone on the long path to precision medicine in oncology.

4.1 Epistasis: undetected ≠ nonexistent

In **Chapter 1**, I presented progress toward constructing a genetic landscape for the cisplatin tolerance network in *Saccharomyces*. The design of this experiment advances the dE-MAP approach for detecting treatment specific pathway membership, and overall pathway relationships (Bandyopadhyay et al. 2010; Ideker and Krogan 2012). Specifically, we took advantage of the null model for genetic interactions of colony fitness to account for the fact that detection of genetic relationships between two alleles critically depends on the phenotype of each allele (**Figure 1.2**). While this fact has been understood by geneticists for decades, it has some unappreciated and important consequences for interpreting existing genetic interaction networks. Perhaps the most important consequence of requiring a phenotype to detect epistasis is that an undetected interaction is *not* equivalent to no interaction. In fact, conspicuous absence of an expected synergistic interaction may be weak evidence for epistasis in an as yet unknown treatment condition. Two epistatic alleles will *not* show any genetic interaction if neither allele has a phenotype in the tested condition. To detect the genetic relationship between these alleles necessitates identification of a condition that generates a phenotype for at least one allele. This limitation does not exist for synergistic alleles, which can account for the fact that the number of negative genetic interactions far exceed the number of positive genetic interactions in the genetic landscape of *S. cerevisiae* (Costanzo et al. 2016). For our dE-MAP, we treated mutants with optimized concentrations of cisplatin to enhance detection of both positive and negative genetic interactions. The current network supports the notion that nucleotide excision repair (NER) is the primary pathway responsible for *repairing* cisplatin lesions. However, the lesion *bypassing* postreplication repair (PRR) pathway appears to coordinate at least three critical nodes in the cisplatin tolerance network, and is thus the most essential pathway for maintaining fitness in the presence of cisplatin. While our network was able to reveal these insights, it is clear that, even with optimized treatment conditions, detecting epistasis remains challenging using the dE-MAP approach alone. An approach to overcome this limitation may be to use the conspicuous absence of genetic interactions between otherwise highly correlated interaction profiles to pick pairs of alleles for further fitness analysis using a quantitative spot assay and a full cisplatin titration.

4.2 TCGA, BioGRID & The Cell Map: insights from public data

Rad5, a member of PRR, was the focus of **Chapter 2**. During a visit with Dr. Karlene Cimprich, she mentioned that there was some evidence that the *RAD5* homolog, *HLTF*, was overexpressed in cancer (also noted in the discussion by Kile et al. (2015)). We therefore turned to publicly available sequencing data from The Cancer Genome Atlas (TCGA) using software that enabled programmatic access to gene-level summaries of this data (Cancer Genome Atlas Research Network 2008; Jacobsen and Questions 2018; Cerami et al. 2012; Gao et al. 2013). Indeed, the *RAD5* homolog, was frequently amplified and overexpressed in several classes of squamous cell carcinomas. This data-driven insight motivated the study of the genomic repercussions of *RAD5* overexpression. By screening the yeast gene disruption and temperature-sensitive mutant collections we identified a clear signature of genetic requirements for *RAD5* overexpression. This signature led us to the discovery that, when overexpressed, Rad5 can bypass canonical PRR regulatory signals to cause hyperactive recombination at replication forks. This hyperactive recombination is also accompanied by genome instability and cisplatin sensitivity.

What was especially fascinating about the *RAD5* story was that both deletion and overexpression caused cisplatin sensitivity, but only overexpression could drive recombination. This distinction was revealed by a careful comparison of the *RAD5* overexpression and deletion genetic interaction profiles. To compare interaction profiles, I used publicly available data from the BioGRID interaction database (Stark et al. 2006), and I developed a landscape enrichment analysis that uses easily accessible genetic interaction data from The Cell Map (Usaj et al. 2017; Costanzo et al. 2016; van der Maaten and Hinton 2008) (**Methods section 2.4.4**). This landscape enrichment analysis proved to be an intuitive way to visualize pathway enrichment for a genetic interaction profile, and allowed facile exploration of the genetic landscape to address questions such as, "what interactions are conspicuously absent?" In the case of *rad5*Δ, interactions with crossover resolution were clearly absent even though many of the mutant alleles in that region of the landscape had been identified as having negative fitness interactions with *rad5*Δ (**Figure 2.4**). Here, I would like to highlight that, in constructing our cisplatin dE-MAP, we also observed no interaction between *rad5*Δ and crossover resolution in the untreated network, but, when we treated with cisplatin, we observed epistasis – an example of conspicuous absence of synergy as weak evidence for epistasis (**Figures 1.4 & 2.4A**)!

4.3 Landscape enrichment: embracing perplexity

As mentioned above, the landscape enrichment analysis from **Figure 2.1C** enabled facile exploration of publicly available genetic interaction networks. This analysis is similar to Systematic Functional Annotation and Visualization of Biological Networks (SAFE), described by Baryshnikova (2016), but differs in its approach to dimension reduction (i.e. the two-dimensional clustering approach used to generate the "landscape"). The implementation presented in the SAFE publication uses a spring-embedded network layout, whereas the landscape presented for the *RAD5* chapter was generated using t-SNE (van der Maaten and Hinton 2008). The t-SNE method for dimension reduction is best summarized by Laurens van der Maaten himself (https://lvdmaaten.github.io/tsne), which can take an $A \times N$ matrix and reduce it to an $A \times 2$ matrix where the two dimensions attempt to preserve local similarities between the profiles of A. For generating the genetic landscape of *Saccharomyces cerevisiae*, rows in the matrix (A) are genes, and columns in the matrix (N) are genetic profile correlation values. Importantly, N can be any number of features that might provide information for generating clusters of genes or alleles. Thus, the landscape enrichment approach can be easily adapted to integrate other sources of data, and could be used to integrate existing biological information to generate a genetic landscape for human genes. t-SNE is lauded for its ability to reveal structure at many different scales. An example of this can be seen in **Figure 2.1C middle and right**, where zooming in on the dominant cluster, reveals additional local clustering of HDR, DNA replication, and crossover resolution. When viewing a t-SNE map it is important to be aware of the limitations of the approach, and avoid some common misinterpretations, nicely described by Wattenberg et al. (2016). Notably, relative distances between clusters on the landscape are largely meaningless. Clusters may appear on opposite sides of the landscape, but this does not mean they are more dissimilar than a distinct cluster present half-way across the landscape. t-SNE also has a parameter, called "perplexity", which loosely corresponds to the expected number of members in a major cluster. This parameter can have a major impact on clustering and must be lower than the number of rows in the matrix (i.e. the size of A, or the number of genes). Recommended values are between 5 and 50. The default value of perplexity for the implementation of t-SNE in the R programming language is 30 (Krijthe 2015), and, unaware of its meaning, my first attempt to run t-SNE on a matrix with only 29 rows resulted in the following message, "`Error: Perplexity is too large`." After a brief existential crisis, I lowered the perplexity, which resulted in beautiful clustering similar to that observed in **Figure 1.6**. Paraphrasing a comment from Dr. Eugene Koonin on the quality of clusters, "you know them when you see them."

