SYNTHETIC HYPERMUTATION: GENE-DRUG MUTATION RATE SYNERGY REVEALS A TRANSLESION SYNTHESIS MECHANISM

by

ROMULO M. SEGOVIA UGARTE

B.Sc., San Antonio Abad del Cusco, 2003
M.Sc., Iowa State University, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2017

Abstract

Gene-gene or gene-drug interactions are typically quantified using fitness as a readout because the data is continuous and easily measured in high-throughput. However, to what extent fitness captures the range of other phenotypes that show synergistic effects is usually unknown. Using *Saccharomyces cerevisiae*, and focusing on a matrix of DNA repair mutants and genotoxic drugs, I quantified 76 gene-drug interactions based on both mutation rate and fitness and find that these parameters are not connected. Independent of fitness defects I identified seven cases of synthetic hypermutation, where the combined effect of the drug and mutant on mutation rate was greater than predicted. One example occurred when yeast lacking *RAD1* were exposed to cisplatin and I characterized this interaction using whole-genome sequencing. Our sequencing results indicate mutagenesis by cisplatin in *rad1*Δ cells depended almost entirely on interstrand crosslinks at GpCpN motifs. Interestingly, our data suggest that the 3' base in this motif templates the addition of the mutated base. This result differs from cisplatin mutation signatures in XPF-deficient *C. elegans* and supports a model in which translesion synthesis polymerases perform a slippage and realignment extension across from the damaged base. Accordingly, DNA polymerase zeta activity was essential for mutagenesis in cisplatin-treated *rad1*Δ cells. Together these data reveal the potential to gain new mechanistic insights from non-fitness measures of gene-drug interactions and extend the use of mutation accumulation and whole-genome sequencing analysis to define DNA repair mechanisms.

Preface

Two versions of this work have been published: [Segovia R, Shen Y, Lujan SA, Jones S, Stirling PC. Synthetic hypermutation: gene-drug mutation rate synergy reveals a translesion synthesis mechanism. bioRxiv 086512; doi: http://dx.doi.org/10.1101/086512. 2016] and [Segovia R, Shen Y, Lujan SA, Jones S, Stirling PC. Hypermutation signature reveals a slippage and realignment model of translesion synthesis by Rev3 polymerase in cisplatin-treated yeast. PNAS USA. 114(10):2663-2668. 2017]. This thesis is a work by Romulo Segovia Ugarte. The project was conceived by Dr. Peter C. Stirling and I. The experimental work and analysis of data were carried out by Romulo Segovia Ugarte, except for the variant calling analysis that was performed by Dr. Yaoqing Shen from Dr. Steven Jones lab, and the analysis shown in figure 8C by Dr. Peter C. Stirling.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AML: Myeloid leukemia

BER: Base-excision repair

BIR: Break-induced replication

BS: Bloom's syndrome

Cis: Cisplatin

CNV: Copy number variant

CPD: Cyclobutane pyrimidine dimer

Cpt: Camptothecin

CS: Cockayne syndrome

DSB: DNA double-strand break

dHJ: Double-Holliday Junction

Etp: Etoposide

FA: Fanconi anemia

G1000: 1000 generation

GG-NER: Global-genome nucleotide excision repair

HNPCC: Hereditary non-polyposis colorectal cancer

HNSCC: Head and neck squamous cell carcinoma

HR: Homologous recombination

ICLs: Interstrand-crosslinks

MA: Mutation accumulation

MMC: Mitomycin C

MMEJ: Microhomology-mediated end joining

MMR: Mismatch repair

MMS: Methyl methanesulfonate

MSI: Microsatellite instability

NER: Nucleotide-excision repair

NHEJ: Non-homologous end joining

PPR: Post-replication repair

ROS: Reactive oxidative species

SD: Synthetic defined yeast medium

SD+5: SD minimal plus histidine, uracil, leucine, lysine and methionine + 2% dextrose

SHyp: Synthetic hypermutation

SNV: Single nucleotide variant

SSA: Single-strand annealing

SSB: DNA single-strand break

ssDNA: Single-stranded DNA

TC-NER: Transcription-coupled nucleotide excision repair

TLS: Translesion synthesis

TTD: Trichothiodystrophy

WGS: whole-genome sequencing

WS: Werner's syndrome

YPD: Yeast peptone-dextrose medium

XP: Xeroderma pigmentosum

5-FU: 5-fluorouracil

Acknowledgements

Foremost, I would like to express my gratitude to my advisor Dr. Peter Stirling for his continuous support throughout my M.Sc. study and research. His patience, advice and knowledge guided me towards the culmination of this thesis and achievement of scientific publication.

My sincere thanks also go to the members of my thesis committee Dr. Phil Hieter and Dr. Steven Jones for their commitment, insightful comments of my research work and collaborative support. A very special thanks to Dr. Yaoqing Shen for the bioinformatics support provided.

I also thanks to my fellow lab mates Dr. Veena Mathew, Annie Tam, Carolina Novoa, Karissa Milbury and Dr. Emily Chang for their help, advice and all the hard working hours and fun that we shared.

My gratitude and admiration also go to my parents who are always encouraging me and following my steps despite of the distance. My achievements are theirs too.

Finally, thanks to my wife Pauline for the patience, sacrifice and invaluable support that she gave me during these years of learning.

Dedication

To my parents

And my wife Pauline

Chapter 1: INTRODUCTION

1.1 Overview and Summary

Genome maintenance pathways suppress the accumulation of mutations derived from chemical lesions or mismatches in DNA that arise during normal metabolic processes. Despite thousands of potentially mutagenic lesions occurring per cell per day, mitotic cell division exhibits extremely low rates of mutation (10^{-8} – 10^{-10} per bp per generation) under normal conditions in a variety of species (1). Cells with defects in DNA repair pathways are generally more permissive for mutation accumulation and this likely underlies the predisposition of individuals inheriting cellular DNA repair defects to cancer formation (2). Increasing the rate of mutations in a cell population makes it more likely that the necessary mutations in oncogenes and tumour-suppressor genes will arise in a given lineage leading to cellular transformation and proliferation (3, 4).

While DNA repair defects can predispose to cancer formation, when they are acquired somatically they may serve as an 'Achilles Heel' that can be exploited for cancer therapy. This is because chemotherapeutic chemicals often work by damaging DNA, stalling DNA replication or disrupting mitosis, all of which could be potentially deleterious to a cell with an acquired DNA repair deficiency (2). Several well-documented examples of this type of gene-drug interaction exist for the BRCA1 and 2 genes where their mutational status can predict sensitivity of tumors to cisplatin and its derivatives or to PARP inhibitors (5). This latter interaction is considered an example of a synthetic lethal therapy because it recapitulates negative gene-gene interactions between the genes encoding PARP1/2 and BRCA1/2 (6). Indeed there are now multiple efforts underway to inhibit additional DNA

repair proteins themselves to further sensitize cancers to killing with genotoxic agents or to overwhelm tumor cells with already debilitated DNA repair capacity (7).

In cancer, as in model systems, there are typically surviving cells after a genotoxic insult. These cells bear a signature of mutations associated with the genotoxin they survived (e.g. cisplatin, temozolomide (8, 9)). Mutation signature analysis of tumor genomes has been refined in the past 5 years and a set of 30 canonical mutation signatures is now maintained in the COSMIC database (9-11). In some cases the etiology of mutation patterns is strongly linked to specific genetic mutations or environmental exposures, while in other cases the etiology remains unknown. Studies in model organisms have sought to dissect which aspects of a mutation signature are due to specific deficiencies in genome maintenance factors or to specific chemical treatments (12-14). Indeed, the largest such study to date in *C. elegans* characterized both a panel of mutant strains and the effects of Aflatoxin B1, mechlorethamine and cisplatin (14).

Of course the intention of genotoxin treatments clinically is to kill cells rather than mutagenize them. Model organism studies have also provided a means to map genetic networks underlying genotoxin sensitivity. The systematic identification of synthetic lethal interactions or chemical-genetic interactions has been led by studies in budding yeast, *Saccharomyces cerevisiae.* Indeed, a full pairwise gene-gene interaction study is now complete for both essential and non-essential yeast genes (15). In addition several thousand small molecules have been profiled for sensitivity and resistance across the yeast knockout (YKO) collections (16). Now researchers are combining these fields to understand the effects of chemical perturbations on genetic interaction networks and identifying gene-gene synergies in drug sensitivity (17, 18). In each of these studies the primary readout for synergy

between genes and chemicals is fitness as it is quantitative, simple to measure in high-throughput and informative. Nevertheless, other quantitative phenotypic readouts are possible and the YKO collection has been profiled by numerous biochemical, cytological and functional phenotypes (19).

Reasoning that DNA repair deficiencies would result in cell death, mutagenesis of survivors, or both after a genotoxic insult I assessed the overlap of fitness and mutagenesis for representative chemical genotoxins in yeast cells defective for all major DNA repair pathways. Quantifying growth and mutation rates showed strikingly little overlap between these parameters and further revealed a phenomenon we term synthetic hypermutation (SHyp). SHyp refers to mutation rates that exceed the expected combinatorial effects of a DNA repair mutant and drug treatment on mutation rate. I characterize one such interaction between *RAD1* and cisplatin by whole-genome sequencing and uncover evidence of a novel translesion synthesis mechanism for yeast DNA polymerase ζ (Rev3). Together these data define gene-drug interactions in a new way, underscore a novel mutation signature in yeast and apply mutation accumulation and whole genome sequencing to suggest new DNA repair mechanisms.

1.2 DNA Repair Processes and Cancer

Mammalian and yeast cells share a minimum of seven major DNA repair mechanisms to deal with various DNA damages. These dynamic mechanisms of repair are homologous recombination (HR), non-homologous end joining (NHEJ), nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), translesion synthesis (TLS), and the Fanconi anemia (FA) pathways (56).

1.2.1 Homologous Recombination (HR)

The cell uses two pathways to repair double-stranded DNA breaks (DSBs), HR and NHEJ. HR is a very efficient error-free mechanism that operates during S and G2 phases of the cell cycle, due to the need of a sister chromatid to be used as a template (57). After lesion recognition and initial end-processing by the MRN complex, the DSB is extensively resected by the BLM-EXO1 complex (58) to create 3' ssDNA overhangs. Recruited by RAD52, RAD51 forms a nucleoprotein filament around ssDNA and dictates homology search, strand exchange and Holliday junction formation (57). The helicase BLM and endonuclease MUS81 promote dissolution and resolution of Holliday junction intermediates respectively, which prevent the formation of crossover and noncrossover products in the case of MUS81, and noncrossover products by BLM (59, 60).

HR is the frontline mechanism that the cell uses to maintain genome integrity and in the repair of lesions caused by some anticancer drugs. The impact of its absence is reflected by observations in patients with heterozygous germline mutations in *BRCA1, BRCA2*, and *RAD51C* who display an elevated risk for the development of various cancers (61-63). Homozygous mutations in various other components of HR have been found to be recurrent in other types of cancers as well (90). Other germline mutations in members of the HR pathway include Bloom's syndrome (BS) and Werner's syndrome (WS) caused by mutations in the HR genes *BLM* and *WRN* respectively. BS patients are prone to the development of lymphomas, leukemias and carcinomas due to the high level of chromosomal instability and hyperrecombination (87).

1.2.2 Non-Homologous End Joining (NHEJ)

NHEJ is the other major pathway to repair DSB. NHEJ preferentially operates in the G1 phase of the cell cycle as it does not require an intact homologous template (64). DSBs are promptly bound by Ku70 and Ku80 subunits forming a heterodimer around free DNA ends; here the Ku complex recruits and activates the catalytic submit of DNA-PK (65). In a final step the broken ends are religated by XRCC4-LIG4. NHEJ can be an error-prone repair mechanism due to trimming of bases around DSBs and the gap filling of the ends. In addition to the classical NHEJ where no resection is required, two other alternative NHEJ mechanisms that requires end-resection of the ends are present in the cell. Microhomology-mediated end joining (MMEJ) that gives rise to small indels after strand resection and annealing of microhomologies, and single-strand annealing (SSA) that produces larger indels as a consequence of strand resection and large areas of homologies (66). From a medical perspective, defective NHEJ predispose to cancer formation, including lymphoma and brain tumors. From its components, mutations in *LIG4* cause Ligase IV syndrome that is characterized by microcephaly and predisposition to lymphomas and other type of cancers (88).

1.2.3 Nucleotide Excision Repair (NER)

Lesions that distort the DNA double helix configuration is repaired by the NER pathway. These lesions are produced by some alkylating agents and by psoralens and UV-induced lesions (CPDs and (6-4) pyrimidine photoproducts) (72). Two subpathways of NER exist and they primarily differ by the initial recognition stage of the lesion. Global-genome NER (GG-NER) and transcription-coupled NER (TC-NER) which utilizes proteins

XPC/RAD23B and RPA/XPA or RNA polymerase stalling for damaged site recognition respectively. In TC-NER, XPG and CSB are then recruited (32, 67). Subsequently GG-NER and TC-NER employ the same downstream components of the pathway, to first unwind dsDNA around the damaged base by XPB and XPD (recruited by TFIIH). This event triggers the recruitment of the nucleases XPF/ERCC1 and XPG that cut ssDNA at the 5' and 3' ends of the lesion respectively. Lastly, gap filling and ligation are carried out by PCNA, RFC DNA Pol δ, DNA Pol ε or DNA pol κ, and DNA ligase 1 or XRCC1-DNA ligase 3 (32, 68, 69).

Mutations in various genes on the NER pathway have been linked to germline disorders and at least 13 genes are associated with NER-related diseases. Xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD) are three such disorders associated with NER deficiency. XP is a rare autosomal-inherited and neurodegenerative disorder associated with mutations in the *XPA-V* and *XPV* genes. XP is characterized by increased risk for skin cancer after exposition to sunlight including squamous and basal cell carcinomas and melanomas. TTD is associated with mutations in the *XPB, XPD, TTDN1* and *TTDA* genes, whereas CS is associated with the *CSA, CSB, XPD* and *XPG* genes (91).