4.4 Quantifying spots: replacing the drop assay

When working with high-throughput genetics, validating results is essential. This often involves strain verification, back-crossing, and low-throughput recapitulation of high-throughput observations using a gold-standard assay. For yeast colony growth, owing mostly to simplicity and familiarity, the serial dilution drop assay reigns supreme. The serial dilution drop assay certainly works well for detecting obvious fitness defects. However, when combining multiple alleles to confirm potentially modest genetic interactions, the qualitative nature of the drop assay quickly becomes a burden, leading to more eye-squinting and head-scratching than insight. For the *RAD5* manuscript we developed a quantitative colony growth assay for generating high-resolution colony fitness estimates (**Methods section 2.4.5 & Figure 2.3**). In my experience, this assay is able to resolve fitness differences as small as 5%. A good example of this resolution can be seen in **Figure 2.4B**. While the effect is small, a significant positive interaction was detected between *rad5*Δ and *sgs1*Δ in standard laboratory growth conditions – additional evidence for epistasis between *rad5*Δ and crossover resolution without requiring cisplatin treatment! The quantitative colony growth assay, while low in throughput, is essential for validating our genetic observations from high-throughput approaches like SGA and SPA. With the CRISPR revolution beginning to produce combinatorial reverse genetic screens, it will be greatly beneficial for the systems genetics community to develop high-resolution validation approaches that are not subject to some of the biases that can occur in these pooled screening approaches (Shen et al. 2017; Sack et al. 2016).

4.5 The CRISPR revolution: systems genetics in new model systems

Reverse genetic screens in mammalian cell lines have been possible through high-throughput RNA interference experiments for nearly twenty years (Carpenter and Sabatini 2004). Work in this field has progressed methods for performing pooled screening experiments to score fitness changes from a single transfected culture, wherein loss or gain of library representation is detected through next-generation sequencing (Sims et al. 2011). Initially, such experiments incurred high sequencing costs, and were plagued by off-target effects and many irreproducible results (Bhinder and Djaballah 2013). While these limitations have largely been overcome through reduced sequencing costs, formalized analysis pipelines, and improved experimental design (Schaefer et al. 2018), the arrival of CRISPR appears to have taken functional genomics in mammalian cells a major step forward (Evers et al. 2016). The favored approach for CRISPR based screens currently

appears to be using Cas9 to induce double-strand breaks that, when improperly repaired, can result in small insertions and deletions, potentially resulting in frameshift mutations that may cause nonsense and gene disruption. In **Chapter 3**, I present a new CRISPR method, called iSTOP, that allows direct generation of programmable nonsense without requiring a double-strand break and mutagenic repair. This feature may enable iSTOP to overcome the high levels of cell lethality caused by Cas9-induced double-strand breaks (Haapaniemi et al. 2018; Ihry et al. 2018). Importantly, for studying the DNA damage tolerance network in model systems beyond yeast, iSTOP may avoid confounding effects caused by requiring repair machinery to disrupt genes while simultaneously knocking out repair pathways. For this reason, I look forward to analyzing results from experiments comparing pooled screens using standard Cas9 and iSTOP.

Bibliography

Agudelo D, Duringer A, Bozoyan L, Huard CC, Carter S, Loehr J, Synodinou D, Drouin M, Salsman J, Dellaire G, et al. 2017. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nature Methods* **14**: 615–620. https://www.ncbi.nlm.nih.gov/pubmed/28417998.

Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, Cook A, Ha G, Harrington WF, Doshi MB, et al. 2016. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discovery* **6**: 914–29. https://www.ncbi.nlm.nih.gov/pubmed/27260156.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**: 415–21. https://www.ncbi.nlm.nih.gov/pubmed/23945592.

Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W. 2018. *rmarkdown: Dynamic documents for R*. https://CRAN.R-project.org/package=rmarkdown.

Andrews SJ, Rothnagel JA. 2014. Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics* **15**: 193–204. https://www.ncbi.nlm.nih.gov/pubmed/24514441.

Ashworth A, Lord CJ. 2018. Synthetic lethal therapies for cancer: What's next after PARP inhibitors? *Nature Reviews Clinical Oncology* **15**: 564–576. https://www.ncbi.nlm.nih.gov/pubmed/29955114.

Avery OT, Macleod CM, McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. *The Journal of Experimental Medicine* **79**: 137–58. https://www.ncbi.nlm.nih.gov/pubmed/19871359.

Baliga NS, Björkegren JLM, Boeke JD, Boutros M, Crawford NPS, Dudley AM, Farber CR, Jones A, Levey AI, Lusis AJ, et al. 2017. The state of systems genetics in 2017. *Cell Systems* **4**: 7–15. https://www.ncbi.nlm.nih.gov/pubmed/28125793.

Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M, et al. 2010. Rewiring of genetic networks in response to DNA damage. *Science (New York, NY)* **330**: 1385–9. https://www.ncbi.nlm.nih.gov/pubmed/21127252.

Bansbach CE, Bétous R, Lovejoy CA, Glick GG, Cortez D. 2009. The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes & Development* **23**: 2405–14. https://www.ncbi.nlm.nih.gov/pubmed/19793861.

Barrangou R, Doudna JA. 2016. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology* **34**: 933–941. https://www.ncbi.nlm.nih.gov/pubmed/27606440.

Baryshnikova A. 2016. Systematic functional annotation and visualization of biological networks. *Cell Systems* **2**: 412–21. https://www.ncbi.nlm.nih.gov/pubmed/27237738.

Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn J-Y, Ou J, Luis B-JS, Bandyopadhyay S, et al. 2010. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods* **7**: 1017–24. https://www.ncbi.nlm.nih.gov/pubmed/21076421.

Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C. 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **21**: 3329–30. https://www.ncbi.nlm.nih.gov/pubmed/8341614.

Baute J, Depicker A. 2008. Base excision repair and its role in maintaining genome stability. *Critical Reviews in Biochemistry and Molecular Biology* **43**: 239–276. http://www.tandfonline.com/doi/full/10.1080/

10409230802309905 (Accessed November 15, 2018).

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**: 289–300. http://www.jstor.org/stable/2346101.

Bhinder B, Djaballah H. 2013. Systematic analysis of RNAi reports identifies dismal commonality at gene-level and reveals an unprecedented enrichment in pooled shRNA screens. *Combinatorial Chemistry & High Throughput Screening* **16**: 665–81. https://www.ncbi.nlm.nih.gov/pubmed/23848309.

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017. The SWISS-MODEL repository-new features and functionality. *Nucleic Acids Research* **45**: D313–D319. https://www.ncbi.nlm.nih.gov/pubmed/27899672.

Billon P, Bryant EE, Joseph SA, Nambiar TS, Hayward SB, Rothstein R, Ciccia A. 2017. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Molecular Cell* **67**: 1068–1079.e4. https://www.ncbi.nlm.nih.gov/pubmed/28890334.

Birrell GW, Brown JA, Wu HI, Giaever G, Chu AM, Davis RW, Brown JM. 2002. Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 8778–83. https://www.ncbi.nlm.nih.gov/pubmed/12077312.

Blastyák A, Pintér L, Unk I, Prakash L, Prakash S, Haracska L. 2007. Yeast Rad5 protein required for postreplication repair has a DNA helicase activity specific for replication fork regression. *Molecular Cell* **28**: 167–75. https://www.ncbi.nlm.nih.gov/pubmed/17936713.

Blastyák A, ú IH, Unk I, Haracska L. 2010. Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. *Molecular and Cellular Biology* **30**: 684–93. https://www.ncbi.nlm.nih.gov/pubmed/19948885.

Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S. 2015. msa: An R package for multiple sequence alignment. *Bioinformatics* **31**: 3997–3999.

Boiteux S, Jinks-Robertson S. 2013. DNA repair mechanisms and the bypass of DNA damage in *Saccharomyces cerevisiae*. *Genetics* **193**: 1025–64. https://www.ncbi.nlm.nih.gov/pubmed/23547164.

Bordeira-Carriço R, Pêgo AP, Santos M, Oliveira C. 2012. Cancer syndromes and therapy by stop-codon readthrough. *Trends in Molecular Medicine* **18**: 667–78. https://www.ncbi.nlm.nih.gov/pubmed/23044248.

Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, Benschop JJ, Morris JH, Qiu C, Hu F, et al. 2013. From structure to systems: High-resolution, quantitative genetic analysis of rna polymerase ii. *Cell* **154**: 775–88. https://www.ncbi.nlm.nih.gov/pubmed/23932120.

Branzei D, Seki M, Enomoto T. 2004. Rad18/Rad5/Mms2-mediated polyubiquitination of PCNA is implicated in replication completion during replication stress. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* **9**: 1031–42. https://www.ncbi.nlm.nih.gov/pubmed/15507115.

Brendel M, Haynes RH. 1973. Interactions among genes controlling sensitivity to radiation and alkylation in yeast. *Molecular & General Genetics: MGG* **125**: 197–216. https://www.ncbi.nlm.nih.gov/pubmed/4359562.

Brewer BJ, Fangman WL. 1987. The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* **51**: 463–71. https://www.ncbi.nlm.nih.gov/pubmed/2822257.

Brown T. 2001. Southern blotting. *Current Protocols in Molecular Biology* **Chapter 2**: Unit2.9A. https://www.ncbi.nlm.nih.gov/pubmed/18265188.

Cancer Genome Atlas Research Network. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**: 1061–8. https://www.ncbi.nlm.nih.gov/pubmed/18772890.

Capouillez A, Decaestecker C, Filleul O, Chevalier D, Coppée F, Leroy X, Belayew A, Saussez S. 2008.

Helicase-like transcription factor exhibits increased expression and altered intracellular distribution during tumor progression in hypopharyngeal and laryngeal squamous cell carcinomas. *Virchows Archiv: an International Journal of Pathology* **453**: 491–9. https://www.ncbi.nlm.nih.gov/pubmed/18825407.

Carpenter AE, Sabatini DM. 2004. Systematic genome-wide screens of gene function. *Nature Reviews Genetics* **5**: 11–22. https://www.ncbi.nlm.nih.gov/pubmed/14708012.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. 2012. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* **2**: 401–4. https://www.ncbi.nlm.nih.gov/pubmed/22588877.

Chavez DA, Greer BH, Eichman BF. 2018. The HIRAN domain of helicase-like transcription factor positions the DNA translocase motor to drive efficient DNA fork regression. *The Journal of Biological Chemistry* **293**: 8484–8494. https://www.ncbi.nlm.nih.gov/pubmed/29643183.

Chen S, Davies AA, Sagan D, Ulrich HD. 2005. The RING finger ATPase Rad5p of *Saccharomyces cerevisiae* contributes to DNA double-strand break repair in a ubiquitin-independent manner. *Nucleic Acids Research* **33**: 5878–86. https://www.ncbi.nlm.nih.gov/pubmed/16224103.

Cheng H, Zhang Z, Borczuk A, Powell CA, Balajee AS, Lieberman HB, Halmos B. 2013. PARP inhibition selectively increases sensitivity to cisplatin in ERCC1-low non-small cell lung cancer cells. *Carcinogenesis* **34**: 739–49. https://www.ncbi.nlm.nih.gov/pubmed/23275151.

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim J-S. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* **24**: 132–41. https://www.ncbi.nlm.nih.gov/pubmed/24253446.

Choi PS, Meyerson M. 2014. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature Communications* **5**: 3728. https://www.ncbi.nlm.nih.gov/pubmed/24759083.

Ciccia A, Bredemeyer AL, Sowa ME, Terret M-E, Jallepalli PV, Harper JW, Elledge SJ. 2009. The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. *Genes & Development* **23**: 2415–25. https://www.ncbi.nlm.nih.gov/pubmed/19793862.

Ciccia A, Elledge SJ. 2010. The DNA damage response: Making it safe to play with knives. *Molecular Cell* **40**: 179–204. https://www.ncbi.nlm.nih.gov/pubmed/20965415.

Cleaver JE. 2016. Profile of Tomas Lindahl, Paul Modrich, and Aziz Sancar, 2015 Nobel laureates in chemistry. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 242–5. https://www.ncbi.nlm.nih.gov/pubmed/26715755.

Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, et al. 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**: 806–10. https://www.ncbi.nlm.nih.gov/pubmed/17314980.

Collins SR, Schuldiner M, Krogan NJ, Weissman JS. 2006. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol* **7**: R63.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–23. https://www.ncbi.nlm.nih.gov/pubmed/23287718.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, et al. 2010. The genetic landscape of a cell. *Science* **327**: 425–31. https://www.ncbi.nlm.nih.gov/pubmed/20093466.

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**. https://www.ncbi.nlm.nih.gov/pubmed/27708008.

Cox BS, Parry JM. 1968. The isolation, genetics and survival characteristics of ultraviolet light-sensitive

mutants in yeast. *Mutation Research* **6**: 37–55. https://www.ncbi.nlm.nih.gov/pubmed/5708072.

Crick F. 1970. Central dogma of molecular biology. *Nature* **227**: 561–3. https://www.ncbi.nlm.nih.gov/pubmed/4913914.

Crick FH, Barnett L, Brenner S, Watts-tobin RJ. 1961. General nature of the genetic code for proteins. *Nature* **192**: 1227–32. https://www.ncbi.nlm.nih.gov/pubmed/13882203.

Dahm R. 2005. Friedrich Miescher and the discovery of DNA. *Developmental Biology* **278**: 274–88. https://www.ncbi.nlm.nih.gov/pubmed/15680349.

Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, Elledge SJ. 2013. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**: 948–62. https://www.ncbi.nlm.nih.gov/pubmed/24183448.

Deans AJ, West SC. 2011. DNA interstrand crosslink repair and cancer. *Nature Reviews Cancer* **11**: 467–80. https://www.ncbi.nlm.nih.gov/pubmed/21701511.

Debauve G, Nonclercq D, Ribaucour F, Wiedig M, Gerbaux C, Leo O, Laurent G, é FJ, Belayew A, Toubeau G. 2006. Early expression of the helicase-like transcription factor (HLTF/SMARCA3) in an experimental model of estrogen-induced renal carcinogenesis. *Molecular Cancer* **5**: 23. https://www.ncbi.nlm.nih.gov/pubmed/16762066.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602–7. https://www.ncbi.nlm.nih.gov/pubmed/21455174.

Dietlein F, Thelen L, Reinhardt HC. 2014. Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends in Genetics* **30**: 326–39. https://www.ncbi.nlm.nih.gov/pubmed/25017190.

Dittmar JC, Pierce S, Rothstein R, Reid RJD. 2013. Physical and genetic-interaction density reveals functional organization and informs significance cutoffs in genome-wide screens. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 7389–94. https://www.ncbi.nlm.nih.gov/pubmed/23589890.

Eastman A. 1986. Reevaluation of interaction of cis-dichloro(ethylenediamine)platinum(II) with DNA. *Biochemistry* **25**: 3912–5. https://www.ncbi.nlm.nih.gov/pubmed/3741840.

Ellis B, Haaland P, Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, Jiang M. 2017. *flowCore: Basic structures for flow cytometry data*. http://bioconductor.org/packages/release/bioc/html/flowCore.html.

Elsässer SJ, Ernst RJ, Walker OS, Chin JW. 2016. Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nature Methods* **13**: 158–64. https://www.ncbi.nlm.nih.gov/pubmed/26727110.

Elserafy M, Abugable AA, Atteya R, El-Khamisy SF. 2018. Rad5, HLTF, and SHPRH: A fresh view of an old story. *Trends in Genetics* **34**: 574–577. https://www.ncbi.nlm.nih.gov/pubmed/29807746.

Evers B, Jastrzebski K, Heijmans JPM, Grernrum W, Beijersbergen RL, Bernards R. 2016. CRISPR knockout screening outperforms shRNA and crispri in identifying essential genes. *Nature Biotechnology* **34**: 631–3. https://www.ncbi.nlm.nih.gov/pubmed/27111720.

Fan HY, Cheng KK, Klein HL. 1996. Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant hpr1 delta of *Saccharomyces cerevisiae*. *Genetics* **142**: 749–59. https://www.ncbi.nlm.nih.gov/pubmed/8849885.

Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, Silva AC, Shales M, Collins SR, van Wageningen S, et al. 2009. Functional organization of the *S. cerevisiae* phosphorylation network. *Cell* **136**: 952–63. https://www.ncbi.nlm.nih.gov/pubmed/19269370.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L,

Mistry J, et al. 2014. Pfam: The protein families database. *Nucleic Acids Research* **42**: D222–30. https://www.ncbi.nlm.nih.gov/pubmed/24288371.

Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, et al. 2017. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Research* **45**: D777–D783. https://www.ncbi.nlm.nih.gov/pubmed/27899578.

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, et al. 2015. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research* **43**: D805–11. https://www.ncbi.nlm.nih.gov/pubmed/25355519.

Fortuna M, Sousa MJ, Côrte-Real M, Leão C, Salvador A, Sansonetty F. 2001. Cell cycle analysis of yeasts. *Current Protocols in Cytometry* **Chapter 11**: Unit 11.13. https://www.ncbi.nlm.nih.gov/pubmed/18770687.

Frock RL, Hu J, Meyers RM, Ho Y-J, Kii E, Alt FW. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnology* **33**: 179–86. https://www.ncbi.nlm.nih.gov/pubmed/25503383.

Galanski M, Jakupec MA, Keppler BK. 2005. Update of the preclinical situation of anticancer platinum complexes: Novel design strategies and innovative analytical approaches. *Current Medicinal Chemistry* **12**: 2075–94. https://www.ncbi.nlm.nih.gov/pubmed/16101495.

Game JC, Mortimer RK. 1974. A genetic study of x-ray sensitive mutants in yeast. *Mutation Research* **24**: 281–92. https://www.ncbi.nlm.nih.gov/pubmed/4606119.

Gangavarapu V, Haracska L, Unk I, Johnson RE, Prakash S, Prakash L. 2006. Mms2-Ubc13-dependent and -independent roles of Rad5 ubiquitin ligase in postreplication repair and translesion DNA synthesis in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **26**: 7783–90. https://www.ncbi.nlm.nih.gov/pubmed/16908531.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* **6**: pl1. https://www.ncbi.nlm.nih.gov/pubmed/23550210.

Garneau JE, Dupuis M-', Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67–71. https://www.ncbi.nlm.nih.gov/pubmed/21048762.

Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E2579–86. https://www.ncbi.nlm.nih.gov/pubmed/22949671.

Gentile F, Tuszynski JA, Barakat KH. 2016. New design of nucleotide excision repair (ner) inhibitors for combination cancer therapy. *Journal of Molecular Graphics & Modelling* **65**: 71–82. https://www.ncbi.nlm.nih.gov/pubmed/26939044.

Ghezraoui H, Piganeau M, Renouf B, Renaud J-B, Sallmyr A, Ruis B, Oh S, Tomkinson AE, Hendrickson EA, Giovannangeli C, et al. 2014. Chromosomal translocations in human cells are generated by canonical non-homologous end-joining. *Molecular Cell* **55**: 829–842. https://www.ncbi.nlm.nih.gov/pubmed/25201414.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–91. https://www.ncbi.nlm.nih.gov/pubmed/12140549.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al. 1996. Life with 6000 genes. *Science* **274**: 546, 563–567.

Goffeau A, Vassarotti A. 1991. The European project for sequencing the yeast genome. *Res Microbiol* **142**: 901–903. https://www.ncbi.nlm.nih.gov/pubmed/1784828 (Accessed October 18, 2018).

Goldstein AL, McCusker JH. 1999. Three new dominant drug resistance cassettes for gene disruption in

Saccharomyces cerevisiae. *Yeast* **15**: 1541–53. https://www.ncbi.nlm.nih.gov/pubmed/10514571.

Gong X, Kaushal S, Ceccarelli E, Bogdanova N, Neville C, Nguyen T, Clark H, Khatib ZA, Valentine M, Look AT, et al. 1997. Developmental regulation of Zbu1, a DNA-binding member of the SWI2/SNF2 family. *Developmental Biology* **183**: 166–82. https://www.ncbi.nlm.nih.gov/pubmed/9126292.

Grimm D. 2006. A mouse for every gene. *Science* **312**: 1862–1866. http://www.sciencemag.org/cgi/doi/10.1126/science.312.5782.1862 (Accessed October 19, 2018).

Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30 Suppl 1**: S162–73. https://www.ncbi.nlm.nih.gov/pubmed/19517507.

Guénolé A, Srivas R, Vreeken K, Wang ZZ, Wang S, Krogan NJ, Ideker T, van Attikum H. 2013. Dissection of DNA damage responses using multiconditional genetic interaction maps. *Molecular cell* **49**: 346–58. https://www.ncbi.nlm.nih.gov/pubmed/23273983.

Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. 2018. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine* **24**: 927–930. https://www.ncbi.nlm.nih.gov/pubmed/29892067.

Harrison MM, Jenkins BV, O'Connor-Giles KM, Wildonger J. 2014. A CRISPR view of development. *Genes & Development* **28**: 1859–72. https://www.ncbi.nlm.nih.gov/pubmed/25184674.

Hartwell LH, Szankasi P, Roberts CJ, Murray AW, Friend SH. 1997. Integrating genetic approaches into the discovery of anticancer drugs. *Science* **278**: 1064–8. https://www.ncbi.nlm.nih.gov/pubmed/9353181.

Hastie CJ, Vázquez-Martin C, Philp A, Stark MJR, Cohen PTW. 2006. The *Saccharomyces cerevisiae* orthologue of the human protein phosphatase 4 core regulatory subunit R2 confers resistance to the anticancer drug cisplatin. *The FEBS Journal* **273**: 3322–34. https://www.ncbi.nlm.nih.gov/pubmed/16857015.

Haynes R, Kunz B. 1981. DNA repair and mutagenesis in yeast. *Cold Spring Harbor Monograph Archive* **11A**. https://cshmonographs.org/index.php/monographs/article/view/4235/3411.

Herrero E, Thorpe PH. 2016. Synergistic control of kinetochore protein levels by Psh1 and Ubr2. *PLoS Genetics* **12**: e1005855. https://www.ncbi.nlm.nih.gov/pubmed/26891228.

Hess GT, Frésard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC. 2016. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature Methods* **13**: 1036–1042. https://www.ncbi.nlm.nih.gov/pubmed/27798611.

Heyer W-D, Ehmsen KT, Liu J. 2010. Regulation of homologous recombination in eukaryotes. *Annual Review of Genetics* **44**: 113–39. https://www.ncbi.nlm.nih.gov/pubmed/20690856.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, Onge RPS, Tyers M, Koller D, et al. 2008. The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* **320**: 362–5. https://www.ncbi.nlm.nih.gov/pubmed/18420932.

Hishiki A, Hara K, Ikegaya Y, Yokoyama H, Shimizu T, Sato M, Hashimoto H. 2015. Structure of a novel DNA-binding domain of helicase-like transcription factor (HLTF) and its functional implication in DNA damage tolerance. *The Journal of Biological Chemistry* **290**: 13215–23. https://www.ncbi.nlm.nih.gov/pubmed/25858588.

Hoege C, Pfander B, Moldovan G-L, Pyrowolakis G, Jentsch S. 2002. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**: 135–41. https://www.ncbi.nlm.nih.gov/pubmed/12226657.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262–78. https://www.ncbi.nlm.nih.gov/pubmed/24906146.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31**: 827–32.

https://www.ncbi.nlm.nih.gov/pubmed/23873081.