1.2.4 Base Excision Repair (BER)

BER corrects single-stranded DNA breaks (SSBs) and non-distorting DNA lesions, like uracil incorporation in DNA, damaged bases by reactive oxidative species (ROS), hydrolysis and methylation. Damaged bases are recognized and removed by DNA glycosylases, like UNG1, producing an apurinic or apyrimidinic site. PARP 1 and 2 play

important role in detecting SSBs and recruiting other BER repair components such as XRCC1. The APE1 nuclease is then recruited to the site of damage by XRCC1 to cut the abasic site at the 5' coordinate. Subsequently a complex formed by PCNA and DNA-Pol δ/ε is recruited by PCNA. Last, FEN1 cuts the flap oligonucleotide before resealing by ligase I takes place (70, 71).

1.2.5 Mismatch Repair (MMR)

MMR recognizes and repairs misincorporated bases and indels that are neglected by DNA polymerases during replication and recombination (73). Initially the detection of small mismatches is carried out by the MutSα complex (Msh2/Msh6) and large mismatches and insertion loops by the MutSβ complex (Msh2/Msh3). After binding and connection of the MutS with the PCNA/RFC complexes by the MutLα complex (MLH1 and PMS2), the exonuclease EXO1 is recruited to resect the mismatch and flanking DNA, followed by gap filling by the DNA Pol δ before resealing by ligase I (74). Predisposition to inherited and sporadic cancers arises as a consequence of malfunction of various components of the MMR pathway. Hereditary non-polyposis colorectal cancer (HNPCC or Lynch syndrome) is a consequence of germline mutations in either *MSH2, MSH6, MLH1* and *PMS2* (85). The most common feature of sporadic cancer genomes with defects in MMR genes are microsatellite instability (MSI), that is frequent small indels in tandem repeat sequences (microsatellites) (86). MSI in sporadic colorectal, gastric and endometrial cancers correlates with *MLH1* hypermethylation or mutations in various members of the MMR system. An additional feature of MMR deficiency is genome-wide hypermutation that differs from hypermutated genomes caused by aberrant replicative DNA polymerases.

1.2.6 The Fanconi Anemia Pathway (FA)

The Fanconi anemia pathway repairs interstrand-crosslinks (ICLs) and involves the coordination of components of the HR, NER and TLS pathways. ICLs are caused by various agents such as platinum-based compounds, mitomycin C (MMC) and alkylating agents; and in nature, by-products of lipid peroxidation and psoralens. The FA pathway is activated when replication forks encounter an ICL, then FANCM recruits the FA core complex to the damaged site (FANCA, B, C, E, F, G, L, M) (77). The FA core complex monoubiquitinates the ID complex (FANCD2/FANCI) that is relocalized to blocked replication forks (90). This step represents a key event in the FA pathway where monoubiquitinated ID complex promotes fork cleavage by FAN1 (and other nucleases), translesion synthesis and crosslink excision. Double-strand resection (by FAN1, CtlP, BLM and EXO1) is followed by strand invasion and HR-mediated DSB repair (75-77). Germline mutations in various genes including members of the FA core complex (*FANCA, B, C, E, F, G*, and *L*), *FANCD2* and *I*, *BRIP1* (*FANCJ*), *FANCM, PALB2* (*FANCN*), *BRCA2* (*FACND1*), *SLX4* (*FANCP*)-*SLX1* and *RAD51C* (*FANCO*) are known to cause the Fanconi anemia syndrome. FA patients have a predisposition to develop bone marrow failure, acute myeloid leukemia (AML), myelodysplasia and squamous cell carcinoma (77, 90).

1.2.7 Translesion Synthesis (TLS)

Translesion synthesis bypasses DNA damages encountered during replication by using specialized polymerases in situations where the regular replicative DNA polymerases ε and δ would fail. These special TLS polymerases have the ability to replicate past damaged bases due to their less stringent catalytic pockets, however at the cost of having diminished

processivity and fidelity. In humans five different TLS polymerases, each with different substrate specificities, can be called to sites of replication fork stalling; these include Polη, Polι, Polκ, REV1 and Polζ. In the yeast *Saccharomyces cerevisiae,* the two Y-family TLS polymerases Rev1 and Polη (encoded by the *RAD30* gene), and the B-family TLS polymerase Polζ (composed of the Rev3 and Rev7 subunits) can be recruited to sites of replication stalling (78).

Rev1 interacts with Polζ and encodes a deoxycytidyl transferase with a preference for incorporating a dCMP across from abasic sites and other aberrant bases. Rev1 is considered to be an error-prone TLS polymerase. DNA Polη is capable of preferentially incorporating A nucleotides across from various types UV-induced photoproducts and other damaged bases. DNA Polη can be error-prone or error-free depending of the substrate, thus it is an error-free polymerase when inserting two A's opposite UV-induced lesions and T-T CPD. DNA Polζ is a heterodimeric enzyme and Rev3 is the catalytic subunit. Polζ has some efficiency in bypassing damages caused by UV-light and is considered to possess error-prone activity, the latter due to the low mutational frequency phenotype exhibited by Polζ mutants. Polζ has low fidelity and processivity, however it cooperates with Rev1, Polη and Polδ (79). TLS polymerases are a significant source of genome mutability; mutations can be introduced during replication when TLS polymerases assume replicative functions of stretches of undamaged DNA (80).

In cancer, DNA Polη originated mutations in A- or T-coordinated clusters at a [T(A|T)] mutation motif, were identified in multiple myelomas (81). This mutated motif was also enriched in chronic leukemia and in B cell lymphomas (82, 118, 119). Furthermore, in

hepatitis C virus (HCV)-positive hepatocellular carcinomas, overexpression of TLS polymerases seems to originate an ApT>ApC mutation signature (83, 84). Inheritable human syndromes with defects in TLS polymerases exist. For instance mutations in *POLH*, the gene that encodes the human Polη, lead to xeroderma pigmentosum-variant type (XPV). XPV is an autosomal recessive disorder characterized by higher sensitivity to sunlight, DNA repair defects and increased frequency of skin cancers (78).

1.3 DNA Damaging Agents in Cancer Therapy

Several cancer drugs with DNA damaging properties have been used for many years now with various degrees of success (**Table 3**). These drugs represent the frontline of cancer therapy and take advantage of the DNA repair weaknesses that some cancers possess. For treatment they are used as a single agent or in combination therapy with DNA repair inhibitors. (92). Chemotherapeutic drugs can be classified by their mode of action into different categories.

Alkylating agents are a group of agents that covalently bind to DNA directly or after metabolization in the body. They can be subcategorized into monofunctional and bifunctional alkylating agents. Monofunctional alkylating agents have one active moiety that binds single bases. Some agents that belong to this group are temozolomide, nitrosourea compounds, alkylsulphonates and methyl methanesulfonate (MMS). On the other hand, bifunctional alkylators have two reactive sites to covalently bind DNA, creating crosslinks between one DNA strand and proteins, or alternatively intrastrand crosslinks (between two bases in the same strand) or interstrand crosslinks (between two bases on different strands). DNA interstrand crosslinks represent a very toxic DNA lesion as they can cause replication fork

stalling and DSBs. Many cancer drugs belong to this group including platinum-based compounds (i.e. cisplatin, carboplatin, etc.), mitomycin C and nitrogen mustards (i.e. mechlorethamine). Cisplatin is a widely used agent in the treatment of testicular, ovarian and non-small lung cancer (92). Cisplatin bind to purines and results in 65% GpG, 25% ApG, and 5-10% GpNpG intrastrand crosslinks and a lesser extent interstrand crosslinks and monoadducts (14).

Antimetabolites are a category of agents that resemble and compete with cellular nucleotides, nucleotide precursors and intermediates. They act by inhibiting nucleotide biosynthesis, interfering with DNA replication and causing cellular toxicity after genomic incorporation, however their mode of action is still an active area of research. 5-fluorouracil (5-FU), thiopurines and folate analogues are examples of antimetabolites used in cancer therapies. 5-FU is widely used in the treatment of colorectal, breast, and head and neck cancers and possesses various modes of action. It inhibits thymidylate synthase leading to so-called 'thymineless death' and causing imbalances in the pyrimidine nucleotide pool and uracil incorporation in DNA, and the accumulation of toxic DNA repaired intermediates leading to cell death (93).

Another strategy for cancer treatment is interfering with the ability of the cell to repair strand breaks that normally arise during replication. Topoisomerases are enzymes with the function of untangling supercoiled DNA that arise during replication, they resolve these structures by producing and resealing SSBs and DSBs. Some chemotherapies act by poisoning topoisomerases and trapping them bound to DNA. Camptothecin is a drug that selectively binds to topoisomerase I and DNA causing SSBs and replication-associated DSBs. Etoposide inhibits topoisomerase II causing DSBs (92).

Replication inhibition is another strategy commonly used by some chemotherapeutic drugs. Some drugs in this group are aphidicolin that inhibits DNA polymerases and hydroxyurea that inhibits ribonucleotide reductase, an enzyme required for the production of dNTPs (94, 95). Thus, by impairing DNA replication fork progression these drugs cause DSBs and other replication lesions. Finally, ionizing radiation and radiomimetic agents like bleomycin kill cells by causing SSBs and DSBs that are independent of replication.

1.4 Mutation Signature Analysis

1.4.1 Mutation Signatures in Cancer Genomes

The first mutational process discovered in a cancer was mutation caused by UV irradiation (96, 97). Here, UV light was shown to cause C>T or CC>TT substitutions. Later the *TP53* locus from several cancer patients was sequenced which confirmed the significance of DNA mutational signature analysis in cancer. The *TP53* signature of skin cancers shown C>T and CC>TT mutations at dipyrimidines. The *TP53* spectrum from lung cancers were C>A substitutions consisting with the effects of tobacco on guanines (98).

These early single-locus studies established the ground for the expansion of mutational signature extraction of cancer genomes to exome and whole-genome sequencing scale. Exome sequencing was preferred due to its low cost and to the predominance for cancers for mutations at protein-coding genes. Initial next-generation sequencing studies revealed tobacco signature in small cell lung cancer and UV light signature in melanoma, validating single-locus based results (100, 101). Other important signatures were extracted from various cancers that were driven by specific mutational processes. Most notably C>T in temozolomide-treated glioblastoma multiforme cancer (102), ultra-hypermutations in

endometrial and colorectal cancers associated with mutations in polymerase ε (103, 104), microsatellites with increase C>T and C>A in gastric cancer, and T>A at CpTpG sites in aristolochic acid-associated urothelial carcinoma (105).

Using computational and mathematical approaches to analyze initially a set of 21 breast cancer genomes and more recently >7,000 primary human cancers of 30 different types, 21 different mutational signatures were defined based on somatic substitutions in a 96-trinucleotide format (11, 118, 119). This format included the immediate 5' and 3' flanking sequence to the mutated base; and since there are 6 classes of base substitutions and 16 possible sequence contexts for a mutated base, 96 different mutated trinucleotides are possible. The immediate sequence context information was included to better differentiate operative mutational processes. While some of these 21 signatures were observed before, many others were novel.

Some mutational signatures were caused by endogenous DNA-damaging factors. For example, several cancers showed signatures of spontaneous deamination (signatures 1A and 1B). Namely C>T substitutions at NpCpG motifs. This pattern was consistent with C>T transitions at methylated CpG dinucleotides. Another source of deamination is caused by members of the cytidine deaminase family of enzymes AID and APOBEC. AID showed a preference for deaminating cytosines when they are 5' flanked by a purine. APOBECs on the other hand have a preference for deaminating cytosines in a TpC motif. In relation to cancer, the APOBEC signature was found in breast and other cancer types (signatures 2 and 13), and showed strand coordination and C>G specific to signature 13 (118, 119). In addition to this, a localized hypermutation (named *kataegis*) was identified in breast cancers that was

dominated by C>T and C>G at TpC. Given the resemblance it was proposed that the APOBEC enzymes were behind cancer *kataegis*.

Other signatures were also defined by the effects of exogenous DNA damage, either physical or chemical. Signature 7 that was defined by C>T mutations at dipyrimidines and CC>TT, was associated with UV exposure in cutaneous cancers (squamous cell skin carcinomas and malignant melanomas). This process also displayed mutational strand-bias suggesting transcriptional couple repair activity (TC-NER) (106, 118, 119).

The mutational spectrum caused by environmental and chemotherapeutic compounds were also identified and defined other mutational signatures. For instance, the alkylating agent temozolomide produced C>T transitions (signature 11), the tobacco smoking by-product benzo[a]pyrene caused G>T transversions (signature 4) and the plant extract aristolochic acid generated T>A transversions (signature 22) (118, 119).

Deficiencies in DNA repair pathways are also associated with mutational signatures in cancer genomes. The mismatch repair pathway (MMR) represents a frontline system to repair misincorporated bases and indels that arise during replication. The landmark mutation signature of MMR deficiency is microsatellite instability and some cancers harbored thousands of small indels. C>T transitions at NpCpG and C>A transversions at CpCpC from some cancers were also associated to MMR deficiency (signature 6) (118, 119).

1.4.2 Mutation Signatures in Model Organisms

Model organisms have been historically used to investigate cellular processes due to their ease of use comparing to their human counterparts. Similarly, model organisms are being increasingly employed to investigate mutational signatures and their processes

operating in cancer genomes resulting from genetic deficiencies and endogenous or chemical mutagens. In this effort various organisms were used, however the yeast *S. cerevisiae* and the worm *C. elegans* represent the most used systems. In this regard, mutation accumulation (MA) experiments and whole-genome sequencing (WGS) play an important role to discover genomic variants that can be attributed to cancer mechanisms. A mutation accumulation approach consists of exposing a certain number of yeast cells or *C. elegans* worms with genetic deficiencies to a series of single-cell passages for several generations, after which the complete genomes of single clones are sequenced (MA-WGS) (40). This approach allows the capture of genomic disturbances and their association to cancer genome features. Alternatively, cells can be subjected to single or repetitive exposure to genotoxins to study their effects, or a combination of genetic deficiency and genotoxic. In this manner, various genetic deficiencies with relevance to cancer and other processes were recapitulated in model organisms, as well as the genomic footprints of environmental, chemotherapeutic and other relevant compounds were captured.