Huang M-E, Rio A-G, Galibert M-D, Galibert F. 2002. Pol32, a subunit of *Saccharomyces cerevisiae* DNA polymerase delta, suppresses genomic deletions and is involved in the mutagenic bypass pathway. *Genetics* **160**: 1409–22. https://www.ncbi.nlm.nih.gov/pubmed/11973297.

Huang R-Y, Eddy M, Vujcic M, Kowalski D. 2005. Genome-wide screen identifies genes whose inactivation confer resistance to cisplatin in *Saccharomyces cerevisiae*. *Cancer Research* **65**: 5890–7. https://www.ncbi.nlm.nih.gov/pubmed/15994967.

Hustedt N, Seeber A, Sack R, Tsai-Pflugfelder M, Bhullar B, Vlaming H, van Leeuwen F, Guénolé A, van Attikum H, Srivas R, et al. 2015. Yeast PP4 interacts with ATR homolog Ddc2-Mec1 and regulates checkpoint signaling. *Molecular cell* **57**: 273–89. https://www.ncbi.nlm.nih.gov/pubmed/25533186.

Ideker T, Krogan NJ. 2012. Differential network biology. *Molecular Systems Biology* **8**: 565. https://www.ncbi.nlm.nih.gov/pubmed/22252388.

Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, Kommineni S, Chen J, Sondey M, Ye C, et al. 2018. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nature Medicine* **24**: 939–946. https://www.ncbi.nlm.nih.gov/pubmed/29892062.

Iyer LM, Babu MM, Aravind L. 2006. The HIRAN domain and recruitment of chromatin remodeling and repair activities to damaged DNA. *Cell Cycle* **5**: 775–82. https://www.ncbi.nlm.nih.gov/pubmed/16627993.

Jacobsen A, Questions. 2018. *cgdsr: R-based API for accessing the MSKCC cancer genomics data server (CGDS)*. https://CRAN.R-project.org/package=cgdsr.

Jaiswal BS, Janakiraman V, Kljavin NM, Chaudhuri S, Stern HM, Wang W, Kan Z, Dbouk HA, Peters BA, Waring P, et al. 2009. Somatic mutations in p85alpha promote tumorigenesis through class IA PI3K activation. *Cancer Cell* **16**: 463–74. https://www.ncbi.nlm.nih.gov/pubmed/19962665.

Jasin M, Haber JE. 2016. The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair* **44**: 6–16. https://www.ncbi.nlm.nih.gov/pubmed/27261202.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–21. https://www.ncbi.nlm.nih.gov/pubmed/22745249.

Johnson RE, Henderson ST, Petes TD, Prakash S, Bankmann M, Prakash L. 1992. *Saccharomyces cerevisiae RAD5*-encoded DNA repair protein contains DNA helicase and zinc-binding sequence motifs and affects the stability of simple repetitive sequences in the genome. *Molecular and Cellular Biology* **12**: 3807–18. https://www.ncbi.nlm.nih.gov/pubmed/1324406.

Jones HD. 2015. Regulatory uncertainty over genome editing. *Nature Plants* **1**: 14011. https://www.ncbi.nlm.nih.gov/pubmed/27246057.

Kelland L. 2007. The resurgence of platinum-based cancer chemotherapy. *Nature Reviews Cancer* **7**: 573–84. https://www.ncbi.nlm.nih.gov/pubmed/17625587.

Kile AC, Chavez DA, Bacal J, Eldirany S, Korzhnev DM, Bezsonova I, Eichman BF, Cimprich KA. 2015. HLTF's ancient HIRAN domain binds 3' DNA ends to drive replication fork reversal. *Molecular Cell* **58**: 1090–100. https://www.ncbi.nlm.nih.gov/pubmed/26051180.

Kim D, Lim K, Kim S-T, Yoon S-H, Kim K, Ryu S-M, Kim J-S. 2017a. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nature Biotechnology* **35**: 475–480. https://www.ncbi.nlm.nih.gov/pubmed/28398345.

Kim K, Ryu S-M, Kim S-T, Baek G, Kim D, Lim K, Chung E, Kim S, Kim J-S. 2017b. Highly efficient RNA-guided base editing in mouse embryos. *Nature Biotechnology* **35**: 435–437. https://www.ncbi.nlm.nih.gov/pubmed/28244995.

Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. 2017c. Increasing the genome-targeting scope

and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology* **35**: 371–376. https://www.ncbi.nlm.nih.gov/pubmed/28191901.

Knoch J, Kamenisch Y, Kubisch C, Berneburg M. 2012. Rare hereditary diseases with defects in DNA-repair. *Eur J Dermatol* **22**: 443–455.

Kobayashi T, Horiuchi T. 1996. A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* **1**: 465–74. https://www.ncbi.nlm.nih.gov/pubmed/9078378.

Koller BH, Marrack P, Kappler JW, Smithies O. 1990. Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells. *Science* **248**: 1227–30. https://www.ncbi.nlm.nih.gov/pubmed/2112266.

Koller BH, Smithies O. 1989. Inactivating the beta 2-microglobulin locus in mouse embryonic stem cells by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 8932–5. https://www.ncbi.nlm.nih.gov/pubmed/2682666.

Komor AC, Badran AH, Liu DR. 2017. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**: 20–36. https://www.ncbi.nlm.nih.gov/pubmed/27866654.

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**: 420–4. https://www.ncbi.nlm.nih.gov/pubmed/27096365.

Krijthe JH. 2015. *Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation*. https://github.com/jkrijthe/Rtsne.

Kroll ES, Hyland KM, Hieter P, Li JJ. 1996. Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics* **143**: 95–102. https://www.ncbi.nlm.nih.gov/pubmed/8722765.

Kuang L, Kou H, Xie Z, Zhou Y, Feng X, Wang L, Wang Z. 2013. A non-catalytic function of Rev1 in translesion DNA synthesis and mutagenesis is mediated by its stable interaction with Rad5. *DNA Repair* **12**: 27–37. https://www.ncbi.nlm.nih.gov/pubmed/23142547.

Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. 2014. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods* **11**: 319–24. https://www.ncbi.nlm.nih.gov/pubmed/24487582.

Kuscu C, Parlak M, Tufan T, Yang J, Szlachta K, Wei X, Mammadov R, Adli M. 2017. CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nature Methods* **14**: 710–712. https://www.ncbi.nlm.nih.gov/pubmed/28581493.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921. https://www.ncbi.nlm.nih.gov/pubmed/11237011.

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**: 214–218. https://www.ncbi.nlm.nih.gov/pubmed/23770567.

Lee JA, Carvalho CMB, Lupski JR. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**: 1235–47. https://www.ncbi.nlm.nih.gov/pubmed/18160035.

Lee W, Onge RPS, Proctor M, Flaherty P, Jordan MI, Arkin AP, Davis RW, Nislow C, Giaever G. 2005. Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *PLoS Genetics* **1**: e24. https://www.ncbi.nlm.nih.gov/pubmed/16121259.

Lemontt JF. 1971. Mutants of yeast defective in mutation induced by ultraviolet light. *Genetics* **68**: 21–33. https://www.ncbi.nlm.nih.gov/pubmed/17248528.

Li F, Dong J, Eichmiller R, Holland C, Minca E, Prakash R, Sung P, Shim EY, Surtees JA, Lee SE. 2013. Role

of Saw1 in Rad1/Rad10 complex assembly at recombination intermediates in budding yeast. *The EMBO Journal* **32**: 461–72. https://www.ncbi.nlm.nih.gov/pubmed/23299942.

Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, Vizeacoumar FS, Min R, Vandersluis B, Bellay J, Devit M, et al. 2011. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nature Biotechnology* **29**: 361–7. https://www.ncbi.nlm.nih.gov/pubmed/21441928.

Liao C, Hu B, Arno MJ, Panaretou B. 2007. Genomic screening *in vivo* reveals the role played by vacuolar H+ ATPase and cytosolic acidification in sensitivity to DNA-damaging agents such as cisplatin. *Molecular Pharmacology* **71**: 416–25. https://www.ncbi.nlm.nih.gov/pubmed/17093137.

Liberi G, Maffioletti G, Lucca C, Chiolo I, Baryshnikova A, Cotta-Ramusino C, Lopes M, Pellicioli A, Haber JE, Foiani M. 2005. Rad51-dependent DNA structures accumulate at damaged replication forks in *sgs1* mutants defective in the yeast ortholog of BLM RecQ helicase. *Genes & Development* **19**: 339–50. https://www.ncbi.nlm.nih.gov/pubmed/15687257.

Lin J-R, Zeman MK, Chen J-Y, Yee M-C, Cimprich KA. 2011. SHPRH and HLTF act in a damage-specific manner to coordinate different forms of postreplication repair and prevent mutagenesis. *Molecular Cell* **42**: 237–49. https://www.ncbi.nlm.nih.gov/pubmed/21396873.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* **362**: 709–15. https://www.ncbi.nlm.nih.gov/pubmed/8469282.

Llorente B, Smith CE, Symington LS. 2008. Break-induced replication: What is it and what is it for? *Cell Cycle* **7**: 859–64. https://www.ncbi.nlm.nih.gov/pubmed/18414031.

Lord CJ, Tutt ANJ, Ashworth A. 2015. Synthetic lethality and cancer therapy: Lessons learned from the development of PARP inhibitors. *Annual Review of Medicine* **66**: 455–70. https://www.ncbi.nlm.nih.gov/pubmed/25341009.

Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nature reviews Molecular Cell Biology* **16**: 665–77. https://www.ncbi.nlm.nih.gov/pubmed/26397022.

Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X. 2016. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature Methods* **13**: 1029–1035. https://www.ncbi.nlm.nih.gov/pubmed/27723754.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823–6. https://www.ncbi.nlm.nih.gov/pubmed/23287722.

Marians KJ. 2018. Lesion bypass and the reactivation of stalled replication forks. *Annual Review of Biochemistry* **87**: 217–238. https://www.ncbi.nlm.nih.gov/pubmed/29298091.

Mayle R, Campbell IM, Beck CR, Yu Y, Wilson M, Shaw CA, Bjergbaek L, Lupski JR, Ira G. 2015. DNA REPAIR. Mus81 and converging forks limit the mutagenicity of replication fork breakage. *Science* **349**: 742–7. https://www.ncbi.nlm.nih.gov/pubmed/26273056.

Mirzoeva OK, Petrini JHJ. 2003. DNA replication-dependent nuclear dynamics of the Mre11 complex. *Molecular Cancer Research* **1**: 207–18. https://www.ncbi.nlm.nih.gov/pubmed/12556560.

Moinova HR, Chen W-D, Shen L, Smiraglia D, Olechnowicz J, Ravi L, Kasturi L, Myeroff L, Plass C, Parsons R, et al. 2002. HLTF gene silencing in human colon cancer. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 4562–7. https://www.ncbi.nlm.nih.gov/pubmed/11904375.

Morgan-Wall T. 2018. *rayshader: Create and visualize hillshaded maps from elevation matrices*. https://github.com/tylermorganwall/rayshader.

Motegi A, Liaw H-J, Lee K-Y, Roest HP, Maas A, Wu X, Moinova H, Markowitz SD, Ding H, Hoeijmakers JHJ, et al. 2008. Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic

instability from stalled replication forks. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 12411–6. https://www.ncbi.nlm.nih.gov/pubmed/18719106.

Neculai D, Walker JR, Weigelt J, Bountra C, Edwards AM, Arrowsmith CH, Dhe-Paganon S. 2015. Co-crystal structure of the HLTF HIRAN domain with a ssDNA fragment. http://www.rcsb.org/structure/5BNH.

Neelsen KJ, Chaudhuri AR, Follonier C, Herrador R, Lopes M. 2014. Visualization and interpretation of eukaryotic DNA replication intermediates *in vivo* by electron microscopy. *Methods in Molecular Biology* **1094**: 177–208. https://www.ncbi.nlm.nih.gov/pubmed/24162989.

Nickoloff JA, Jones D, Lee S-H, Williamson EA, Hromas R. 2017. Drugging the cancers addicted to DNA repair. *Journal of the National Cancer Institute* **109**. https://www.ncbi.nlm.nih.gov/pubmed/28521333.

Nirenberg M, Caskey T, Marshall R, Brimacombe R, Kellogg D, Doctor B, Hatfield D, Levin J, Rottman F, Pestka S, et al. 1966. The RNA code and protein synthesis. *Cold Spring Harb Symp Quant Biol* **31**: 11–24.

Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, et al. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**. https://www.ncbi.nlm.nih.gov/pubmed/27492474.

Noguchi E, Noguchi C, Du L-L, Russell P. 2003. Swi1 prevents replication fork collapse and controls checkpoint kinase Cds1. *Molecular and Cellular Biology* **23**: 7861–74. https://www.ncbi.nlm.nih.gov/pubmed/14560029.

O'Connor MJ. 2015. Targeting the DNA damage response in cancer. *Molecular Cell* **60**: 547–60. https://www.ncbi.nlm.nih.gov/pubmed/26590714.

Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. 2006. A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**: 1069–81. https://www.ncbi.nlm.nih.gov/pubmed/16487579.

Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. 2010. EBImage: An R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**: 979–981.

Pfister R, Janczyk M. 2013. Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology* **9**: 74–80. https://www.ncbi.nlm.nih.gov/pubmed/23826038.

Plosky BS. 2016. CRISPR-mediated base editing without DNA double-strand breaks. *Molecular Cell* **62**: 477–8. https://www.ncbi.nlm.nih.gov/pubmed/27203175.

Popp MW, Maquat LE. 2016. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* **165**: 1319–1322. https://www.ncbi.nlm.nih.gov/pubmed/27259145.

Postow L, Woo EM, Chait BT, Funabiki H. 2009. Identification of SMARCAL1 as a component of the DNA damage response. *The Journal of Biological Chemistry* **284**: 35951–61. https://www.ncbi.nlm.nih.gov/pubmed/19841479.

R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**: 2281–2308. https://www.ncbi.nlm.nih.gov/pubmed/24157548.

Rees HA, Komor AC, Yeh W-H, Caetano-Lopes J, Warman M, Edge ASB, Liu DR. 2017. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature Communications* **8**: 15790. https://www.ncbi.nlm.nih.gov/pubmed/28585549.

Reid RJD, Du X, Sunjevaric I, Rayannavar V, Dittmar J, Bryant E, Maurer M, Rothstein R. 2016. A synthetic dosage lethal genetic interaction between *CKS1B* and *PLK1* is conserved in yeast and human cancer cells. *Genetics* **204**: 807–819. https://www.ncbi.nlm.nih.gov/pubmed/27558135.

Reid RJD, González-Barrera S, Sunjevaric I, Alvaro D, Ciccone S, Wagner M, Rothstein R. 2011. Selective ploidy ablation, a high-throughput plasmid transfer protocol, identifies new genes affecting topoisomerase

I-induced DNA damage. *Genome Research* **21**: 477–86. https://www.ncbi.nlm.nih.gov/pubmed/21173034.

Roberts SA, Gordenin DA. 2014. Hypermutation in human cancer genomes: Footprints and mechanisms. *Nature Reviews Cancer* **14**: 786–800. https://www.ncbi.nlm.nih.gov/pubmed/25568919.