Likewise, the spontaneous mutation rates constitute an important metric for analysis that can be extracted using a MA-WGS approach. This information can be used as a reference to compare other sources of mutation. In this regard and using this approach, the substitution mutation rate of diploid *S. cerevisiae* was estimated in 1.67×10^{-10} per base per generation (107). This rate was lower than estimates from haploid yeast (0.33×10^{-9} per base per generation) (108), but close to other estimates from diploid yeast using a WGS approach (2.9×10^{-10} per base per generation) (109).

The power of yeast genetics combined with MA-WGS method revealed the signatures of cancer-relevant and other genetic perturbations. For example, an increase of

indels at homopolymeric tracts (HP) and C>T transitions was attributed to deficiencies in the MMR pathway in yeast (41, 110, 111). These findings were consistent with genomic features present in colorectal and other cancers (112, 118, 119). MMR deficiency was also combined with deficiencies in replicative DNA polymerases to reveal the strand specificity of the three major DNA polymerases (113, 114). These studies confirmed the primary roles for Polα and Polδ in lagging-strand replication and Polε for leading-strand replication. Further studies revealed the genomic consequences of several genome instability and DNA damage response deficiencies in yeast (12, 13).

In a comprehensive study in the worm *C. elegans*, the mutational signatures of three carcinogens and various DNA repair pathway deficiencies were revealed using MA-WGS (14). Genome rearrangements at telomeres and C>T transitions were observed in *mrt-2* and *ung-1* deficient lines respectively. In addition to genetic deficiencies, the mutational spectrum of cisplatin, aflatoxin B1 and mechlorethamine were investigated. Notably, aflatoxin produced a significant increase in C>T and C>A base substitutions in GpCpG context, and cisplatin an even higher C>A base substitutions in CpC context in *xpf-1*-deficient worms. Cisplatin also produced rare dinucleotide substitution, in particular CT>AC and CA>AC. In either case, they resembled the cancer footprints seen in aflatoxin-induced hepatocellular carcinoma and melanoma respectively.

Other mutagen spectra were also obtained using *C. elegans*. Enrichment for C>T transitions in ethyl-methane sulfonate (EMS) and T>A and C>T substitutions in *N*-ethyl-*N*-nitrosourea (ENU) treated animals were observed (116, 117). In yeast, MMS was found to induce clusters of strand-coordinated multiple point mutations within ssDNA that is formed during replication (81). Similarly, MMS-associated mutations were also found in multiple

myelomas, prostate cancer, and head and neck squamous cell carcinomas (HNSCCs). Lastly, overexpression of human APOBEC family of enzymes in yeast recapitulated the *kataegis* signature seen in breast cancer; namely, clustered C>T transitions that were associated with DNA breaks (42).

1.5 Thesis Objectives

The first objective of this project was to identify cases of synthetic hypermutation caused by gene-drug interactions from a screen of a combination of 12 DNA-repair deficiencies and 5 DNA-damaging agents; and compared the results of this mutation-based method to a fitness-based screen of the same interactions in order to determine mutation and fitness relationship.

The second objective was to characterize a synthetic hypermutation phenotype by whole-genome sequencing to determine the mutation landscape produced by the effects of a drug in a DNA-repair impaired genome. I sequenced 12 genomes deficient in Rad1 that were treated with cisplatin and extracted single nucleotide variants, small indels, copy number variants and the sequence context of these mutations.

The third objective of this project was to use the mutation information extracted from whole-genome sequencing to give mechanistic insights of how mutations occur; and to better understand functions of proteins involved in DNA damage pathways. I further investigated the peculiarity of the role of DNA polymerase ζ (Rev3) in causing specific mutations in nucleotide excision repair deficient cells treated with cisplatin.

Chapter 2: MATERIALS & METHODS

2.1 Yeast Strains and Growth Curve Analysis

Yeast strains are listed in **Table 1**. All gene deletions were PCR-confirmed. Unless otherwise indicated yeast were grown on SD minimal media plus histidine, uracil, leucine, lysine and methionine + 2% dextrose (SD+5) at 30°C. Growth curves were conducted in triplicate in 96-well plates using a TECAN M200 plate reader at 30°C (20). OD600 readings were taken every 30 minutes for 24 hours. The area under the growth curve was used as a measure of fitness and compared to the expected fitness value (multiplicative model) using a t-test (21). Unless otherwise indicated, drug concentrations were determined empirically by permitting well saturation of *rad52*Δ cells over 72 hours, to be 200 μM etoposide (Etp), 1 μM camptothecin (Cpt), 10 μM 5-fluorouracil (5FU), 0.0005% methyl methanesulfonate (MMS), and 10 μM cisplatin (Cis).

Strain name	Genotype	Source
PSY1064	MATα *ura3Δ0 leu2Δ0 his3Δ1 rad30Δ::KanMX*	Open Biosystems
PSY1065	MATα *ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 met15Δ0 rev3Δ::KanMX*	This study
PSY1705	MATα *ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 rev1Δ::KanMX*	Open Biosystems
PSY1522	MATα *ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 mlh1Δ::KanMX*	Open Biosystems
PSY1536	MATα *ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 yku80Δ::KanMX*	Open Biosystems
PSY1067	MATα *ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 met15Δ0 exo1Δ::KanMX*	This study
PSY758	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 apn1Δ::KanMX*	Phil Hieter
PSY759	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rad2Δ::KanMX*	Phil Hieter
PSY760	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rad23Δ::KanMX*	Phil Hieter
PSY761	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rad26Δ::KanMX*	Phil Hieter
PSY762	MATa *ura3Δ0 leu2Δ0 his3Δ1 sgs1Δ::KanMX*	This study
PSY764	MATa *ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 rad1Δ::KanMX*	This study
PSY766	MATa *ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 mus81Δ::KanMX*	This study
PSY768	MATa *ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 mph1Δ::KanMX*	This study
PSY771	MATα *ura3Δ0 leu2Δ0 his3Δ1 rad1Δ::KanMX rev1Δ::KanMX*	This study
PSY772	MATa *ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 rad1Δ::KanMX rev3Δ::KanMX*	This study
PSY774	MATa *ura3Δ0 leu2Δ0 his3Δ1 rad1Δ::KanMX rad30Δ::KanMX*	This study
BY4741	MATa *ura3Δ0 leu2Δ0 his3Δ1 met15Δ0*	Open Biosystems
PSY776	MATa/α *ura3Δ0/ura3Δ0 leu2Δ0/LEU2 his3Δ1/his3Δ1 lys2Δ0/LYS2 met15Δ0/MET15 can1Δ0::LEU2-MFA1pr::HIS3/CAN1 sgs1Δ::KanMX/sgs1Δ::KanMX*	This study
PSY778	MATa/α *ura3Δ0/ura3Δ0 leu2Δ0/LEU2 his3Δ1/his3Δ1 lys2Δ0/LYS2 met15Δ0/MET15 can1Δ0::LEU2-MFA1pr::HIS3/CAN1 rad1Δ::KanMX/rad1Δ::KanMX*	This study
PSY780	MATa/α *ura3Δ0/ura3Δ0 leu2Δ0/LEU2 his3Δ1/his3Δ1 lys2Δ0/LYS2 met15Δ0/met15Δ0 can1Δ0::LEU2-MFA1pr::HIS3/CAN1 mus81Δ::KanMX/mus81Δ::KanMX*	This study
PSY783	MATa/α *ura3Δ0/ura3Δ0 leu2Δ0/LEU2 his3Δ1/his3Δ1 lys2Δ0/LYS2 met15Δ0/MET15 can1Δ0::LEU2-MFA1pr::HIS3/CAN1*	This study

Table 1. *S. cerevisiae* strains used in this study. All strains are derived from knockout collections available at Open Biosystems.

2.2 Fluctuation and Synthetic Hypermutation Analysis

Fluctuation analysis at *CAN1* was performed in 96-well plates, with a minimum of 18 replicates per condition, exactly as described (22), except that cells were grown to saturation in the presence of drugs in SD+5 for 48 hours at 30°C. Average cell numbers in each well were determined using a TC-20 cell counter (BioRad), and canavanine-resistant colony counts were converted to mutation rates using the Ma-Sandri-Sarkar maximum-likelihood estimator calculator (FALCOR program) (23). Synthetic hypermutation was calculated as:

$$SHyp = O(M_{ikj}) > E(M_{ij,kj})_P$$

Where M_i is the mutation rate of mutant *i;* M_k is the mutation rate caused by drug *k* in WT; M_{ik} is the mutation rate caused by drug *k* in mutant *i*; each are expressed as $_{j,}$ the fold increase over untreated WT, to give observed mutation rates $O(M)$. $E(M_{ij,kj})_P$ is the expected mutation rate as the product model that is defined as $E(M_{ij,kj})_P = M_{ij} \cdot M_{kj}$. Alternatively, the additive expected mutation rate was also considered and is defined as $E(M_{ij,kj})_A = M_{ij} + M_{kj}$ (**Fig. 2**). Synthetic hypermutation was considered if the lower 95% confidence interval of the observed mutation rate fell above the expected value.

2.3 Mutation Accumulation and Whole-Genome Sequencing

Diploid WT or homozygous diploid mutant strains were pulsed with 100 μM cisplatin for 2 hours prior to plating on YPD plates to rescue surviving cells. Diploids were chosen in order to buffer potentially deleterious mutations and a pulse of cisplatin ensured that resistance mutations could not accumulate to alter the mutation spectrum. For mutation accumulation (MA) lines single colonies were plated on YPD and passaged 40 times by

restreaking to fresh YPD every 2-3 days (an estimated 1000 generations). 12 independent lines of WT or *rad1*Δ/*rad1*Δ that went through MA and 12 independent cisplatin-treated WT or mutant lines were outgrown for genomic DNA preparation (24). Isolated genomic DNA was subjected to 125 bp paired end sequencing at the Michael Smith Genome Sciences Centre using the Illumina HiSeq2000 platform. Sequence files are deposited at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) (SRP091984). Reads were aligned to the Saccer3 reference genome using Burrow-Wheeler Aligner 0.5.7 (25). Variants were identified exactly as described previously (13). Mutations were extracted by comparing evolved to parental genomes. Variant calling was carried out using the mpileup utility of the SAMtools package 0.1.18 (26) with a threshold of 20. Copy-number variants were analyzed using a locally modified version of CNAseq (27). In addition, we manually verified variant calls with the Integrated Genomics Viewer (Broad Institute).

2.4 Mutation Signature Analysis

Flanking sequence was extracted from the Saccer3 genome based on the mutation positions determined (See Supplementary Table S2 at http://biorxiv.org/content/early/2016/11/08/086512). The enrichment of trinucleotide motifs within the diploid *rad1*Δ+cisplatin genomes was calculated as described based on the 20 bases either side of a mutation to give a set of 41mers (28). Briefly, enrichment of a mutation signature is the ratio of the number of mutations at a motif multiplied by the total number of C or G bases in 41mers flanking the mutation, divided by the number of mutations of that type (e.g. C>N) multiplied by the total number of motifs present in the 41mers. The significance of these enrichments was calculated using the hypergeometric distribution.

pLogos were generated online http://plogo.uconn.edu/ (29). As foreground input n(fg), 41-mer sequences from the *rad1*Δ-cisplatin genomes were used. These sequences were centered at the mutated C for all substitutions with fixed positions at C and GC. As background input n(bg), 41-mer sequences automatically generated by the pLogo tool from the Saccer3 yeast genome were used after removing duplicate sequences. Reanalysis of variants from Stone et al. was performed by manually scoring whether the SNV type matched base immediately 3' of the mutation (30). The proportions of these 3'-templated SNVs were compared with a Fisher exact test.

Chapter 3: RESULTS

3.1 A Network of Genotoxin-induced Fitness Defects and Mutation Rates

To investigate gene-drug interactions, I first established a panel of DNA repair mutants representing the major DNA repair pathways in yeast (**Table 2**). Haploid yeast bearing gene deletions impairing homologous recombination (HR), non-homologous end joining (NHEJ), nucleotide-excision repair (NER), base-excision repair (BER), mismatch repair (MMR), translesion synthesis (TLS), Fanconi Anemia-like (FA) and post-replication repair (PRR) pathways were chosen to represent loss or deficiencies of each pathway. To account for the multiple steps of, and routes for, DSB (DNA double-strand break) repair by HR we also included *exo1*Δ, *rad52*Δ, *sgs1*Δ, *mus81*Δ and *mph1*Δ lines as these mutants specifically impair end resection, strand exchange, double Holliday junction dissolution, resolution or other HR steps respectively (31). This panel of wildtype (WT) and 12 mutant strains (**Table 2**) was exposed to five classes of DNA damaging agents, represented by the bifunctional alkylator cisplatin (Cis), the antimetabolite 5-fluorouracil (5FU), the topoisomerase I inhibitor camptothecin (Cpt), the topoisomerase II inhibitor etoposide (Etp), and the monofunctional alkylator methyl methanesulfonate (MMS) (**Table 3**). We screened 78 pairwise gene-drug combinations plus WT for changes in fitness and, in 76 cases, were also able to measure mutation rates (*rad5*Δ growth was impaired to a degree that prevented fluctuation analysis).

Yeast gene	Human homolog	DNA repair pathway	Functions of gene product
APN1	*APE1*	BER	Abasic site endonuclease and 3'-5' exonuclease involved in repair various damages during BER
EXO1	*EXO1*	MMR, HR	5'-3' exonuclease and flap-endonuclease involved in recombination and DSB repair
MLH1	*MLH1*	MMR	Protein required for MMR by binding PMS2 or other MMR proteins
MPH1	*FANCM*	FA, HR	5'-3' translocase and branch migration activity during HR, recruitment of FA core complex and BLM
MUS81	*MUS81*	HR, BIR	3'-flap endonuclease required for repair of DSB during HR by dHJ resolution and stalled replication forks in BIR
RAD1	*ERCC4* (XPF)	NER, FA	Endonuclease required for incision on the 5' side of a DNA lesion during NER
RAD5	*HLTF*	PRR	Helicase/ubiquitin ligase involved in replication fork regression during PRR
RAD30	*POLH* (Polη)	TLS	DNA polymerase eta that participates in TLS to bypass specific DNA lesions, commonly CPD
RAD52	*RAD52*	HR, SSA	Protein involved in DSB repair by facilitating Rad51 to bind ssDNA
REV3	*REV3L* (Polζ)	TLS	Catalytic subunit of DNA polymerase zeta that participates in TLS and DSB repair
SGS1	*BLM*	HR	5'-3' helicase that promotes dissolution of dHJ intermediates during HR
YKU80	*XRCC5* (KU80)	NHEJ	Protein subunit of Ku complex, it relocates to DSB during NHEJ

Table 2. Genes and DNA repair pathways investigated for gene-drug interactions.

DNA-damaging agent	Category	DNA lesions
Cisplatin	Bifunctional alkylator	Intra and interstrand crosslinks, DSBs, replication fork stalling, bulky adducts
Methyl methanesulfonate	Monofunctional alkylator	Methylation, replication fork stalling, DSBs
5-Flurouracil	Antimetabolite, TS inhibitor	dTTP depletion, dUTP misincorporation, mutations
Camptothecin	Topoisomerase I inhibitor	Replication-associated DSBs, SSBs
Etoposide	Topoisomerase II inhibitor	DSBs, SSBs

Table 3. DNA-damaging agents tested in this study.

Growth rates were measured over a period of 24 hours and the area under the curve was calculated and normalized to the untreated WT to measure fitness (**Fig. 1A**). Overall, I observed that Cpt and Etp had no significant additional effects on any strain using the concentrations tested and although 5FU inhibited growth of all lines irrespective of the genetic background it only showed a synergistic interaction with *apn1*Δ. I also observed that cisplatin had synergistic gene-drug interactions with *rad1*Δ, *rad5*Δ, *rad52*Δ, *rev3*Δ and *sgs1*Δ and that MMS had synergistic interactions with *rad5*Δ and *sgs1*Δ ($p<0.05$; **Fig. 1A**). These interactions are consistent with, for example, the known roles for BER in repair of uracil in DNA; or the known complexity of repairing cisplatin interstrand crosslinks (ICLs) which involves at least HR, NER, FA and TLS. Moreover, this screen may under-represent chemical sensitivities because very low concentrations of drug were used to allow comparison with fluctuation assays below.

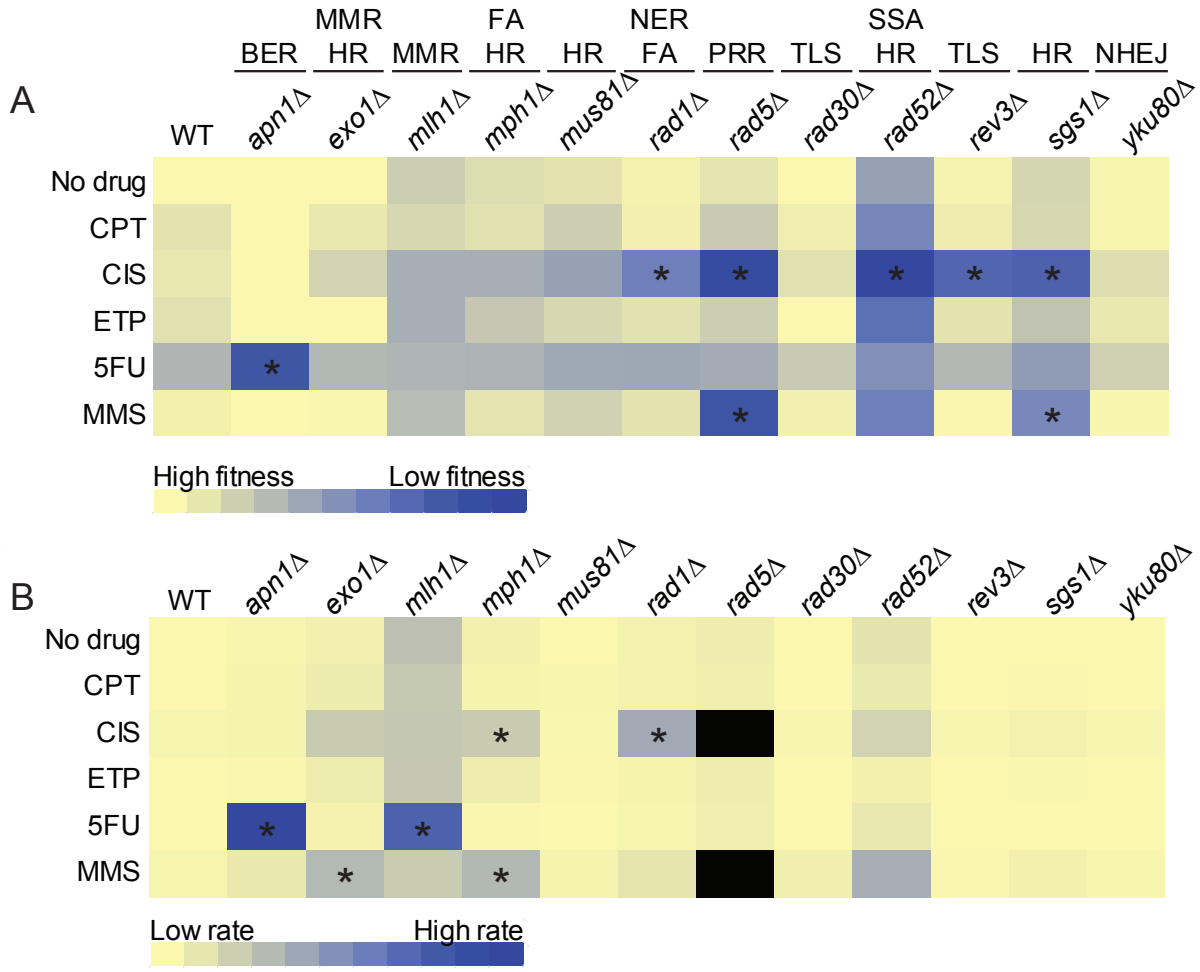


Figure 1. Gene-drug interactions measured by fitness and mutation rates. Heat maps illustrating (A) fitness, and (B) *CAN1* mutation rate, relative to WT. Interactions significantly greater than expected ($p>0.05$) are indicated *. The transition from yellow to blue indicates greater fitness defects or higher mutation rates. Black boxes for *rad5*Δ indicate not done.

Mutation rates of the same matrix were quantified at *CAN1* using a well-plate fluctuation assay (**Fig. 1B** and **Appendix A**) (22). In untreated cells the baseline mutation rates matched with previously reported rates for each mutant (**Table 4**). Again I observed that Cpt and Etp had no major mutator effects regardless of the genetic background at the given drug concentrations. On the contrary, cisplatin, 5FU and MMS increased the mutation rates of specific repair mutants. When this increase in mutation rate exceeded the expected effects of multiplying the effect of the repair deficiency and the effect of the chemical on WT, we defined it as synthetic hypermutation (SHyp). Applying this metric I identified six cases of SHyp: *mph1*Δ-Cis, *rad1*Δ-Cis, *apn1*Δ-5FU, *mlh1*Δ-5FU, *exo1*Δ-MMS and *mph1*Δ-MMS (**Fig. 1B** and **Fig. 2)**. Of these cases *rad1*Δ-Cis, *apn1*Δ-5FU and *mlh1*Δ-5FU showed the most severe SHyp phenotype (17.8, 45.1 and 33.4 relative rates respectively). Importantly, cases of SHyp did not overlap with fitness defects except for *rad1*Δ-Cis and *apn1*Δ-5FU and conversely many strains with fitness defect exhibited no increase in mutation rate. Indeed, survival analysis of cisplatin sensitivity of *sgs1*Δ, *mus81*Δ and *rad1*Δ show nearly identical viability curves over cisplatin concentration but different mutational outcomes (**Fig. 3**). Together the data show that decreased fitness and increased mutability by genotoxins in DNA repair deficient cells are not perfectly linked. This is likely both an issue of dose and DNA repair mechanisms wherein some gene-drug combinations are more likely to elicit toxic or irreparable intermediates.

Yeast genotype	CAN1^R rate (x10^{-7})		Increase over WT	Previous estimates
	This study	95% CI		
Wild-type	3.55	2.86-4.31	1.00	5.9 (5); 3.8 (7); 3.4 (3); 3.27 (4); 3 (8); 2.80 (1); 1.52 (2)
*apn1*Δ	7.93	3.86-12.99	2.23	6.9 (8)
*exo1*Δ	11.25	7.78-15.17	3.17	11.47 (4)
*mlh1*Δ	45.84	30.04-63.98	12.91	53 (3)
*mph1*Δ	11.16	7.74-15.02	3.14	12-46 (8); 11.1 (9)
*mus81*Δ	4.50	2.7-6.60	1.27	
*rad1*Δ	10.98	7.70-14.66	3.09	
*rad5*Δ	13.89	9.54-18.81	3.91	21.5 (3); 5.1,7.4, 9.8 (8)
*rad30*Δ	3.35	1.93-5.04	0.95	
*rad52*Δ	21.07	14.70-28.25	5.94	46.4 (3); 27,29 (6)
*rev1*Δ	1.13	0.52-1.91	0.32	1.1 (8)
*rev3*Δ	2.52	1.34-3.94	0.71	
*sgs1*Δ	4.14	2.47-6.11	1.17	
*yku80*Δ	2.82	1.51-4.41	0.79	2.3 (8)

(1) Serero A, *et al.* (2014) *Proc Natl Acad Sci USA* 111(5):1897-1902.
(2) Lang GI & Murray AW (2008) *Genetics* 178(1):67-82.
(3) Huang ME, *et al.* (2003) *Proc Natl Acad Sci USA* 100(20):11529-11534.
(4) Doerfler L, *et al*. (2011) *Genome Integr* 2:8. doi: 10.1186/2041-9414-2-8.
(5) Schmuckli-Maurer J, *et al*. (2003) *Nucleic Acid Res* 31(3):1013-1023.
(6) Kokoska RJ, *et al*. (2000) *Mol Cell Biol* 20(20):7490-7504.
(7) Johnson RE, *et al*. (1996) *J Biol Chem* 271(13):7285-7288.
(8) Schurer KA, *et al*. (2004) *Genetics* 166(4):1673-1686.
(9) Ang JS, *et al*. (2016) *Genetics* pii: genetics.116.192211.

Table 4. Baseline mutation rates of untreated strains.

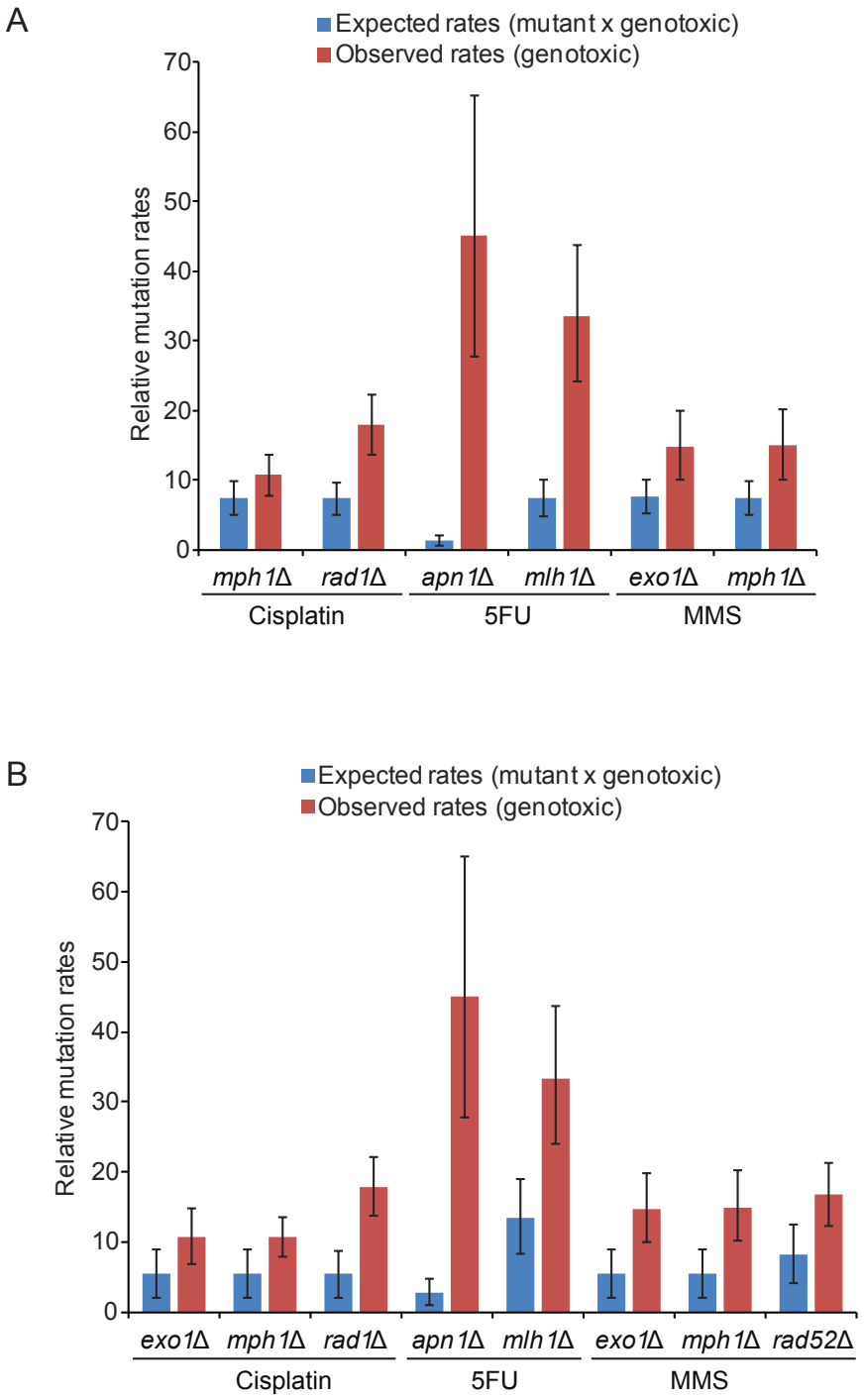


Figure 2. Summary of synthetic hypermutation hits by two models. The multiplicative model defines six cases of SHyp (A) and the additive model eight cases of SHyp (B). Synthetic hypermutation was considered when the lower 95% confidence interval of the observed mutation rate fell above the calculated expected value. See methods for calculations.