Rocha CRR, Silva MM, Quinet A, Cabral-Neto JB, Menck CFM. 2018. DNA repair pathways and cisplatin resistance: An intimate relationship. *Clinics* **73**: e478s. https://www.ncbi.nlm.nih.gov/pubmed/30208165.

Rosenberg B, Vancamp L, Krigas T. 1965. Inhibition of cell division in *Escherichia coli* by electrolysis products from a platinum electrode. *Nature* **205**: 698–9. https://www.ncbi.nlm.nih.gov/pubmed/14287410.

Rosenberg B, VanCamp L, Trosko JE, Mansour VH. 1969. Platinum compounds: A new class of potent antitumour agents. *Nature* **222**: 385–6. https://www.ncbi.nlm.nih.gov/pubmed/5782119.

Rothstein RJ. 1983. One-step gene disruption in yeast. *Methods in Enzymology* **101**: 202–211. https://www.ncbi.nlm.nih.gov/pubmed/6310324 (Accessed October 18, 2018).

Roukos V, Misteli T. 2014. The biogenesis of chromosome translocations. *Nature Cell Biology* **16**: 293–300. https://www.ncbi.nlm.nih.gov/pubmed/24691255.

Roy N, Runge KW. 2000. Two paralogs involved in transcriptional silencing that antagonistically control yeast life span. *Current Biology* **10**: 111–4. https://www.ncbi.nlm.nih.gov/pubmed/10662670.

Roy R, Chun J, Powell SN. 2011. BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nature Reviews Cancer* **12**: 68–78. https://www.ncbi.nlm.nih.gov/pubmed/22193408.

RStudio Team. 2016. *RStudio: Integrated development environment for R*. RStudio, Inc., Boston, MA http://www.rstudio.com/.

Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, Eliceiri KW. 2017. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**: 529. https://www.ncbi.nlm.nih.gov/pubmed/29187165.

Sack LM, Davoli T, Xu Q, Li MZ, Elledge SJ. 2016. Sources of error in mammalian genetic screens. *G3* **6**: 2781–90. https://www.ncbi.nlm.nih.gov/pubmed/27402361.

Saini N, Ramakrishnan S, Elango R, Ayyar S, Zhang Y, Deem A, Ira G, Haber JE, Lobachev KS, Malkova A. 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**: 389–92. https://www.ncbi.nlm.nih.gov/pubmed/24025772.

Sakofsky CJ, Malkova A. 2017. Break induced replication in eukaryotes: Mechanisms, functions, and consequences. *Critical Reviews in Biochemistry and Molecular Biology* **52**: 395–413. https://www.ncbi.nlm.nih.gov/pubmed/28427283.

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* **39**: 9275–82. https://www.ncbi.nlm.nih.gov/pubmed/21813460.

Schaefer C, Mallela N, Seggewiß J, Lechtape B, Omran H, Dirksen U, Korsching E, Potratz J. 2018. Target discovery screens using pooled shRNA libraries and next-generation sequencing: A model workflow and analytical algorithm. *PloS One* **13**: e0191570. https://www.ncbi.nlm.nih.gov/pubmed/29385199.

Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, et al. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**: 507–19. https://www.ncbi.nlm.nih.gov/pubmed/16269340.

Schuldiner M, Collins SR, Weissman JS, Krogan NJ. 2006. Quantitative genetic analysis in *Saccharomyces cerevisiae* using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. *Methods* **40**: 344–52. https://www.ncbi.nlm.nih.gov/pubmed/17101447.

Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, Licon K, Klepper K, Pekin D, Beckett AN, et al. 2017. Combinatorial CRISPR-Cas9 screens for *de novo* mapping of genetic interactions.

Nature Methods **14**: 573–576. https://www.ncbi.nlm.nih.gov/pubmed/28319113.

Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, Teramura H, Yamamoto T, Komatsu H, Miura K, et al. 2017. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nature Biotechnology* **35**: 441–443. https://www.ncbi.nlm.nih.gov/pubmed/28346401.

Shin S, Hyun K, Kim J, Hohng S. 2018. ATP binding to Rad5 initiates replication fork reversal by inducing the unwinding of the leading arm and the formation of the Holliday junction. *Cell Reports* **23**: 1831–1839. https://www.ncbi.nlm.nih.gov/pubmed/29742437.

Siddik ZH. 2003. Cisplatin: Mode of cytotoxic action and molecular basis of resistance. *Oncogene* **22**: 7265–79. https://www.ncbi.nlm.nih.gov/pubmed/14576837.

Sievers F, Higgins DG. 2014. Clustal omega, accurate alignment of very large numbers of sequences. *Methods in Molecular Biology* **1079**: 105–16. https://www.ncbi.nlm.nih.gov/pubmed/24170397.

Simon JA, Szankasi P, Nguyen DK, Ludlow C, Dunstan HM, Roberts CJ, Jensen EL, Hartwell LH, Friend SH. 2000. Differential toxicities of anticancer agents among DNA repair and checkpoint mutants of *Saccharomyces cerevisiae*. *Cancer Research* **60**: 328–33. https://www.ncbi.nlm.nih.gov/pubmed/10667584.

Sims D, Mendes-Pereira AM, Frankum J, Burgess D, Cerone M-A, Lombardelli C, Mitsopoulos C, Hakas J, Murugaesu N, Isacke CM, et al. 2011. High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. *Genome Biology* **12**: R104. https://www.ncbi.nlm.nih.gov/pubmed/22018332.

Snow R. 1967. Mutants of yeast sensitive to ultraviolet light. *Journal of Bacteriology* **94**: 571–5. https://www.ncbi.nlm.nih.gov/pubmed/6035260.

Srivas R, Costelloe T, Carvunis A-R, Sarkar S, Malta E, Sun SM, Pool M, Licon K, van Welsem T, van Leeuwen F, et al. 2013. A UV-induced genetic network links the RSC complex to nucleotide excision repair and shows dose-dependent rewiring. *Cell Reports* **5**: 1714–24. https://www.ncbi.nlm.nih.gov/pubmed/24360959.

St. Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, Nislow C, Roth FP, Giaever G. 2007. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature Genetics* **39**: 199–206. https://www.ncbi.nlm.nih.gov/pubmed/17206143.

Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: A general repository for interaction datasets. *Nucleic Acids Research* **34**: D535–9. https://www.ncbi.nlm.nih.gov/pubmed/16381927.

Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM. 1949. *The american soldier: Adjustment during army life (studies in social psychology in world war ii)*. Princeton University Press.

Sustic T, van Wageningen S, Bosdriesz E, Reid RJD, Dittmar J, Lieftink C, Beijersbergen RL, Wessels LFA, Rothstein R, Bernards R. 2018. A role for the unfolded protein response stress sensor ERN1 in regulating the response to MEK inhibitors in KRAS mutant colon cancers. *Genome Medicine* **in press**.

Szalat R, Samur MK, Fulciniti M, Lopez M, Nanjappa P, Cleynen A, Wen K, Kumar S, Perini T, Calkins AS, et al. 2018. Nucleotide excision repair is a potential therapeutic target in multiple myeloma. *Leukemia* **32**: 111–119. https://www.ncbi.nlm.nih.gov/pubmed/28588253.

Tenenbaum D, Team B. 2017. *BiocInstaller: Install/update bioconductor, CRAN, and github packages*.

Thomas BJ, Rothstein R. 1989. The genetic control of direct-repeat recombination in Saccharomyces: The effect of rad52 and rad1 on mitotic recombination at GAL10, a transcriptionally regulated gene. *Genetics* **123**: 725–38. https://www.ncbi.nlm.nih.gov/pubmed/2693208.