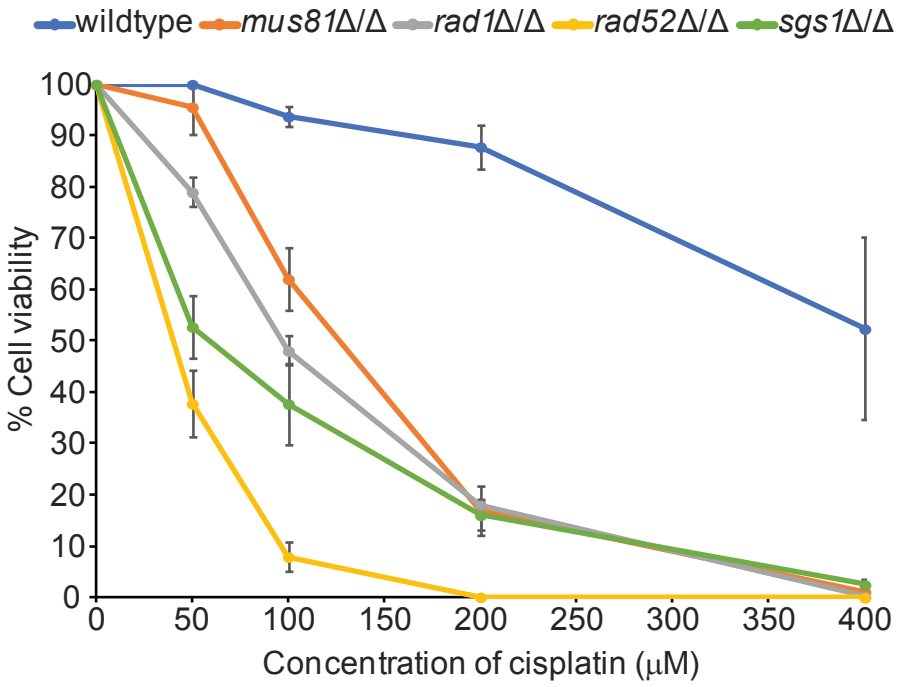


Figure 3. Dose viability curves of cisplatin sensitive mutants. For survival analysis, diploid strains were exposed to the indicated concentrations of cisplatin for 2 hours in SD minimal media. $6x10^6$ treated cells were diluted and grown as single colonies on YPD plates to score the proportion of surviving cells.

3.2 The Incision Step of NER Mediates Synthetic Hypermutation Upon Cisplatin Treatment

Rad1 is a nuclease involved in cleaving DNA flanking bulky lesions during NER. Recognition of these lesions can occur during transcription or globally due to the action of upstream NER components that recognize distorted DNA helices (32). To determine which aspects of NER exhibit SHyp with cisplatin we tested mutation rates in three other NER mutants, *rad2*Δ, *rad26*Δ and *rad23*Δ. Deletion of *RAD2* which, like Rad1, encodes a nuclease required for DNA incision and is related to human XPG, also exhibited SHyp with cisplatin (**Fig. 4**). However, deletion of *RAD26*, which selectively impairs transcription-coupled (TC) NER, or *RAD23* which partially impairs global and TC-NER but allows efficient DNA incision, had no SHyp phenotype with cisplatin (**Fig. 4**). Since Rad1 acts first, and Rad2 is required in a structural role to position Rad1 (32), loss of either would block incision and alternative mechanisms of unhooking of the cisplatin cross-link must be error prone.

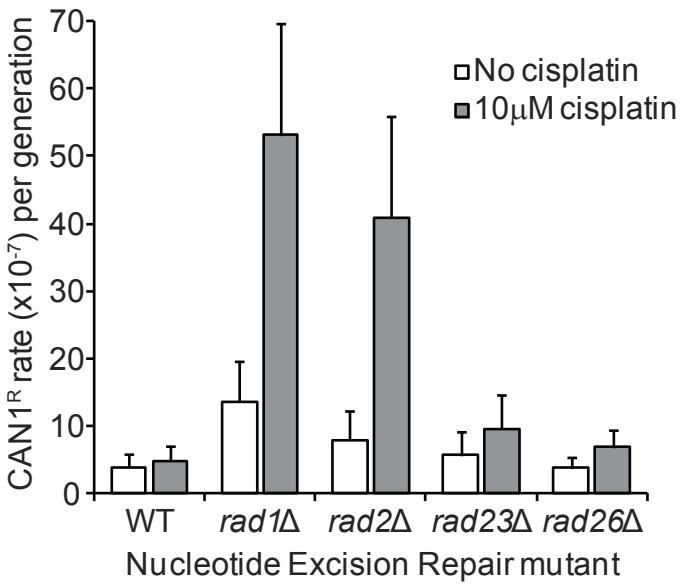


Figure 4. *CAN1* mutation rates of other NER deficient strains in cisplatin. *rad1*Δ and *rad2*Δ exhibit SHyp ($p<0.05$).

3.3 Mutation Accumulation and Whole Genome Sequencing

It is possible that *CAN1* showed variable mutation rates in repair deficient strains due to a locus specific effect. To explore the genomic spectrum of mutations and identify any evidence of such bias, I pulsed diploid WT, *rad1*Δ, *sgs1*Δ and *mus81*Δ with a higher dose of cisplatin and sequenced the whole genomes of 12 independent survivors for each strain (**Fig. 5A**). Consistent with the *CAN1* fluctuation analysis data, the *rad1*Δ mutant strain acquired significantly more mutations than WT or the other mutants (**Fig. 5B**). As expected based on the preference of cisplatin for crosslinking guanine residues, the most common mutation types were single-nucleotide variants (SNVs) at C:G basepairs. Moreover, indels seen in *rad1*Δ, which were primarily -1 deletions, occurred at C:G basepairs, and normally rare dinucleotide substitutions were also evident at motifs containing C:G basepairs (**Fig. 5C and 5D**). To ensure that the rate and type of mutations were due to cisplatin and not simply *rad1*Δ, we passaged diploid WT and *rad1*Δ cells through 40 single-cell bottlenecks and sequenced 12 independent whole genomes after this 1000 generation mutation accumulation (MA) experiment (**Fig. 5A**). The mutation rate and type evident in the *rad1*Δ MA lines was similar to the *CAN1* rate in untreated *rad1*Δ cells, showed proportionally more mutations at T:A basepairs and no dinucleotide substitutions (**Fig. 6**). Thus the specific combination of *RAD1* deletion and cisplatin treatment engage an error prone repair process that causes a global increase in specific types of mutations.

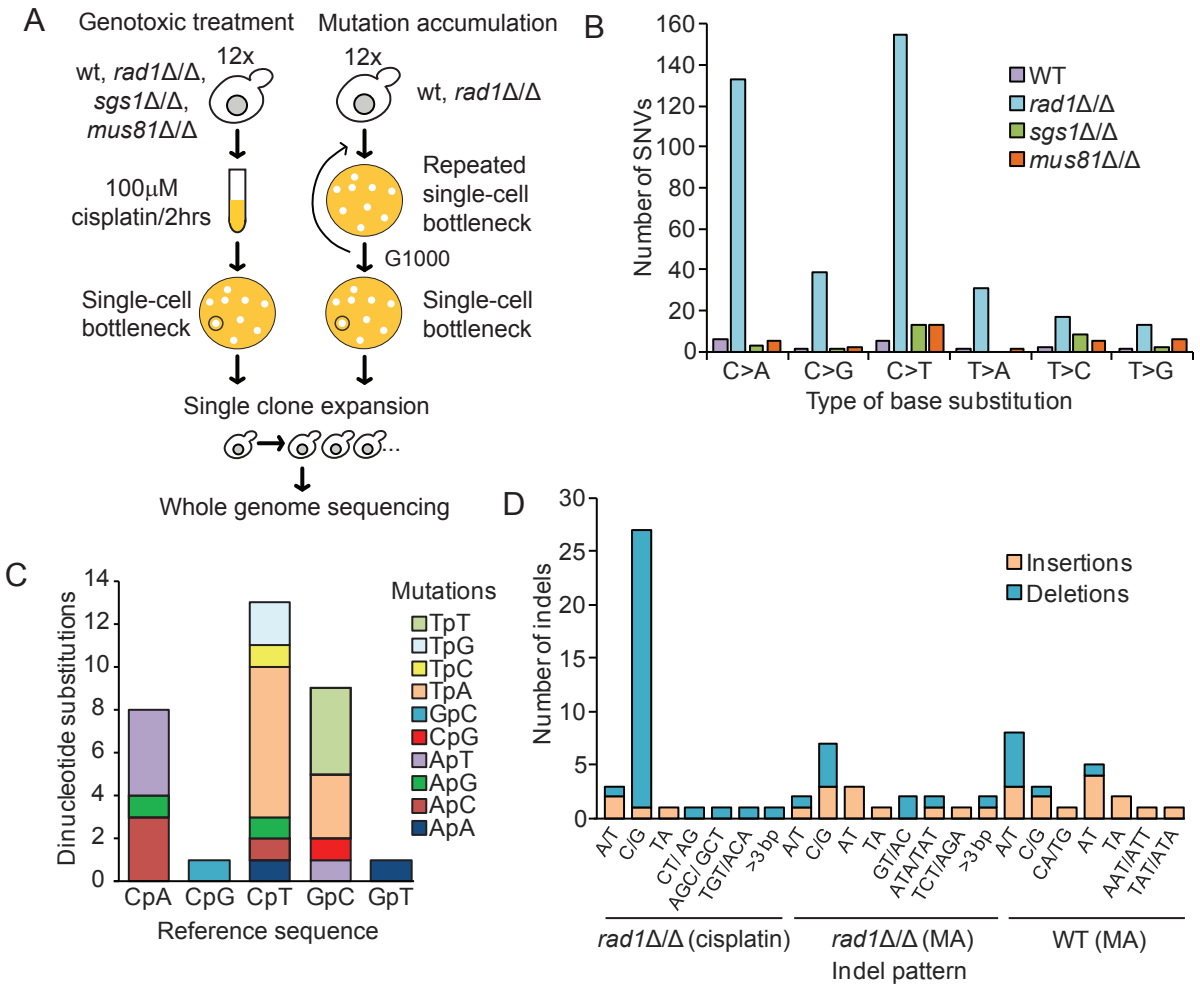


Figure 5. Mutation accumulation and whole genome sequencing. (A) Experimental design of drug treatment and 1000 generation (G1000) approaches for the indicated strains. (B) Summary of WGS results of single nucleotide variants in cisplatin treated WT, *rad1*Δ, *sgs1*Δ and *mus81*Δ homozygous diploid strains. SNV number and type are summarized in the bar graphs color-coded as indicated. SNVs are oriented to initiate with the pyrimidine base by convention (i.e. C>A encompasses C>A and G>T). (C) Dinucleotide substitutions in *rad1*Δ-Cis genomes. The parental dinucleotide sequences are on the x-axis and the mutated sequence is represented by the color of the bar. (D) Pattern of small indels in *rad1*Δ-Cis and mutation accumulation (MA) genomes. Indel size and sequence are on the x-axis and indel types are color coded.

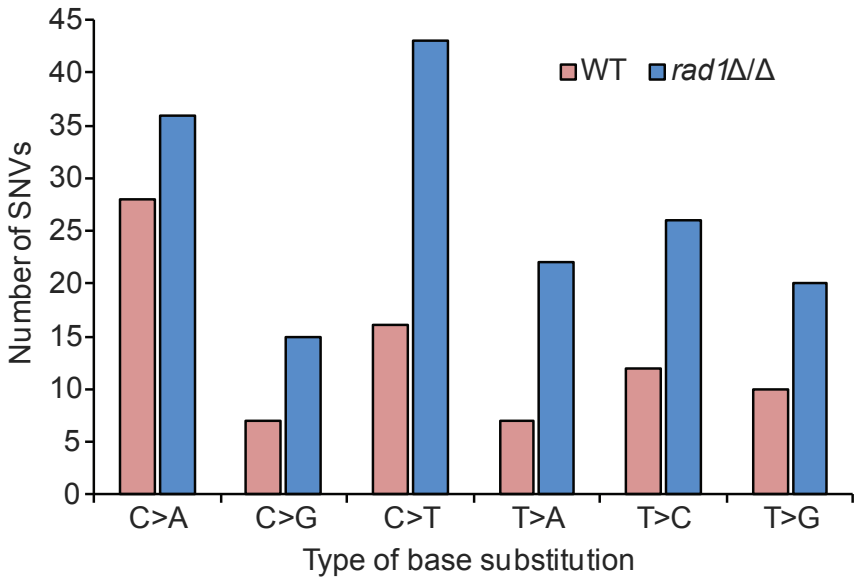
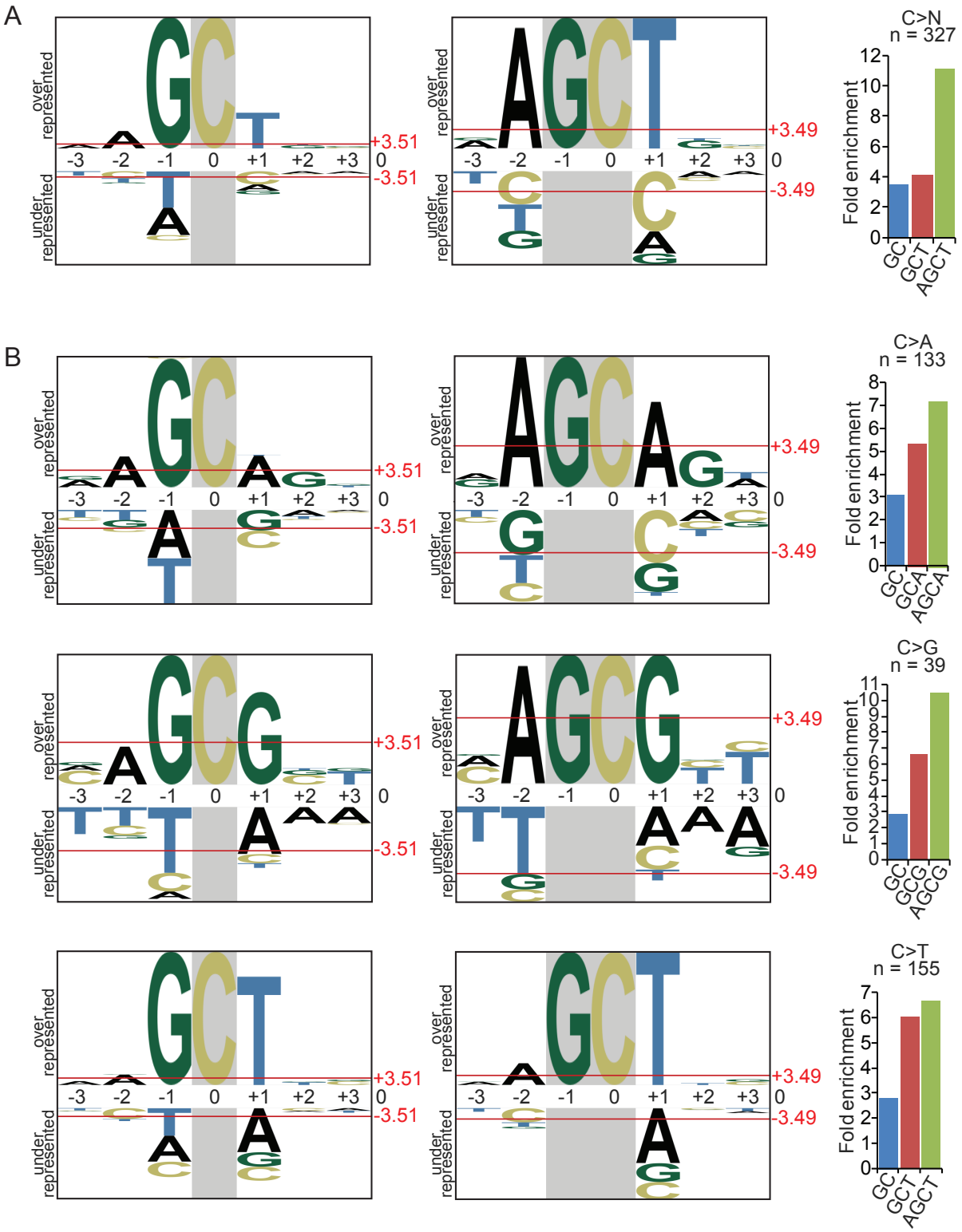


Figure 6. Spectrum of mutations in untreated 1000 generation passaged *rad1*Δ cells. SNVs are oriented to initiate with the pyrimidine base by convention (i.e. C>A encompasses C>A and G>T).

3.4 Cisplatin-*rad1*Δ Mutation Signatures Suggest 3' Templating of Base Substitutions

The flanking sequences around mutations can define a signature which is associated with a particular genetic or chemical perturbation. Indeed, previous work in *Caenorhabditis elegans* has examined cisplatin treatment of several mutant backgrounds including *XPF-1* (*i.e. C. elegans RAD1*) deficient animals (14). We extracted flanking sequences for the predominant C>N mutations in cisplatin-treated genomes and performed pLogo and enrichment analysis (**Fig. 7A** and **Appendix B1**) (28). This analysis revealed that GpCpT motifs were favored when all C>N mutations were considered. Since cisplatin is expected to preferentially crosslink guanine residues, this signature suggests that rarer interstrand crosslinks are the mutagenic lesion in yeast *RAD1* mutants. The observed GpCpN motif differs from *C. elegans* where CpCpT motifs were preferentially mutated in *xpf-1* worms (14). Since intrastrand crosslinks are much more common than interstrand crosslinks (33), our data suggest that *rad1*Δ yeast are either able to repair these in an error-free manner, even when lacking a key NER protein, or that they die and are removed from the experiment. Further analysis showed that, for each mutation type, adenine was significantly enriched at the -2 position (i.e. ApGpCpN) (**Fig. 7A center** and **Appendix B1**). When we analyzed the 3' position a different picture emerged, the mutation type varied with the +1 base, such that C>T mutations preferentially occurred in GpCpT motifs, and C>A mutations in GpCpA motifs and C>G mutations occurred at GpCpG motifs (**Fig. 7B** and **Appendix B2**). Thus, the base which is erroneously inserted at the damaged site is selected by basepairing with the 3' nucleotide.



Legend on next page.

Figure 7. Mutation signature analysis induced by cisplatin in *rad1*Δ diploid yeast. (A) Flanking nucleotides for C>N substitutions are presented as pLogo plots and fold enrichment. The mutated C is centered in each plot with fixed positions at C (left) and GC (center). The red line indicates significant enrichment $p<0.05$ and the font size indicates the magnitude of enrichment. The increasing fold enrichment for 2- 3- and 4-nucleotide motifs containing a mutated C is shown (left). All analysis was performed with a set of 41-mers and cropped for visualization (see methods) (B) Flanking analysis for C>A, C>G and C>T mutations as presented in A. Note the +1 base changing with the mutation type.

3.5 DNA Polymerase ζ Drives *rad1*Δ-cisplatin Hypermutation

The mutation signature analysis suggests error-prone DNA replication is likely causing mutations in *rad1*Δ cells treated with cisplatin. To identify the relevant polymerase I deleted all enzymatic TLS components individually in the *rad1*Δ background. Loss *of RAD30,* which slightly positively affected fitness when compared with the *rad1*Δ single mutant, showed no effect on *rad1*Δ hypermutation (**Fig. 8A** and **8B**), however additional loss of *REV1* or *REV3* led to inviability when treated with 10 μM cisplatin (**Fig. 8A**). This is consistent with the DNA damage tolerance role of *REV3* and the idea that TLS is supporting viability in *RAD1* deficient cells. I lowered the cisplatin concentration to a dose that still revealed SHyp with *rad1*Δ but now allowed *rad1*Δ*rev3*Δ cells to divide sufficiently for a fluctuation assay. Loss of *REV3* or its partner *REV1* led to complete suppression of *rad1*Δ-cisplatin SHyp (**Fig. 8B**), indicating that POLζ TLS was responsible for the mutations. To determine if endogenous DNA lesions could promote 3'-templated mutations, we scored a published dataset of *URA3* mutations sequenced in unperturbed NER-deficient (*rad14*Δ) cells with or without *TLS* polymerase deletions (30). Remarkably, 3' templating was observed at 43% of SNVs, deletion of *RAD30* had no effect on this frequency while *REV3* deletion reduced the frequency of SNVs matching the 3' base to 29% (**Fig. 8C**). Together these data

led me to propose a model in which slippage and realignment of the DNA template within Rev3 led to the observed mutation signature and is generalizable to endogenous lesions (**Fig. 8D** and discussed below).

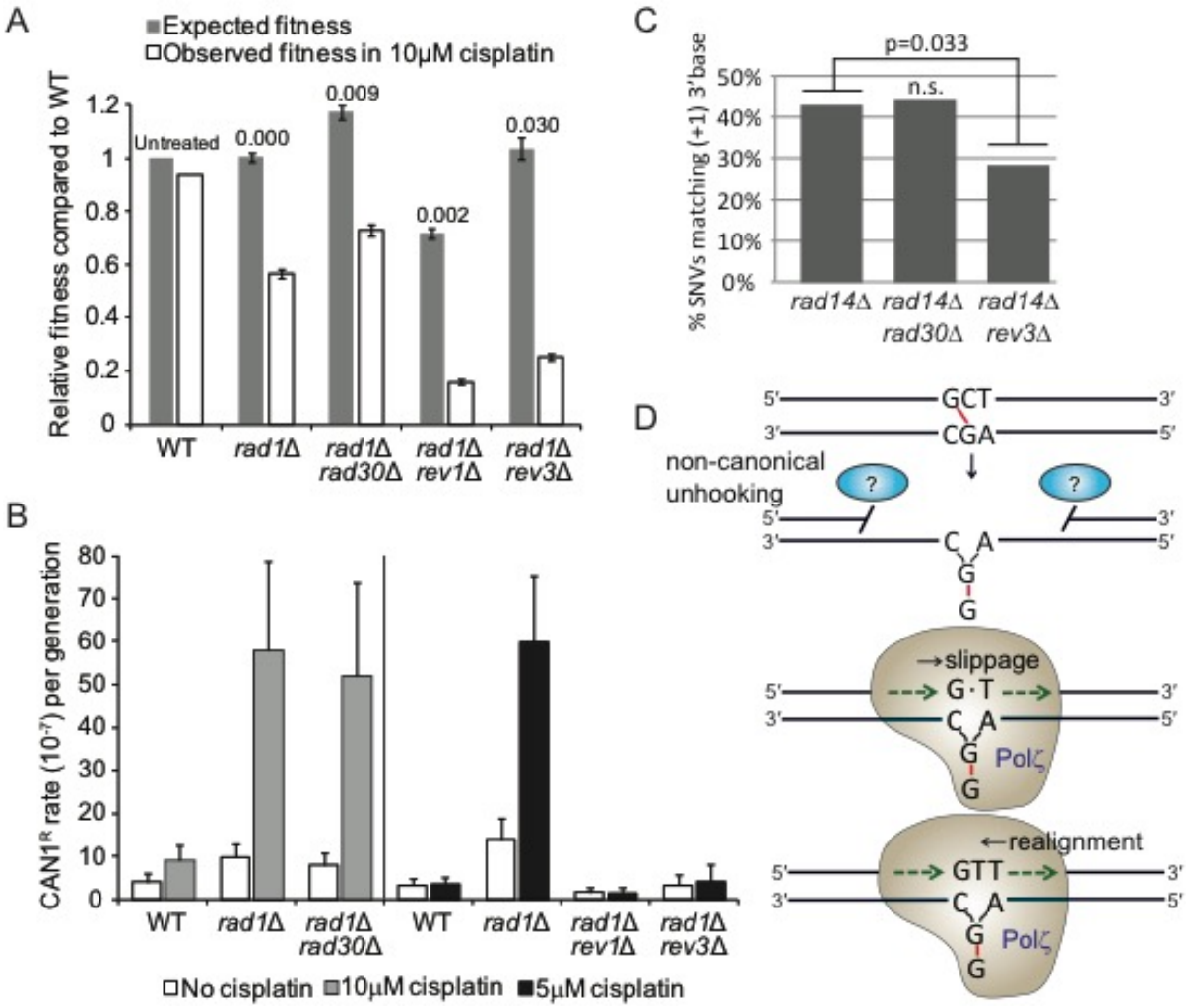


Figure 8. Role of translesion synthesis in hypermutation of cisplatin treated *rad1*Δ cells. (A) Fitness measurements and (B) mutation rates in *rad1*Δ cells lacking the indicated TLS polymerase subunit. (C) Frequency of SNVs identical to 3' base in *URA3* from the indicated strains from (23). (D) Model for slippage and realignment of DNA polymerase ζ (Rev3) at a cisplatin damaged base. Non-canonical and presumably not optimal unhooking of the ICL creates a substrate for error-prone TLS that promotes slippage of Polζ at the lesion and subsequent realignment upon lesion bypass leading to a 3'-templated SNV.

Chapter 4: CONCLUDING CHAPTER

4.1 Conclusions and Discussion

4.1.1 Synthetic Hypermutation as an Alternative Measure of Phenotypic Enhancement in Genome Maintenance

As simple fitness datasets on gene-gene and gene-drug interactions accumulate (16, 34) researchers are turning to measuring trigenic interactions (35), or the ways in which chemical stresses shift genetic-interaction networks (e.g. synthetic cytotoxicity (17, 18)) to enhance the wiring diagram of the cell (36). Still other approaches have combined high-content imaging with genetic and chemical perturbations to create conditional phenotypic networks (e.g. Lsm1-GFP imaged in the YKO collection +/- hydroxyurea (37)). For genome maintenance pathways one of the most relevant phenotypes beyond fitness is the rate of mutations. While there is prior evidence that ultra-high mutation rates can drive severe fitness consequences, even lethality (38), the overlap of these phenotypes broadly was not known. By systematically ablating the major DNA repair pathways in yeast and measuring the effects of diverse genotoxic chemicals we show that fitness and chemically induced hypermutation are not necessarily linked.

SHyp interactions between 5FU and *mlh1*Δ and *apn1*Δ are largely consistent with the role of MMR and BER respectively in removing uracil incorporated into DNA (39). Indeed, these data show that 5FU can overwhelm the repair capacity of MMR or BER deficient cells without additionally affecting the fitness of MMR impaired cells. This is consistent with evidence that 5FU has multiple modes of action interfering with both DNA and RNA metabolism (39). On the other hand, while both NER and HR are required for repair of

cisplatin interstrand crosslinks and mutants in these pathways had fitness defects, *rad1*Δ but not *sgs1*Δ or other HR mutants showed SHyp with cisplatin. This suggests that while survival in *rad1*Δ cells is supported by an error prone mechanism, that same pathway is insufficient to support *sgs1*Δ cells which either die, likely due to toxic recombination intermediates, or repair cisplatin damage in an error-free way.