Tittel-Elmer M, Alabert C, Pasero P, Cobb JA. 2009. The MRX complex stabilizes the replisome independently of the S phase checkpoint during replication stress. *The EMBO Journal* **28**: 1142–56. https:

//www.ncbi.nlm.nih.gov/pubmed/19279665.

Todd RC, Lippard SJ. 2010. Structure of duplex DNA containing the cisplatin 1,2-Pt(NH3)22+-d(GpG) cross-link at 1.77 Å resolution. *Journal of Inorganic Biochemistry* **104**: 902–8. https://www.ncbi.nlm.nih.gov/pubmed/20541266.

Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. 2016. Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 14330–14335. https://www.ncbi.nlm.nih.gov/pubmed/27911828.

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, é NP, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**: 2364–8. https://www.ncbi.nlm.nih.gov/pubmed/11743205.

Tong AHY, Boone C. 2006. Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol Biol* **313**: 171–192. https://www.ncbi.nlm.nih.gov/pubmed/16118434.

Torres R, Martin MC, Garcia A, Cigudosa JC, Ramirez JC, Rodriguez-Perales S. 2014. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nature Communications* **5**: 3964. https://www.ncbi.nlm.nih.gov/pubmed/24888982.

Ulrich HD. 2003. Protein-protein interactions within an E2-RING finger complex. implications for ubiquitin-dependent DNA damage repair. *The Journal of Biological Chemistry* **278**: 7051–8. https://www.ncbi.nlm.nih.gov/pubmed/12496280.

Ulrich HD, Jentsch S. 2000. Two RING finger proteins mediate cooperation between ubiquitin-conjugating enzymes in DNA repair. *The EMBO Journal* **19**: 3388–97. https://www.ncbi.nlm.nih.gov/pubmed/10880451.

Unk I, ú IH, Fátyol K, Hurwitz J, Yoon J-H, Prakash L, Prakash S, Haracska L. 2008. Human HLTF functions as a ubiquitin ligase for proliferating cell nuclear antigen polyubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 3768–73. https://www.ncbi.nlm.nih.gov/pubmed/18316726.

Usaj M, Tan Y, Wang W, VanderSluis B, Zou A, Myers CL, Costanzo M, Andrews B, Boone C. 2017. TheCellMap.org: A web-accessible database for visualizing and mining the global yeast genetic interaction network. *G3* **7**: 1539–1549. https://www.ncbi.nlm.nih.gov/pubmed/28325812.

van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**: 2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html (Accessed July 29, 2018).

van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B, et al. 2016. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Molecular Cell* **63**: 633–646. https://www.ncbi.nlm.nih.gov/pubmed/27499295.

Vogel G. 2007. Nobel prizes. a knockout award in medicine. *Science* **318**: 178–9. https://www.ncbi.nlm.nih.gov/pubmed/17932258.

Wach A, Brachat A, Pöhlmann R, Philippsen P. 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**: 1793–808. https://www.ncbi.nlm.nih.gov/pubmed/7747518.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**: W296–W303. https://www.ncbi.nlm.nih.gov/pubmed/29788355.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–62. https://www.ncbi.nlm.nih.gov/pubmed/12466850.

Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**: 737–8. https://www.ncbi.nlm.nih.gov/pubmed/13054692.

Wattenberg M, Viégas F, Johnson I. 2016. How to use t-sne effectively. *Distill*. http://distill.pub/2016/

misread-tsne.

Wickham H. 2017. *tidyverse: Easily install and load the 'tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Williamson L, Saponaro M, Boeing S, East P, Mitter R, Kantidakis T, Kelly GP, Lobley A, Walker J, Spencer-Dene B, et al. 2017. UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene. *Cell* **168**: 843–855.e13. https://www.ncbi.nlm.nih.gov/pubmed/28215706.

Wilson MA, Kwon Y, Xu Y, Chung W-H, Chi P, Niu H, Mayle R, Chen X, Malkova A, Sung P, et al. 2013. Pif1 helicase and polδ promote recombination-coupled DNA synthesis via bubble migration. *Nature* **502**: 393–6. https://www.ncbi.nlm.nih.gov/pubmed/24025768.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**: 901–6. https://www.ncbi.nlm.nih.gov/pubmed/10436161.

Wong JHY, Brown JA, Suo Z, Blum P, Nohmi T, Ling H. 2010. Structural insight into dynamic bypass of the major cisplatin-DNA adduct by Y-family polymerase Dpo4. *The EMBO Journal* **29**: 2059–69. https://www.ncbi.nlm.nih.gov/pubmed/20512114.

Wu HI, Brown JA, Dorie MJ, Lazzeroni L, Brown JM. 2004. Genome-wide identification of genes conferring resistance to the anticancer agents cisplatin, oxaliplatin, and mitomycin C. *Cancer Research* **64**: 3940–8. https://www.ncbi.nlm.nih.gov/pubmed/15173006.

Xiao W, Chow BL, Broomfield S, Hanna M. 2000. The *Saccharomyces cerevisiae RAD6* group is composed of an error-prone and two error-free postreplication repair pathways. *Genetics* **155**: 1633–41. https://www.ncbi.nlm.nih.gov/pubmed/10924462.

Xie Y. 2015. *Dynamic documents with R and knitr*. 2nd ed. Chapman; Hall/CRC, Boca Raton, Florida https://yihui.name/knitr/.

Xie Y, Kerscher O, Kroetz MB, McConchie HF, Sung P, Hochstrasser M. 2007. The yeast Hex3.Slx8 heterodimer is a ubiquitin ligase stimulated by substrate sumoylation. *The Journal of Biological Chemistry* **282**: 34176–84. https://www.ncbi.nlm.nih.gov/pubmed/17848550.

Yang L, Briggs AW, Chew WL, Mali P, Guell M, Aach J, Goodman DB, Cox D, Kan Y, Lesha E, et al. 2016. Engineering and optimising deaminase fusions for genome editing. *Nature Communications* **7**: 13330. https://www.ncbi.nlm.nih.gov/pubmed/27804970.

Yuan J, Ghosal G, Chen J. 2009. The annealing helicase HARP protects stalled replication forks. *Genes & Development* **23**: 2394–9. https://www.ncbi.nlm.nih.gov/pubmed/19793864.

Yusufzai T, Kong X, Yokomori K, Kadonaga JT. 2009. The annealing helicase HARP is recruited to DNA repair sites via an interaction with RPA. *Genes & Development* **23**: 2400–4. https://www.ncbi.nlm.nih.gov/pubmed/19793863.

Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nature Genetics* **41**: 849–53. https://www.ncbi.nlm.nih.gov/pubmed/19543269.

Zhang J. 2013. The role of BRCA1 in homologous recombination repair in response to replication stress: Significance in tumorigenesis and cancer therapy. *Cell & Bioscience* **3**: 11. https://www.ncbi.nlm.nih.gov/pubmed/23388117.

Zhou J, Monson EK, Teng SC, Schulz VP, Zakian VA. 2000. Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science* **289**: 771–4. https://www.ncbi.nlm.nih.gov/pubmed/10926538.

Zijlstra M, Bix M, Simister NE, Loring JM, Raulet DH, Jaenisch R. 1990. Beta 2-microglobulin deficient mice lack CD4-8+ cytolytic T cells. *Nature* **344**: 742–6. https://www.ncbi.nlm.nih.gov/pubmed/2139497.

Zijlstra M, Li E, Sajjadi F, Subramani S, Jaenisch R. 1989. Germ-line transmission of a disrupted beta 2-microglobulin gene produced by homologous recombination in embryonic stem cells. *Nature* **342**: 435–8.

https://www.ncbi.nlm.nih.gov/pubmed/2685607.

Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, Qiu J-L, Wang D, Gao C. 2017. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology* **35**: 438–440. https://www.ncbi.nlm.nih.gov/pubmed/28244994.