4.1.2 A Mutation Signature of TLS Polymerase Slippage and Realignment

Whole genome sequencing analysis of model organism genomes that have been mutagenized under specific conditions is a powerful way to link mutation signatures to specific causal phenomena (Reviewed in (40)). This approach has been used to shed new light on DNA repair mechanism (MMR,(41)), the effects of genotoxic chemicals (14), and the role of cytidine deaminase enzymes in cancer (42-44). Characterization of the sequencing data revealed that in *rad1*Δ yeast interstrand crosslinks were the predominant source of mutations since most variants were found at GpC motifs. This differs from the cisplatin mutation signature of both *C. elegans* and humans where intrastrand crosslinks seem to drive mutagenesis (i.e. mutations occur at CpC motifs (8, 14)). There are several possible reasons for this including the complexity of the Fanconi Anemia pathway for interstrand crosslink repair in humans that is only partly represented in yeast (45), or the high divergence of *REV3* itself (i.e. human REV3L protein is >twice as long, and exhibits only 25% similarity to REV3L). While Rev1 plays an important structural role, we predict Rev3 is catalyzing the mutation signature observed in this study, since the catalytic role of Rev1 to insert C opposite the damaged guanine would be error-free and thus not detectable in this assay. We also found evidence of Rev3 slippage and realignment in mutations from untreated cells, suggesting that endogenous, non-cisplatin-induced lesions can promote the same mechanism (30).

The mutation signature analysis revealed not only a different type of mutagenic lesion in yeast, but also a mechanism involving yeast Polζ, Rev3, undergoing a slippage and realignment event during replication. Previous work *in vitro* has identified this activity in another TLS polymerase, human DNA polκ (46). We propose that (**Fig. 8D**), at least in the specific context of cisplatin adducts, the guanine lesion is bypassed by Rev3 in a manner that uses Watson-Crick basepairing with the undamaged +1 base. This may occasionally lead to an indel, but more often the newly inserted base shifts to mispair with the damaged G residue and is extended with the correct base at the 3' position. Whether loss of *RAD1* simply increases the frequency of such events, or whether suboptimal DNA incision by compensatory nucleases favor this misincorporation event is an active area of research. Indeed, the nature of DNA incision in the absence of Rad1 is unclear, although structure-specific nucleases such as Slx4, Mus81 and Yen1 have activities that could in principle complement Rad1 loss at some structures (47). Alternatively, new data in *Xenopus* extracts implicate the N-glycosylase activity of NEIL3, an upstream BER factor, in unhooking psoralen and other interstrand crosslinks (48). Therefore it is also possible that yeast N-glycosylases could act on cisplatin crosslinks in the absence of the preferred Rad1-pathway and create a unique chemical moiety that drives slippage-realignment TLS by POLζ *in vivo*.

4.1.3 Synthetic Hypermutation and Cancer

The framework established here sought to establish the gene-drug relationships driving hypermutation in cells surviving genotoxin treatment beginning with canonical DNA repair pathways and drugs of clinical relevance. There remains a large space to be explored; in yeast there are many hundreds of genes whose loss of function causes genome instability and hundreds of others whose increased dosage causes genome instability (13, 20, 49). The

ways in which these will combine with each other and with various chemicals to permit mutagenesis remains incomplete.

Hypermutation of surviving cancer cells after genotoxic chemotherapy could be viewed as negative if it permits the acquisition of chemoresistant mutations. However, current evidence suggests that resistance mutations are often pre-existing in tumor cell populations and their emergence is a clonal selection process rather than chemotherapy induced mutations (50). Indeed, even for therapy-related leukemias it was recently shown that they bear the same number of somatic SNVs as *de novo* leukemias and that p53 mutations driving leukemogenesis were pre-existing when chemotherapy occurred (51). Thus, synthetic hypermutation may not be a major concern for acquiring chemoresistant mutations when treating repair-deficient cancers with genotoxic chemotherapies. While chemotherapy has been shown to minimally impact chemoresistant mutations, the DNA damaging nature of chemotherapeutic agents can increase the overall mutational load. (8) This work proposes that the hypermutation of surviving cells could ultimately be beneficial in contexts where an immune response to tumor neo-antigens is desired as is the case for therapies targeting immune checkpoints (52). The efficacy of immune checkpoint blockade has been correlated with mutational load in lung cancer (53), melanoma (54), and MMR deficient colorectal cancers (55). Understanding more about the interactions of genotoxic chemotherapies with genome maintenance defects in cancer could ultimately support interventions where it is desirable to control or even increase the mutational load of cancer cells following frontline therapy.

4.2 Significance

Cancer cells often have defects in DNA repair and are killed effectively by drugs that damage DNA. However surviving cells can acquire additional mutations after treatment with these genotoxic chemicals. This work applies a simple model system to reveal synergy between specific DNA repair mutations and genotoxic drugs that occurs independently of fitness defects. Moreover, by analyzing the entire genome of a mutagenized cell population this work identifies a signature of mutations that informs the mechanism of the translesion synthesis DNA damage tolerance pathway. This work establishes a conceptual framework for predicting the mutational burden of cells surviving genotoxin treatment and demonstrates the utility of model organism mutation signature analysis for generating mechanistic insights.

4.3 Future Directions

In order to elucidate the nature of the incision that is made in the absence of Rad1, it would be necessary to test candidate yeast nucleases like Slx4, Mus81 and Yen1 and N-glycosylases Ntg1 and Ntg2 in the *rad1*Δ-cisplatin context. This could reveal the identity/ies of the protein/s replacing Rad1 in its role of incision at or around damaged bases. Even though this work already covers major groups of chemotherapeutic agents, radiomimetic and replication inhibitor groups of drugs are not covered in this study. In this regard, it would be important to add at least bleomycin (radiomimetic) and hydroxyurea (replication inhibitor) to the screens of gene-drug interactions. This addition would further complete this work by covering all major DNA-damaging agents. In the future, the network of chemicals tested could be expanded immensely to include novel DNA damaging agents, environmental genotoxins or other molecules. Whole-genome sequencing of other SHyp phenotypes, in

particular differentiating SHyp in *apn1*Δ-5FU and *mlh1*Δ-5FU, would extend this project and potentially reveal other signatures that can be used to suggest new DNA repair mechanisms and functions. Finally, analyzing cancer genome databases for signatures similar to those in this thesis could demonstrate that a similar mechanism is operational in some cancer genomes. This would open to the door to translating the SHyp phenotypes into human cell lines to reproduce the synthetic hypermutator effect and could yield important insights for cancer therapy and outcome.

BIBLIOGRAPHY

1. Farlow A*, et al.* (2015) The Spontaneous Mutation Rate in the Fission Yeast Schizosaccharomyces pombe. *Genetics* 201(2):737-744.
2. Curtin NJ (2012) DNA repair dysregulation from cancer driver to therapeutic target. *Nature reviews.Cancer* 12(12):801-817.
3. Loeb LA (2011) Human cancers express mutator phenotypes: origin, consequences and targeting. *Nature reviews.Cancer* 11(6):450-457.
4. Stratton MR, Campbell PJ, & Futreal PA (2009) The cancer genome. *Nature* 458(7239):719-724.
5. Lord CJ & Ashworth A (2013) Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nature medicine* 19(11):1381-1388.
6. Farmer H*, et al.* (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434(7035):917-921.
7. Lord CJ & Ashworth A (2012) The DNA damage response and cancer therapy. *Nature* 481(7381):287-294.
8. Huang KK*, et al.* (2016) Exome sequencing reveals recurrent REV3L mutations in cisplatin-resistant squamous cell carcinoma of head and neck. *Scientific reports* 6:19552.
9. Alexandrov LB*, et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.
10. Forbes SA*, et al.* (2010) COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic acids research* 38(Database issue):D652-657.
11. Nik-Zainal S*, et al.* (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5):979-993.
12. Serero A, Jubin C, Loeillet S, Legoix-Ne P, & Nicolas AG (2014) Mutational landscape of yeast mutator strains. *Proceedings of the National Academy of Sciences of the United States of America* 111(5):1897-1902.
13. Stirling PC, Shen Y, Corbett R, Jones SJ, & Hieter P (2014) Genome destabilizing mutator alleles drive specific mutational trajectories in Saccharomyces cerevisiae. *Genetics* 196(2):403-412.
14. Meier B*, et al.* (2014) C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome research* 24(10):1624-1636.
15. Costanzo M*, et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353(6306).
16. Lee AY*, et al.* (2014) Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* 344(6180):208-211.
17. Martin H*, et al.* (2015) Differential genetic interactions of yeast stress response MAPK pathways. *Molecular systems biology* 11(4):800.
18. Li X, O'Neil NJ, Moshgabadi N, & Hieter P (2014) Synthetic cytotoxicity: digenic interactions with TEL1/ATM mutations reveal sensitivity to low doses of camptothecin. *Genetics* 197(2):611-623.

19. Giaever G & Nislow C (2014) The yeast deletion collection: a decade of functional genomics. *Genetics* 197(2):451-465.
20. Stirling PC*, et al.* (2011) The Complete Spectrum of Yeast Chromosome Instability Genes Identifies Candidate CIN Cancer Genes and Functional Roles for ASTRA Complex Components. *PLoS genetics* 7(4):e1002057.
21. McLellan J*, et al.* (2009) Synthetic lethal genetic interactions that decrease somatic cell proliferation in Caenorhabditis elegans identify the alternative RFC CTF18 as a candidate cancer drug target. *Molecular biology of the cell* 20(24):5306-5313.
22. Lang GI & Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. *Genetics* 178(1):67-82.
23. Hall BM, Ma CX, Liang P, & Singh KK (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics (Oxford, England)* 25(12):1564-1565.
24. Hoffman CS & Winston F (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene* 57(2-3):267-272.
25. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25(14):1754-1760.
26. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25(16):2078-2079.
27. Jones SJ*, et al.* (2010) Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome biology* 11(8):R82-2010-2011-2018-r2082. Epub 2010 Aug 2019.
28. Chan K*, et al.* (2015) An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* 47(9):1067-1072.
29. O'Shea JP*, et al.* (2013) pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods* 10(12):1211-1212.
30. Stone JE, Lujan SA, Kunkel TA, & Kunkel TA (2012) DNA polymerase zeta generates clustered mutations during bypass of endogenous DNA lesions in Saccharomyces cerevisiae. *Environmental and molecular mutagenesis* 53(9):777-786.
31. Jasin M & Rothstein R (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology* 5(11):a012740.
32. Marteijn JA, Lans H, Vermeulen W, & Hoeijmakers JH (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature reviews. Molecular cell biology* 15(7):465-481.
33. Eastman A (1987) The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. *Pharmacology & therapeutics* 34(2):155-166.
34. Costanzo M*, et al.* (2010) The genetic landscape of a cell. *Science (New York, N.Y.)* 327(5964):425-431.
35. Braberg H*, et al.* (2014) Quantitative analysis of triple-mutant genetic interactions. *Nature protocols* 9(8):1867-1881.
36. Boone C (2014) Yeast systems biology: our best shot at modeling a cell. *Genetics* 198(2):435-437.

37. Tkach JM, *et al.* (2012) Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature cell biology* 14(9):966-976.
38. Herr AJ, Kennedy SR, Knowels GM, Schultz EM, & Preston BD (2014) DNA replication error-induced extinction of diploid yeast. *Genetics* 196(3):677-691.
39. Wyatt MD & Wilson DM, 3rd (2009) Participation of DNA repair in the response to 5-fluorouracil. *Cellular and molecular life sciences : CMLS* 66(5):788-799.
40. Segovia R, Tam AS, & Stirling PC (2015) Dissecting genetic and environmental mutation signatures with model organisms. *Trends Genet*.
41. Ma X, *et al.* (2012) Mutation hot spots in yeast caused by long-range clustering of homopolymeric sequences. *Cell reports* 1(1):36-42.
42. Taylor BJ, *et al.* (2013) DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* 2:e00534.
43. Lada AG, *et al.* (2012) AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biology direct* 7:47; discussion 47-6150-6157-6147.
44. Roberts SA, *et al.* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 45(9):970-976.
45. Daee DL & Myung K (2012) Fanconi-like crosslink repair in yeast. *Genome integrity* 3(1):7.
46. Mukherjee P, Lahiri I, & Pata JD (2013) Human polymerase kappa uses a template-slippage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. *Nucleic Acids Res* 41(9):5024-5035.
47. Munoz-Galvan S, *et al.* (2012) Distinct roles of Mus81, Yen1, Slx1-Slx4, and Rad1 nucleases in the repair of replication-born double-strand breaks by sister chromatid exchange. *Mol Cell Biol* 32(9):1592-1603.
48. Semlow DR, Zhang J, Budzowska M, Drohat AC, & Walter JC (2016) Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell* 167(2):498-511 e414.
49. Duffy S, *et al.* (2016) Overexpression screens identify conserved dosage chromosome instability genes in yeast and human cancer. *Proceedings of the National Academy of Sciences of the United States of America*.
50. Bhang HE, *et al.* (2015) Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nature medicine* 21(5):440-448.
51. Wong TN, *et al.* (2015) Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* 518(7540):552-555.
52. Schumacher TN & Schreiber RD (2015) Neoantigens in cancer immunotherapy. *Science* 348(6230):69-74.
53. Rizvi NA, *et al.* (2015) Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348(6230):124-128.
54. Snyder A, *et al.* (2014) Genetic basis for clinical response to CTLA-4 blockade in melanoma. *The New England journal of medicine* 371(23):2189-2199.
55. Le DT, *et al.* (2015) PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England journal of medicine* 372(26):2509-2520.
56. Dietlin F, *et al.* (2014) Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends Genet* 30(8):326-39.

57. Chapman JR, *et al.* (2012) Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* 47(4):497-510.
58. Nimonkar AV, *et al.* (2011) BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA-break repair. *Genes Dev* 25(4):350-62.
59. Brosh M (2013) DNA helices involved in DNA repair and their roles in cancer. *Nat Rev Cancer* 13(8): 542-558.
60. Sarbajna S & West SC (2014) Holliday junction processing enzymes as guardians of genome instability. *Trends Biochem Sci* 39(9):409-19.
61. Meindl A, *et al.* (2010) Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* 42(5):410-4.
62. Al-Sukhni W, *et al.* (2008) Germline BRCA1 mutations predispose to pancreatic adenocarcinoma. *Human Genet* 124(3):271-8.
63. Greer JB, *et al.* (2007) Role of BRCA1 and BRCA2 mutations in pancreatic cancer. *Gut* 56(5):601-5.
64. Hartlerode AJ, *et al.* (2009) Mechanisms of double-strand break repair in somatic mammalian cells. *Biochem J* 423(2):157-68.
65. Mahaney BL, *et al.* (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417(3):639-50.
66. Deriano L, *et al.* (2013) Modernizing the nonhomologus end-joining repertoire: alternative and classical NHEJ share the stage. *Ann Rev Genet* 47:433-55.
67. Shuck SC, *et al.* (2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res* 18(1):64-72.
68. Ogi T, *et al.* (2010) Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell* 37(5):714-27.
69. Moser J, *et al.* (2007) Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III α in a cell-cycle-specific manner. *Mol Cell* 27(2):311-23.
70. Hegde ML, *et al.* (2008) Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res* 18(1):27-47.
71. David SS, *et al.* (2007) Base-excision repair of oxidative DNA damage. *Nature* 447(7147):941-50.
72. Nouspikel T (2009) DNA repair in mammalian cells: nucleotide excision repair: variations on versatility. *Cell Mol Life Sci* 66(6):994-1009.
73. Jiricny J (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* 7(5):335-46.
74. Li GM (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res* 18(1):85-98.
75. Kim H, *et al.* (2012) Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Gene Dev* 26(13):1393-408.
76. Kottemann MC, *et al.* 2013. Fanconi anemia and the repair of Watson and Crick crosslinks. *Nature* 493(7432):356-63.
77. Deans AJ, *et al.* (2011) DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* 11(7):467-80.
78. Sale JE, *et al.* (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol* 13(3):141-52.

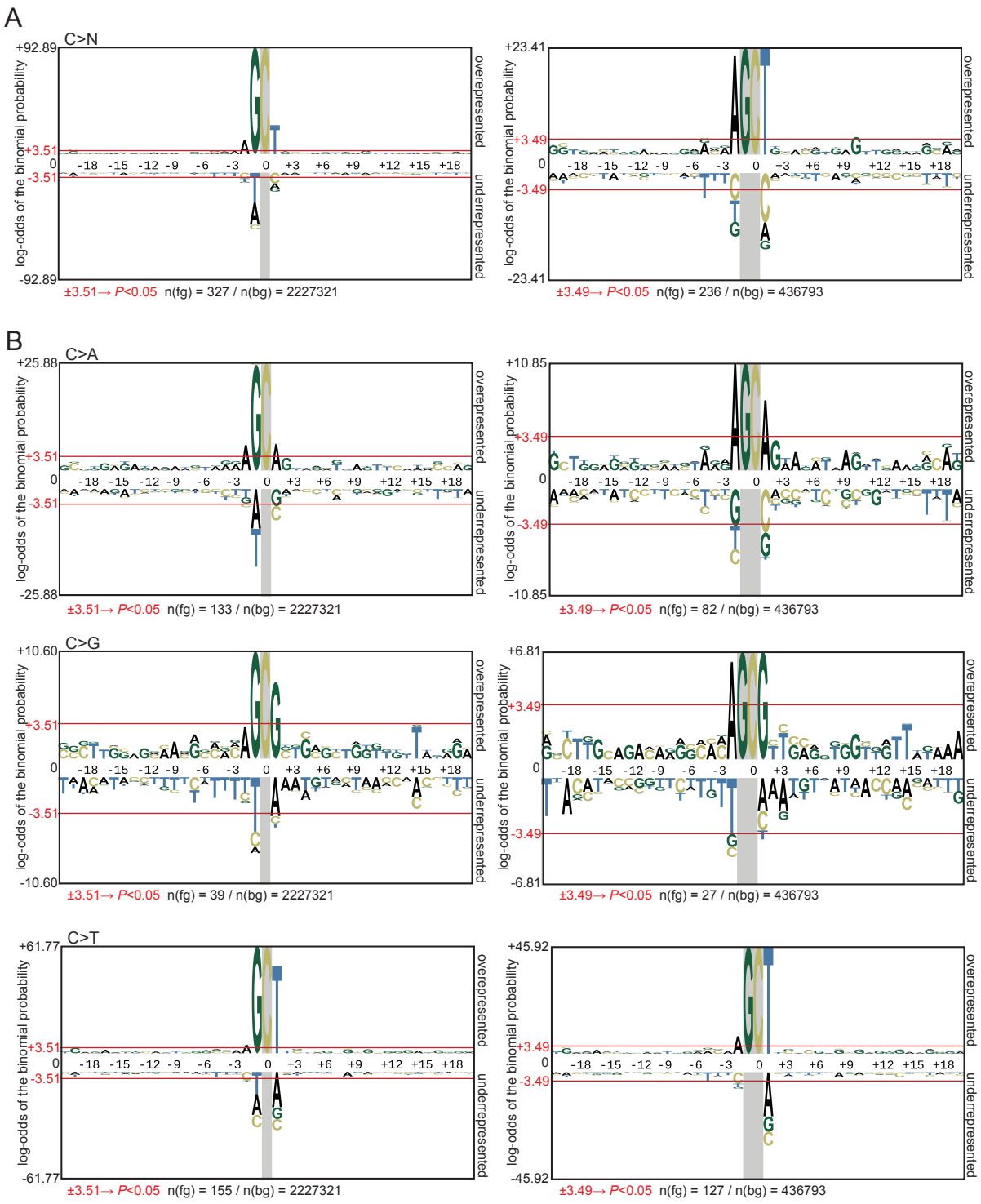
79. Prakash S, *et al.* (2005) Eukaryotic translesion synthesis DNA polymerases: specificity of structure. *Annu Rev Biochem* 74:317-53.
80. Shcherbakova PV, *et al.* (2006) Translesion synthesis DNA polymerases and control of genome stability. *Front Biosci* 11:2496-517.
81. Roberts SA, *et al.* (2012) Clustered mutations in yeast and in human cancers can arise from damaged long single-stranded DNA regions. *Mol Cell* 46(4):424-35.
82. Puente XS, *et al.* (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475(7354):101-5.
83. Machida K, *et al.* 2004. Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and protooncogenes. *Proc Natl Acad Sci USA* 101(12):4262-7.
84. Totoki Y, *et al.* (2011) High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* 43(5):464-9.
85. Kunkel TA, *et al.* (2005) DNA mismatch repair. *Annu Rev Biochem* 74:681-710.
86. Vilar E, *et al.* (2010) Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol* 7(3):153-62.
87. Chu WK, *et al.* (2009) RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer* 9(9):644-54.
88. Altmann T, *et al*. (2016) DNA ligase IV syndrome; a review. *Orphanet J Rare Dis* 11(1):137.
89. Moshous D, *et al.* (2003) Partial T and B lymphocite immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in Artemis. *J Clin Invest* 111(3):381-7.
90. Ciccia A, *et al.* (2010) The DNA damage response: making it safe to play with knives. *Mol Cell* 40(2):179-204.
91. Cleaver JE, *et al.* (2009) Disorders of the nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet* 10(11):756-68.
92. Helleday T, *et al*. (2008) DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* 8(3):193-204.
93. Seiple L, *et al.* (2006) Linking uracil base excision repair and 5-fluorouracil toxicity in yeast. *Nucleic Acid Res* 34(1):140-51.
94. Ikegami S, *et al.* (1978) Aphidicolin prevents mitotic cell division by interfering with the ability of DNA polymerase-alpha. *Nature* 275(5679):458-60.
95. Bianchi V, *et al.* (1986) Changes of deoxyribonucleoside triphosphate pools induced by hydroxyurea and their relation to DNA synthesis. *J Biol Chem* 261(34):16037-42
96. Howard BD, *et al.* (1964) Identification of the altered bases in mutated single-stranded DNA. II. In vivo mutagenesis by 5- bromodeoxyuridine and 2-aminopurine. J Mol Biol 9:364-71.
97. Setlow RB, *et al.* (1966) Pyrimidine dimers in ultraviolet-irradiated DNA's. *J Mol Biol* 17(1):237-54.
98. Rodin SN, *et al.* (2005) Origins and selection of p53 mutations in lung carcinogenesis. *Semi Cancer Biol* 15(2):103-12.
99. Wogan GN (1992) Aflatoxins as risk factors for hepatocellular carcinoma in humans. *Cancer Res* 52(7 Suppl):2114s-2118s.
100. Pleasance ED, *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463(7278):191-6.

101. Pleasance ED, *et al.* (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463(7278):184-90.
102. Parsons DW, *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321(5897):1807-12.
103. Cancer Genome Atlas Research Network (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499(7456):43-9.
104. Cancer Genome Atlas Research Network (2013) Integrated genomic characterization of endometrial carcinoma. *Nature* 497(7447):67-73.
105. Hoang ML, *et al.* (2013) Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Sci Transl Med* 5(197):197ra102.
106. Pfeifer GP, *et al.* (2005) Mutations induced by ultraviolet light. *Murat Res* 571(1-2):19-31.
107. Zhu YO, *et al*. (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* 111(22):E2310-8.
108. Lynch M, *et al.* (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105(27):9272-7.
109. Nishant KT, *et al.* (2010) The Baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLos Genet* 6(9):e1001109.
110. Zanders S, *et al.* (2010) Detection of heterozygous mutations in the genome of mismatch repair defective diploid yeast using a Bayesian approach. *Genetics* 186(2):493-503.
111. Lang GI, *et al.* (2013) Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3 (Bethesda)* 3(9):1453-65.
112. Shlien A, *et al.* (2015) Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet* 47(3):257-62.
113. Larrea AA, *et al.* (2010) Genome-wide model for the normal eukaryotic DNA replication fork. *Proc Natl Acad Sci USA* 107(41):17674-9.
114. Lujan SA, *et al*. (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res* 24(11):1751-64.
115. Roberts SA, *et al.* (2012) Clustered mutations in yeast and in human cancers can arise from damaged long single-stranded DNA regions. *Mol Cell* 46(4):424-35.
116. Flibotte S, *et al.* (2010) Whole-genome profiling of mutagenesis in *Caenorhabditis elegans*. *Genetics* 185(2):431-41.
117. Sarin S, *et al.* (2010) Analysis of multiple ethyl methanesulfonate-mutagenized *Caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* 185(2):417-30.
118. Alexandrov LB, *et al.* (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 3(1):246-59.
119. Alexandrov LB & Stratton MR (2014) Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev* 24:52-60.

APPENDICES

	wildtype	*apn1*Δ	*exo1*Δ	*mlh1*Δ	*mph1*Δ	*mus81*Δ	*rad1*Δ	*rad5*Δ	*rad30*Δ	*rad52*Δ	*rev3*Δ	*sgs1*Δ	*yku80*Δ
No drug	3.56	7.93	11.25	45.84	11.16	4.50	10.98	13.89	3.35	21.07	2.52	4.14	2.82
95% CI	2.86	3.86	7.79	30.04	7.74	2.71	7.70	9.54	1.94	14.70	1.34	2.47	1.51
	4.31	12.99	15.17	63.98	15.02	6.60	14.66	18.81	5.04	28.25	2.52	6.11	4.41
CPT	4.32	9.43	15.15	39.69	9.10	6.61	10.28	12.61	3.76	16.91	3.77	5.13	3.95
95% CI	3.26	4.83	10.80	29.86	6.20	4.20	7.07	8.67	2.21	10.94	2.18	3.05	2.20
	5.49	15.09	20.02	50.51	12.40	6.61	13.91	17.08	5.59	23.81	5.67	7.59	6.07
CIS	8.40	8.77	37.83	41.09	37.90	5.88	63.39	lethal	7.66	32.91	6.82	10.57	7.37
95% CI	6.62	4.45	24.41	30.48	28.08	3.64	48.92	-	4.95	21.82	4.15	6.54	4.67
	10.33	14.10	53.31	52.83	48.77	8.51	79.20	-	10.78	45.61	9.97	15.28	10.51
ETP	4.76	5.07	14.38	40.93	14.06	5.34	9.59	14.02	3.50	11.16	3.27	5.82	3.40
95% CI	3.64	2.17	10.28	30.98	9.96	3.30	6.66	9.52	2.05	6.80	1.87	3.57	1.91
	5.99	8.81	18.97	51.86	18.65	7.73	12.90	19.14	5.23	16.27	4.95	8.46	5.18
5FU	2.01	160.44	10.16	118.82	4.92	1.08	8.58	13.21	1.01	18.18	0.29	4.09	2.00
95% CI	1.23	99.16	6.96	85.86	2.95	0.42	5.67	8.79	0.40	8.92	0.06	2.18	0.95
	2.88	232.04	13.80	155.60	7.25	1.96	11.90	18.27	1.80	29.70	0.64	6.41	3.30
MMS	8.41	17.26	52.39	37.29	53.08	8.90	20.32	lethal	12.40	59.38	5.02	11.12	6.58
95% CI	6.74	10.18	36.05	28.44	36.28	5.8	12.29	-	8.78	44.29	3.02	7.14	3.86
	10.22	25.64	70.91	47	72.14	12.47	29.75	-	16.48	76.04	7.37	15.72	9.81

Appendix A. CAN1 mutation rates ($x10^{-7}$) and 95% confidence intervals.



Legend on next page.

Appendix B. Extended pLogos for cisplatin-induced mutations in *rad1*Δ yeast from figure 7. (A) 41-mers sequence context for C>N fixed at C (left) and GC (right). (B) 41-mers sequence context for C>A, C>G and C>T with fixed positions as in A. Mutated C is centered. Red lines represent cutoffs of 0.05. n(fg) denotes the number of sequences with mutations at C and GC in the set of 41-mers (foreground) and n(bg) the number of 41-mer sequences created by the pLogo tool from the yeast genome (background).